University of Southampton Research Repository

University of Southampton

Faculty of Engineering and Physical Sciences
School of Physics and Astronomy
Southampton High Energy Physics Group

Novel b-jet Analyses for Beyond the Standard Model Higgs Bosons at the LHC

by

Billy George Ford
MPhys
ORCiD: 0000-0003-0061-577X

A thesis for the degree of
Doctor of Philosophy

December 2021

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Physics and Astronomy

Doctor of Philosophy

Novel b-jet Analyses for Beyond the Standard Model Higgs Bosons at the LHC

by Billy George Ford

Despite the many successes of the LHC since running began in 2008 - such as the discovery of the Higgs Boson in 2012 - the hunt for new physics beyond the Standard Model (SM) remains elusive.

One particularly appealing extension to the SM comes in the form of Two-Higgs-Doublet-Models (2HDMs), which provide an enriched scalar sector of additional Higgs particles (the CP-even H and h, CP-odd A, and charged $H^{\pm}$). Where kinematically possible, interactions between these Higgses can lead to high b-jet multiplicity final states via $H \to hh \to bb\bar{b}\bar{b}$ decays.

In this thesis, various novel approaches to observing such new physics are considered. An alternative to traditional jet clustering algorithms, using a variable-R cone dependent on jet p_T, is shown to increase the potential signal significance when modelled against the leading backgrounds.

Furthermore, the use of high level machine learning is investigated. By mapping p_T-weighted pixels in a detector into images, we build a convolutional neural network (CNN) to classify wide cone b-jets in signal events coming from 2HDM decays, against the leading backgrounds.

Finally, we present a novel approach to jet reconstruction using spectral clustering machine learning techniques, and compare the performance with the currently well-established methods in use at the LHC.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as: [1] and [2]

Signed:.. Date:..................

Acknowledgements

There are countless people to whom I would like to extend my deepest gratitude for their support over the course of my post graduate studies.

Firstly, to my supervision team of Stefano Moretti, Srinandan Dasmahapatra, and Claire Shepherd-Themistocleous. For their professional support in my development as a young physicist, along with the entire extended group involved in our research endeavours - Henry Day-Hall, Giorgio Cerro, Shubhani Jain, Emmanuel Olaiya, Amit Chakraborty and Jacan Chaplais.

Having started at the University of Southampton as an undergraduate in 2012, I am possibly the longest serving player (and almost certainly record goalscorer!) for the Physics Society Men's Football Club - Physoc FC. I give thanks to all team mates, past and present, for the great times both on and off the pitch.

During my time as a Ph.D. student in Southampton I have been incredibly lucky to have worked alongside so many tremendous people, with particular mention to Matthew Russell, Michele Santagata, Ross Glew, Simon King, Sam Rowley, Adam Forster, Alex Mitchell, Aaron Poole, Sami Rawash, in addition to the countless others who I have had the pleasure of working alongside.

Away from the university environment, I thank my friends. To the West Ham boys for the constant source of laughter, to the basement boys and our fortnightly lockdown chats. To Mel and George and the countless beers we have shared, and to Ashleigh and Sophie for the encouragement, living room dance floors, and of course the brunches!

I am incredibly grateful to the support received from my family, Mum - for inspiring me with your strength; Dad - for all of the support, and putting up with my rants over the phone!

Finally, to my partner Stacie. You have been by my side every step of the way, from Southampton to Seattle! None of this would have been possible without your constant love, support and understanding. I can't wait to embark on our next adventure together!

Part I

Introduction and Background

Chapter 1

Introduction

The ultimate goal of particle physics is to provide a complete description of the foundational building blocks of our universe. In fact the very beginnings of particle physics can be traced back to an astoundingly simple question posed by Greek philosophers, if I divide an object into progressively smaller pieces, can I continue ad infinitum, or will I reach a fundamental piece that cannot be further broken, an “uncuttable”, or “atom”?

It was in the 19th century when subatomic particles graduated from the realm of philosophical speculation into scientific reality, and since then the following century gave rise to an avalanche of particle discoveries, from the electron in 1897, up to the eventual discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC).

It was the Higgs Boson that provided the final piece of the Standard Model (SM) puzzle, our current best fundamental theory of nature, that has held up remarkably well against LHC experiments. Despite this, there exist well known shortcomings of the SM, and so the hunt for a more complete theory of particle physics continues. Areas of active research range from complete reworks of the language of fundamental physics, such as string theory, or reworks of the group structure driving fundamental interactions, to experimental searches for simpler extensions to the SM, via precision measurements and new particle searches.

In this thesis, approaches of phenomenologically probing new theories beyond the Standard Model (BSM) are considered. In particular Two-Higgs-Doublet-Models (2HDMs) comprised of an extended scalar structure of the SM, yielding additional Higgs particles that can potentially be observed at the LHC via high b-jet multiplicity final states.

The outline of this thesis is as follows; chapters 2-5 introduce the theoretical groundwork and motivation for the research activities undertaken. In chapter 2, a overview of the SM is presented, with focus on the Higgs sector. The shortcomings of the SM are also addressed. In chapter 3 a theoretical review of 2HDMs is presented, and chapter 4 discusses the physics of jets, as well as clustering algorithms, in the context of the LHC.

Finally in chapter 5, the field of machine learning is introduced, and application within high energy physics is discussed.

Chapters 6-8 present published research work, in chapter 6 the methodologies and results of the paper “Revisiting Jet Clustering Algorithms for New Higgs Boson Searches in Hadronic Final States”, where a jet clustering algorithm using a variable-R is compared with traditional fixed cone algorithms in the context of potential 2HDM searches.

Chapter 7 describes recent research activity on jet visualisation, exploring the ability of convolutional neural networks (CNNs) to classify signal 2HDM processes from the leading backgrounds, when represented as jet images mapped in detector (η,ϕ) space.

In chapter 8, a completely novel approach to jet clustering is presented, utilising spectral clustering, from the work “Spectral Clustering for Jet Physics”.

Finally, chapter 9 presents concluding remarks of the various research endeavours, as well as discussing possible future work moving forward into future iterations of the LHC, such as the High Luminosity Large Hadron Collider (HL-LHC).

Chapter 2

The Standard Model of Particle Physics

For many years the Standard Model has been the crowning achievement of theoretical particle physics, and provides the most complete description of nature at a fundamental level we currently have. In field theory language, the SM is a renormalisable QFT with a local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry, corresponding to three of the four fundamental forces of nature, Quantum Chromodynamics (QCD), the Weak Interaction, and Quantum Electrodynamics (QED).

2.1 The Brout-Englert-Higgs Mechanism in the SM

It would be an understatement to say that the discovery of the Higgs boson at the LHC in 2012 had been overdue, in fact its existence had been predicted by Peter Higgs, François Englert and Robert Brout back in 1964 [3], and utilised as a method of explaining the origin of mass for the the weak gauge bosons $W^{\pm}$ and Z^0 by Steven Weinberg in 1967 [4]. Unlike their counterparts the photon γ and gluon g, the weak bosons had experimentally verified masses which had so far been unexplained in the SM.

2.1.1 The Higgs Potential

The scalar sector of the SM Lagrangian takes the following form [5]

$$\mathcal{L}_H = (D^{\mu}\phi)^*(D_{\mu}\phi) - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2. \tag{2.1}$$

If we firstly consider the case where ϕ is a complex scalar field, the lagrangian has a global $U(1)$ invariance under transformations like

$$\phi \rightarrow e^{i\alpha}\phi, \tag{2.2}$$

and D^μ is the covariant derivative $D^\mu = \partial^\mu + iq\mathcal{A}^\mu$. In such a case, for $\mu^2 > 0$, there is a single vacuum state with vanishing field expectation value $\phi_0 = 0$. If however $\mu^2 < 0$, the potential takes on the form (shown in Fig.2.1)

$$V(\phi) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2. \tag{2.3}$$

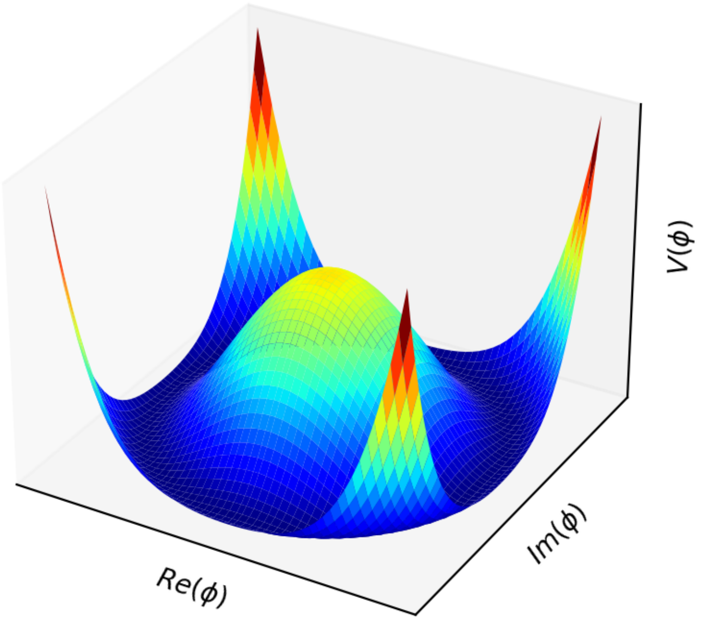


Figure 2.1: The ‘Mexican hat’ or ‘wine bottle’ Higgs potential, with an infinite ring of minima at $\phi_0 \neq 0$.

The shape of the potential is such that the minimum does not occur at $\phi_0 = 0$, but instead there exists an infinite ring of minima. The minimum, known as the vacuum expectation value (v.e.v), is found, setting $\frac{dV(\phi)}{d\phi} = 0$,

$$\phi^0 = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}. \tag{2.4}$$

Moving on to the Standard Model, the gauge group of which is $SU(2)_L \times U(1)_Y$, the Lagrangian is written [6]

$$\mathcal{L}_H = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \tag{2.5}$$

where ϕ is a complex $SU(2)_L$ doublet,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.6}$$

$D_\mu = (\partial_\mu - igA_\mu \tau^a - ig'YB_\mu)$ represents the $SU(2)_L \times U(1)_Y$ covariant derivative. The second and third term are the Higgs potential, $V(\phi)$,

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2. \tag{2.7}$$

The same arguments as before follow for the $\mu^2 > 0$ and $\mu^2 < 0$ cases, where again it can be shown that $\phi^0 = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}$.

2.1.2 Spontaneous Symmetry Breaking

By making a choice of non-zero v.e.v, the gauge symmetry $SU(2)_L \times U(1)_Y$ is spontaneously broken to $U(1)_{EM}$.

By making a choice of a radial direction, which here is taken as ϕ_3 (i.e setting $\phi_{1,2,4} = 0$), one can write [7]

$$\phi^0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{2.8}$$

with $h(x)$ corresponding to the physical Higgs field, appearing as the perturbation about the minimum v. Making this choice is known as the unitary gauge.

2.1.2.1 Gauge Boson Masses

To see the effect of the Higgs on the mass spectrum of gauge bosons, we can expand Eq. 2.1 using 2.8,

$$(D^\mu \phi)^\dagger (D_\mu \phi) = \frac{1}{2}(\partial^\mu h)^\dagger (\partial_\mu h) + \frac{1}{4} g^2 (v+h)^2 W^+_\mu W^{-\mu} + \frac{1}{8}(v+h)^2 (g+g')^2 Z^0_\mu Z^{0\mu} \tag{2.9}$$

where $W^{\pm}_{\mu} = \frac{1}{\sqrt{2}}(A^1_{\mu} \pm A^2_{\mu})$ and $Z^0_{\mu} = \frac{1}{\sqrt{g^2+g'^2}}(gA^3_{\mu} - g'B_{\mu})$ represent the now more familiar $W^{\pm}$ and Z^0 gauge bosons. From terms two and three in Eq. 2.9, the mass terms for the gauge bosons can be read off as

$$M_{W^{\pm}} = \frac{vg}{2};\ M_{Z^0} = \frac{v}{2}\sqrt{g^2 + g'^2}. \tag{2.10}$$

There is also the term $A_{\mu} = \frac{1}{\sqrt{g^2+g'^2}}(g'A^3_{\mu} + gB_{\mu})$, which is identified with the photon with $M_{\gamma} = 0$.

We can perform the same exercise on the potential $V(\phi)$, gathering together h^2 terms,

$$V(\phi_0) = \frac{\mu^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 = ... + \left(\frac{\mu^2}{2} + \frac{6\lambda v^2}{4}\right) h^2 + ... \tag{2.11}$$

to determine the mass of the Higgs boson

$$M_h = \sqrt{2\lambda}v = \sqrt{2}\mu. \tag{2.12}$$

2.1.2.2 Fermion Masses

As we know, particle mass in the Standard Model is not constrained only to gauge bosons. Here we demonstrate how the process of spontaneous symmetry breaking allows the fermions to acquire masses. A simple term of the form

$$\mathcal{L}_{\psi} = m\bar{\psi}\psi \tag{2.13}$$

is not gauge invariant, so we have to think a little more carefully. By coupling fermions to the Higgs boson, Using the example of the up quark, one can write [8]

$$\mathcal{L}_u = -\lambda_u \bar{Q}_L \Phi^c u_R + h.c. \tag{2.14}$$

We have introduced the Yukawa coupling λ_u, and constructed the field (for the first generation of quarks)

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}. \tag{2.15}$$

This yields a mass for the u quark of

$$m_u = \frac{v\lambda_u}{\sqrt{2}}. \tag{2.16}$$

Moving on to leptons, considering electrons for example, one can write down a term

$$\mathcal{L}_e = y_e \bar{L}_L \psi e_R + h.c. \tag{2.17}$$

Which gives rise to a mass term

$$m_e = \frac{y_e v}{\sqrt{2}}. \tag{2.18}$$

2.1.3 Goldstone's Theorem

Goldstone's theorem states that there is a new massless scalar field for each generator of a spontaneously broken continuous, global symmetry. In the case of the Higgs mechanism however, recall that the spontaneously broken $SU(2)_L \times U(1)_Y$ is indeed a gauge symmetry [9].

As we are dealing with a local gauge symmetry, there are no Goldstone bosons. We say instead that, by making a choice of gauge, such as the unitary gauge in Eq. 2.8, the degrees of freedom corresponding to the Goldstone bosons are 'eaten' by the gauge bosons, acquiring a mass.

2.2 QCD and the Strong Force

2.2.1 QCD Lagrangian

The strong sector of the SM is governed by a local non-abelian $SU(3)_C$ gauge symmetry. As such, we expect our quark fields to be invariant under spacetime dependent transformations [10]

$$\psi(x) \to e^{i\alpha^a(x)t^a}\psi(x), \tag{2.19}$$

where t^a are the generators of $SU(3)_C$, which follow the Lie algebra $[t^a, t^b] = if_{abc}t^c$, and f_{abc} the structure constants. The full Lagrangian of QCD is given by

$$\mathcal{L}_{QCD} = \sum_f \bar{\psi}_f (i\not{D}^\mu - m_f)\psi_f - \frac{1}{4}F^{a\mu\nu}F^a_{\mu\nu}, \tag{2.20}$$

where $f = u, d, s, c, b, t$ labels the quark flavour. The ψ_f represent the quark fields with mass m_f, which are acted on by the covariant derivative

$$D_\mu = (\partial_\mu + igt^c A^c). \tag{2.21}$$

The interactions of gluons, the strong gauge bosons, are dictated by the gluon field strength tensor

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - gf^{abc} A^a_\mu A^b_\nu, \tag{2.22}$$

where the label $a = 1, ..., 8(= N_C^2 - 1)$.

2.2.2 QCD Running Coupling

One of the more notable features of QCD, and key differences between an abelian and non-abelian gauge theory, is in the running of the coupling. This is made apparent by computing the β function, which describes the running of the coupling g with energy scale.

2.2.2.1 Renormalisation

It turns out that using Eq. 2.20 to calculate diagrams involving loop induced processes will introduce unwanted divergences. What one will find is that this is due to the fact that the loop momentum is being integrated up to infinity, which implies that the theory is is valid up to arbitrarily large energy scales. It turns out this is not the case, and the prescription for taming these infinities is the process of renormalisation.

In particular, the terms in the Lagrangian written down in Eq. 2.20 are known as the bare fields, whereas to proceed with meaningful calculations one requires the renormalised fields/parameters [11]

$$\psi_B = \sqrt{Z_2}\psi_R;\ A^\mu_B = \sqrt{Z_3}A^\mu_R;\ g_B = Z_g \mu^\epsilon g_R = \frac{Z_2\sqrt{Z_3}}{Z_1} g_R. \tag{2.23}$$

The Z_i are rescaling factors, which can also be parametrised as counter-terms $\delta_i = Z_i = 1$. The renormalised quantities are the quark fields ψ, the gluon fields A^μ, and the coupling g. In the term for the coupling g, μ is a mass scale, and $\epsilon = \frac{4-d}{2}$ is a dimensionful regularisation parameter, and is used to evaluate Euclidean integrals in d-dimensions, noting that as $d = 4 - 2\epsilon$, we recover the $d = 4$ case in the limit $\epsilon \to 0$.

2.2.2.2 The β Function

The β function, which defines how the coupling g varies with energy scale μ, is defined as

$$\beta(g(\mu)) = \frac{\partial g(\mu)}{\partial \log \mu}. \tag{2.24}$$

In QED, the β function is positive, taking the form $\beta_{QED} = \frac{e^3}{12\pi^2}$, and hence the evolution of the QED coupling e with energy scale is increasing. We will see that in QCD this is not the case, which leads to the theory having the property of asymptotic freedom.

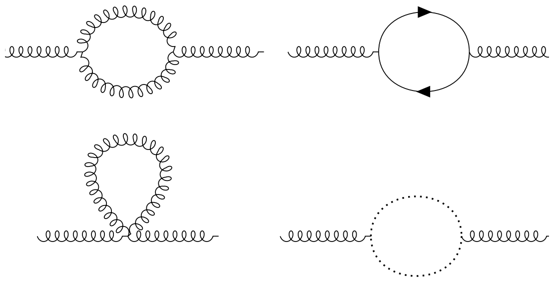
To explicitly evaluate the β function (to first order), we use

$$\beta(g) = g\ res(2\delta_1 - 2\delta_2 - \delta_3), \tag{2.25}$$

where res refers to the residue of the pole in the loop integral proportional to ϵ^{-1}. To compute the δ_i's one must compute the relevant loop diagrams.

(A) The gluon-quark vertex loop diagrams contributing to the δ_1 counter term.

(B) The quark loop diagram corresponding to δ_2.



(C) The gluon loop diagrams for δ_3.

FIGURE 2.2: The loop diagrams contributing to the three counter terms $\delta_{1,2,3}$, required to compute the β function via Eq. 2.25.

By computing these divergent diagrams in Fig. 2.2, one finds [11]

$$\beta_{QCD}(g) = -\frac{g^3}{16\pi^2}\left(\frac{11}{3}N_C - \frac{2}{3}N_f\right) = -\frac{g^3}{16\pi^2}\beta_0, \tag{2.26}$$

for a theory with N_C colours, and N_f quark flavours, which are known to be three and six respectively for the SM. Notice that for $\frac{11}{2}N_C > N_f$, which is satisfied in the SM, the β function is negative.

2.2.2.3 The QCD Coupling α_s

In terms of the QCD strong coupling constant $\alpha_s = \frac{g^2}{4\pi}$, the solution to the the β function, for running energy scale Q and fixed renormalisation scale μ, is given by [11]

$$\alpha_s(Q) = \frac{\alpha_s(\mu)}{1 + \frac{\beta_0}{2\pi}\alpha_s(\mu)\log\frac{Q}{\mu}}, \tag{2.27}$$

such that the running of coupling takes the form shown in Fig. 2.3.

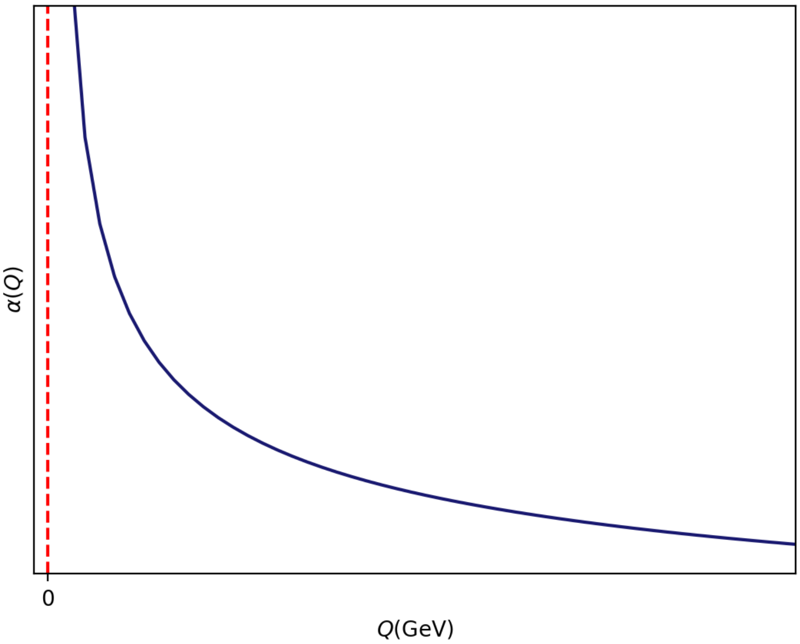


FIGURE 2.3: The running of the QCD coupling constant $\alpha(Q)$ with energy scale Q.

2.2.3 Asymptotic Freedom

As a result of the above, the strong force between quarks q_f and gluons g has the characteristic of being asymptotically free. From Fig. 2.3, we can see that the coupling strength α_s reduces as one increases the energy scale. The increase in energy scale (dimension GeV), corresponds to a decrease in length scale (GeV^{-1}), such that α_s is low at small distances, and grows at larger distances. As a consequence, quarks and gluons cannot be observed as individual objects, existing in nature only in bound states, called hadrons.

One can therefore imagine, as the separation between two quarks is increased, the strong force between them will increase, and in fact this can lead to scenarios where it becomes energetically favourable for the formation a new pair of quarks from the binding energy of the original pair. Later we will see this is an important behaviour in large scale particle physics experiments such as the LHC.

2.2.3.1 Colour Combinations

Another way of expressing this, is that in nature only colour charge neutral states can exist, which in turn comprise of combinations of coloured objects. As quarks and gluons carry colour charge (in contrast with QED, where the photon is chargeless), they both must exist in bound states together. It is here where the colour description of $SU(3)_C$ become rather convenient, as one can imagine how the two types of hadron satisfy this colour neutrality. Mesons, which contain a quark anti-quark pair ($q\bar{q}$), are combinations of a coloured quark, and an anti-quark of the corresponding anti-colour, which combine to a net colour of zero. Likewise, for baryons, which are bound states of three quarks (or three antiquarks), the combination of red, blue, and green also yields a colourless final state (as indeed would anti-red, anti-green and anti-blue).

2.3 Limitations of the SM

As a final discussion on the Standard Model, we look at some of the well established limitations, which act as motivation for the ongoing search for physics beyond the SM. We note here that not all BSM models address all of these issues, and the list provided is not representative of the issues solved by models we consider later, but instead provides a varied list of the kinds of physics yielding tensions with the SM.

2.3.1 Gravity

Perhaps the most famous example of incompleteness within the SM is its failure to incorporate a satisfactory quantum theory of gravity. The theory governing the dynamics of the gravitational force is General Relativity, governed by Einstein's field equations,

$$G_{\mu\nu} + \Lambda g_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{2.28}$$

GR has been remarkably successful, and stood up to countless tests, including the recently discovered Gravitational Waves [12]. However attempts to quantise gravity and incorporate into the SM have been so far unsuccessful, in particular it is found that gravity is perturbatively nonrenormalisable.

Such a quantum theory of gravity would imply the existence of a graviton, a massless, chargeless, spin-2 boson, however any experimental verification of their existence would come with great difficulty.

2.3.2 Dark Matter

Another well known issue in the SM is on the origin of Dark Matter. There is an astounding amount of evidence from both astronomical and cosmological phenomena postulating the existence of additional 'dark' matter in the universe. A few key examples are given below, but this list is non-exhaustive.

2.3.2.1 Galaxy Rotation Curves

A galactic rotation curve compares the velocity of a visible object in a galaxy, with its radial displacement from the galactic core. Kepler's third law, which relates the square of the orbital period T of an object with the cube of its semi-major axis of orbit, is given by

$$T^2 = \alpha r^3 \tag{2.29}$$

where $\alpha = \frac{4\pi^2}{GM}$, with G the gravitational constant and M the galactic mass. With this one can predict the expected orbital period of a star, by measuring the distance to the centre and the mass within the galaxy.

It is found, however, that there is a large discrepancy between the curve predicted, using visible light to compute the mass, and the curve obtained from measuring the velocities [13]. This inconsistency implies that there must be some additional mass in the galaxy which is not observable.

2.3.2.2 Velocity Dispersions

Astronomical observations of galaxies have yielded discrepancies between the observed mass to that predicted by theory [14]. By the Virial theorem, which states that the time averaged kinetic energy $\bar{T}$ and potential energy $\bar{U}$ for a bound system of stars in a galaxy satisfy

$$\bar{T} + \frac{1}{2}\bar{U} = 0, \tag{2.30}$$

along with measurements of the velocity distribution of stars inside a galaxy, one can calculate the expected mass distribution of a galaxy. Such measurements have been made and shown a discrepancy between measured and predicted velocities, in particular that the observed velocities would arise from a total mass greater than is measured - again implying something is missing.

2.3.2.3 Cosmic Microwave Background

A further source of evidence for the existence of Dark Matter comes from measurements of the Cosmic Microwave Background (CMB). The CMB is the earliest window through which we can peer into the early universe, first observed in 1941 by Andrew McKellar [15], and later famously discovered by Anro Penzias and Robert Wilson [16].

Measurements of the power spectrum of the CMB, the fluctuations of the temperature with angular scale coming from the observed anisotropy of the early universe, can be made. These currently indicate that approximately 5% of the universe is comprised of known baryonic matter, and around 26% Dark Matter, in close agreement with the Lambda-Cold-Dark-Matter (ΛCDM) cosmological model [17].

2.3.3 Neutrino Masses

Originally neutrinos were believed to be massless within the Standard Model, however recent experimental observations have suggested that neutrinos are in fact massive. In particular neutrino oscillations in Super-Kamiokande [18], as well as measurements from solar neutrinos [19].

If we label the three neutrino masses as $i = 1, 2, 3$, we can express the oscillation, for a neutrino flavour f, as a superposition of these mass eigenstates

$$|\nu_f\rangle = \sum_i U_{fi} \, |\nu_i\rangle \, , \tag{2.31}$$

where U_{fi} is an element of the Pontecorvo-Maki-Nakagawa-Sakata matrix, giving the probability of each respective mass eignenstate i.

As a result of the above, some extension to the SM is required to account for massive neutrinos, many candidates of which exist, such as Dirac mass, Majorana mass, or the seesaw mechanism. There are also ways in which neutrino masses can be incorporated into 2HDMs [20].

The above are a few examples out of a whole host of issues with the SM, others including matter-antimatter asymmetry, the hierarchy problem, the strong CP problem and dark energy.

Chapter 3

Two-Higgs-Doublet-Models

In this chapter we will briefly review the foundations and phenomenology of Two-Higgs-Doublet-Models (2HDMs).

3.1 A Second Higgs Doublet

The fundamental idea behind 2HDMs is strikingly simple. Instead of a single Higgs doublet, as seen in the SM, that looks like

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{3.1}$$

we add a second, such that we have complex scalar $SU(2)$ doublet fields (with hypercharge $Y_i = 1$)

$$\Phi_i = \begin{pmatrix} \phi_i^+ \\ \phi_i^0 \end{pmatrix}, \tag{3.2}$$

with $i = 1, 2$. This yields a richer phenomenology in the Higgs sector. The most general form of the potential for the new fields Φ_i can be written as [21]

$$\begin{aligned} V_{\text{2HDM}} = {} & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \text{h.c.}\right) + \frac{\lambda_1}{2}\left(\Phi_1^\dagger \Phi_1\right)^2 + \frac{\lambda_2}{2}\left(\Phi_2^\dagger \Phi_2\right)^2 \\ & + \lambda_3 \left(\Phi_1^\dagger \Phi_1\right)\left(\Phi_2^\dagger \Phi_2\right) + \lambda_4 \left(\Phi_1^\dagger \Phi_2\right)\left(\Phi_2^\dagger \Phi_1\right) + \left[\frac{\lambda_5}{2}\left(\Phi_1^\dagger \Phi_2\right)^2 + \text{h.c.}\right] \\ & + \left[\lambda_6 \left(\Phi_1^\dagger \Phi_1\right)\left(\Phi_1^\dagger \Phi_2\right) + \text{h.c.}\right] + \left[\lambda_7 \left(\Phi_2^\dagger \Phi_2\right)\left(\Phi_1^\dagger \Phi_2\right) + \text{h.c.}\right], \end{aligned} \tag{3.3}$$

where $m_{11,22}$ and $\lambda_{1,2,3,4}$ are real, and m_{12} and $\lambda_{5,6,7}$ are complex, totalling fourteen degrees of freedom in V_{2HDM}. The same procedure as for the SM is followed, finding the ground states which minimise V_{2HDM}, such that the ground states are

$$\Phi_i^0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_i \end{pmatrix}. \tag{3.4}$$

We note that Eq. 3.4 only generates mass for the d-type quarks. The u-type quarks gain mass from the isodoublet $\tilde{\Phi}_i$ [22], whose ground states look like

$$\Phi_i^{0*} = \frac{1}{\sqrt{2}} \begin{pmatrix} v_i \\ 0 \end{pmatrix}. \tag{3.5}$$

The real parameters $v_{1,2}$ can be written

$$v_1 = v \cos\beta;\ v_2 = ve^{i\xi} \sin\beta;\ \tan\beta = \frac{v_1}{v_2} \tag{3.6}$$

such that the vev v is

$$v^2 = v_1^2 + v_2^2 = (246\ \text{GeV})^2. \tag{3.7}$$

3.2 Extracting New Physical Higgs States

With two doublet fields, after expanding around the minimum of the potential, we have eight fields [21]

$$\Phi_i = \begin{pmatrix} \phi_i^+ \\ \frac{v_i + \rho_i + i\eta_i}{\sqrt{2}} \end{pmatrix},\ (i = 1, 2). \tag{3.8}$$

In analogue with the SM Higgs mechanism, three Goldstone bosons, $G^\pm$ and G^0, are ‘eaten’ by the weak bosons $W^\pm$ and Z, such that there are five physical Higgs states remaining. These are the neutral CP-even scalars h and H, the charged scalar $h^\pm$ and a neutral CP-odd pseudoscalar A.

By convention we take H to be the heavier of the two neutral scalars, and recognise that presently the Higgs boson discovered within the framework of the SM in 2012 can be identified with either h or H in 2HDMs. These two scenarios correspond to the alignments limits $\cos(\beta - \alpha) \to 0$ and $\sin(\beta - \alpha) \to 0$ respectively.

3.2.1 A Change of Basis

In order to extract physical quantities from the states Φ_i, we must choose a basis to work from. We can transform from a given basis to another by rotating the fields with some 2×2 unitary matrix U

$$\Phi'_i = \sum_j U_{ij}\Phi_j. \tag{3.9}$$

One such choice is the Higgs basis [23], which is reached by rotating the Φ_i fields

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix}. \tag{3.10}$$

After SSB it is found that the H_1 field obtains a vev of $\frac{v}{\sqrt{2}} = \sqrt{\frac{v_1^2+v_2^2}{2}}$, while H_2 has a null vev.

From Eq. 3.8, the Goldstone modes are

$$G^{\pm} = \phi_1^{\pm}\cos\beta + \phi_2^{\pm}\sin\beta;\;\; G^0 = \eta_1\cos\beta + \eta_2\sin\beta, \tag{3.11}$$

the physical charged Higgses are written

$$H^{\pm} = -\phi_1^{\pm}\sin\beta \;+ \phi_2^{\pm}\cos\beta. \tag{3.12}$$

For the neutral Higgses, in addition to $\rho_{1,2}$, we have

$$\rho_3 = -\eta_1\sin\beta + \eta_2\cos\beta. \tag{3.13}$$

To obtain the physical neutral Higgs states, we perform a rotation parametrised by α, such that

$$h = -\rho_1\sin\alpha + \rho_2\cos\alpha;\;\; H = \eta_1\cos\alpha + \eta_2\sin\alpha \tag{3.14}$$

and $A = \rho_3$, in the case where the specific model concerned is CP-conserving, such that the ρ_3 decouples from the gauge boson pairs at tree level.

3.2.2 Higgs Masses

Finally, we can write expressions for the masses of the physical Higgs spectrum arising in the theory. For the pseudoscalar A we have [23]

$$m_A^2 = \frac{m_{12}^2}{\sin\beta\cos\beta} - v^2\left(\lambda_5 + \frac{\lambda_6}{2}\cot\beta + \frac{\lambda_7}{2}\tan\beta\right), \tag{3.15}$$

for the charged Higgses $H^\pm$

$$m_{H^\pm}^2 = \frac{m_{12}^2}{\sin\beta\cos\beta} - v^2\left(\lambda_4 - \frac{\lambda_5}{2} + \frac{\lambda_6}{2}\cot\beta + \frac{\lambda_7}{2}\tan\beta\right). \tag{3.16}$$

For the CP-even states h and H, the mass matrix M

$$M^2 = \begin{pmatrix} M_{11}^2 M_{12}^2 \\ M_{21}^2 M_{22}^2 \end{pmatrix} \tag{3.17}$$

is used, where

$$\begin{aligned} M_{11}^2 &= m_A^2\sin^2\beta + v^2(\lambda_1\cos^2\beta + \lambda_5\sin^2\beta + 2\lambda_6\sin\beta\cos\beta) \\ M_{12}^2 &= M_{21} = -m_A^2\sin\beta\cos\beta + v^2((\lambda_3+\lambda_4)\sin\beta\cos\beta + \lambda_6\cos^2\beta + \lambda_7\sin^2\beta) \\ M_{22}^2 &= m_A^2\cos^2\beta + v^2(\lambda_2\sin^2\beta + \lambda_5\cos^2\beta + 2\lambda_7\sin\beta\cos\beta). \end{aligned} \tag{3.18}$$

The masses for the CP-even scalar are then given by (taking $m_H > m_h$) [24]

$$m_{h,H} = \frac{1}{\sqrt{2}}\left(M_{11}^2 + M_{22}^2 \mp \sqrt{(M_{11}^2) - M_{22}^2)^2 + 4M_{12}^4}\right)^{\frac{1}{2}}. \tag{3.19}$$

3.2.3 Quark Masses

Of course this does not explain how the Higgs gives mass to fermions. As we will see there are several different iterations of 2HDM, which each treat the generation of quark masses differently. In further study we limit ourselves to the so-called type II case.

In the type II 2HDM (or 2HDM-II), the u type quarks (i.e. those with charge $q = +2/3$) couple to one Higgs doublet, and the d type quarks ($q = -1/3$) couple to the other. Conventionally these are taken to be Φ_2/Φ_1 for the u and d types respectively [25].

3.3 FCNCs in 2HDMs

Flavour changing neutral currents (FCNCs), are fermion interactions from one flavour to another, without any change in electric charge - for example an interaction a decay process that converts a μ^- to an e^-. In the SM, tree level FCNCs are not allowed, and beyond tree level FCNCs are suppressed by the Glashow-Iliopoulos-Maiani (GIM) mechanism [26]. In fact the SM is hugely successful in its treatment of FCNCs - and so any new physics should not alter this behaviour.

When introducing a second Higgs doublet in a 2HDM, we introduce new potential sources of FCNCs, which do not agree with current observations, and so a mechanism to suppress FCNCs is required. In the 2HDM framework described above in Eq. 3.8, we invoke a $\mathbb{Z}_2$ symmetry

$$\Phi_1 \rightarrow \Phi_1, \;\; \Phi_2 \rightarrow -\Phi_2. \tag{3.20}$$

With this, the potential becomes

$$\begin{aligned} V_{\text{2HDM}} = {} & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 + \frac{\lambda_1}{2}\left(\Phi_1^\dagger \Phi_1\right)^2 + \frac{\lambda_2}{2}\left(\Phi_2^\dagger \Phi_2\right)^2 + \lambda_3 \left(\Phi_1^\dagger \Phi_1\right)\left(\Phi_2^\dagger \Phi_2\right) \\ & + \lambda_4 \left(\Phi_1^\dagger \Phi_2\right)\left(\Phi_2^\dagger \Phi_1\right) + \left[\frac{\lambda_5}{2}\left(\Phi_1^\dagger \Phi_2\right)^2 + \text{h.c.}\right] - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \text{h.c.}\right), \end{aligned} \tag{3.21}$$

where we have rearranged the terms, demonstrating that the imposed $\mathbb{Z}_2$ symmetry is softly broken by m_{12}^2. There are in fact four types of 2HDM which can account for the softly broken $\mathbb{Z}_2$ - type I, type II (seen above), type X and type Y.

3.4 Phenomenology of 2HDMs

We have laid the foundations of the theoretical framework of 2HDMs, and so we now discuss the phenomenology of 2HDMs, including a discussion of LHC searches for evidence of 2HDM particles.

3.4.1 Theoretical Model Constraints

Before investigating the experimental searches, and hence regions of explored parameter space, there are important constraints on certain parameters.

3.4.1.1 Vacuum Stability

In order for the vacuum to be stable we require that the potential $V_{2\mathrm{HDM}}$ is positive, which implies the following constraints on the λ_i (i=1,..,7) [23]

$$\lambda_{1,2} > 0; \;\; \lambda_3 > -\sqrt{\lambda_1\lambda_2}. \tag{3.22}$$

For the case where either λ_6 or λ_7 are zero, we also require

$$\lambda_3 + \lambda_4 - \lambda_5 > -\sqrt{\lambda_1\lambda_2}, \tag{3.23}$$

and where both $\lambda_6 = \lambda_7 = 0$, λ_5 is replaced with $|\lambda_5|$ in the above expression.

3.4.1.2 Tree-Level Unitarity

A further constraint comes from the scattering matrices, which are required to be unitary. This implies that the eigenvalues, which are given by

$$\begin{aligned}
a_\pm &= 3\left(\lambda_1 + \lambda_2 + 2\lambda_3\right) \pm \sqrt{9(\lambda_1 - \lambda_2)^2 + \left(4\lambda_3 + \lambda_4 + \frac{1}{2}\left(\lambda_5 + \lambda_6\right)\right)^2} \\
b_\pm &= \lambda_1 + \lambda_2 + 2\lambda_3 \pm \sqrt{(\lambda_1 - \lambda_2)^2 - \frac{1}{4}\left(2\lambda_4 - \lambda_5 - \lambda_6\right)^2} \\
c_\pm &= \lambda_1 + \lambda_2 + 2\lambda_3 \pm \sqrt{(\lambda_1 - \lambda_2)^2 + \frac{1}{4}\left(\lambda_5 - \lambda_6\right)^2} \\
e_1 &= 2\lambda_3 - \lambda_4 - \frac{\lambda_5}{2} + \frac{5\lambda_6}{2} \\
e_2 &= 2\lambda_3 + \lambda_4 - \frac{\lambda_5}{2} + \frac{\lambda_6}{2} \\
f_+ &= 2\lambda_3 - \lambda_4 + \frac{5\lambda_5}{2} - \frac{\lambda_6}{2} \\
f_- &= 2\lambda_3 + \lambda_4 + \frac{\lambda_5}{2} - \frac{\lambda_6}{2} \\
f_1 &= f_2 = 2\lambda_3 + \frac{\lambda_5}{2} + \frac{\lambda_6}{2} \\
p_1 &= 2(\lambda_3 + \lambda_4) - \frac{\lambda_5}{2} - \frac{\lambda_6}{2},
\end{aligned} \tag{3.24}$$

obey the conditions [27]

$$|a_\pm|,\; |b_\pm|,\; |c_\pm|,\; |e_1|,\; |e_2|,\; |f_+|,\; |f_-|,\; |f_1|,\; |p_1| \;\leq 8\pi. \tag{3.25}$$

3.4.1.3 Oblique Parameters

The oblique parameters S, T and U [28] (along with their higher order cousins V, W and X [29]) are a powerful tool in measuring deviations from the SM, via precise electroweak interaction measurements.

The parameters themselves encode loop corrections to electroweak gauge interactions, and are computed via a litany of electroweak diagrams. By convention, in the SM $S = T = U = 0$, such that a non-zero value for any of the observables give a large hint on the existence of BSM physics. In particular, for 2HDMs, there are corrections contributing to S, T and U arising explicitly from the extra Higgses, in particular in their loop contributions and couplings to the SM Higgs boson [30].

3.4.2 2HDMs at the LHC

Clearly, as the particle content in 2HDMs are indeed richer than predicted in the SM, perhaps the most obvious route to discovering evidence of 2HDMs is via the detection of one of the unobserved Higgs states, to follow the 125 GeV Higgs boson in 2012.

3.4.3 (Pseudo)Scalar Higgs Decays

3.4.3.1 $H \to hh$ Decays

The process of interest throughout this thesis is entirely between the CP-even, neutral scalars, h and H. Where it is kinematically possible, i.e. for $m_H > 2m_h$, we can observe decays from the heavier Higgs H to a pair of lighter Higgses hh. Assuming for a moment we identify h with the observed Higgs boson, so $m_h = 125$ GeV, then the decay channels are well known [31]. In particular, the dominant decay channel is $h \to b\bar{b}$ (with a branching ratio of around 57%).

With the above, we can therefore observe decay chains of the form $H \to hh \to bb\bar{b}\bar{b}$, leading to high multiplicity b-jet final states at the LHC. Such final states are part of active searches, for example at the ATLAS detector [32], along with mixed final states from $H \to hh$ decays, such as $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$ and $b\bar{b}W^+W^-$. We limit our focus to the $bb\bar{b}\bar{b}$ final states in part II.

3.4.3.2 $h/H \to AA$ Decays

In analogue with the above, there are also experimental searches involving the pseudoscalar A. For example, where kinematically permitted, $h/H \to AA \to \gamma\gamma\gamma\gamma$ [33] or $h/H \to AA \to \tau^+\tau^-\tau^+\tau^-$ [34].

3.4.4 Charged Higgs Decays

As mentioned, the research in this thesis is primarily focused on the decays from the scalar particles in the extended 2HDM Higgs family, however other searches are of course possible, such as via the charged Higgses $H^{\pm}$.

There are a range of possible avenues for charged Higgs searches, depending on the mass $m_{H^{\pm}}$. In the lower mass region, the dominant production mechanism for the charged Higgs comes from top decays $t \to H^{+}b$, with $H^{+} \to \tau^{+}\nu$.

At higher masses the reverse occurs, with $H^{+} \to tb$, as well couplings with other Higgses via $W^{\pm}$ bosons $H^{\pm} \to h/AW^{\pm}$. These only cover a small portion of possibilites, For comprehensive review of charged Higgs searches at the LHC, we refer the reader to [35].

Chapter 4

Jet Physics

We now move on to a review of the physics of jets, which are of paramount importance in mapping between messy final states in high energy experiments, and details of the underlying hard interactions. The latter are of course are required for discovering evidence of BSM physics.

4.1 Jet Formation - Revisiting QCD

We will later see that the a jet is not a fundamental object in a QFT, and instead requires a jet definition to be observed and studied. The foundational basis for such a requirement arises due to nature of the strong force in a high energy particle detector environment, in particular the feature of colour confinement in QCD.

In particular, jets arise from quarks and gluons, collectively referred to as partons in this context. If, in some event in a particle collider, a hard interaction results in partons in the final state, these undergo several processes between the hard interactions and the detector, which we outline below.

4.1.1 Parton Showering

The first step in the genesis of jets is so called parton showering, which is a series of small angle splittings from a parton, as shown in Fig. 4.1.

Consider the production of a quark antiquark pair from an electron positron collider,

$$e^+e^- \to q\bar{q}. \tag{4.1}$$

As the quark and antiquark traverse away from one another over sufficiently large enough distances, we have seen that the strong force between them increases in strength. This can be further expressed in the QCD potential for a $q\bar{q}$ separated by distance r, which contains a term linear in r

$$V_{\mathrm{QCD}} \sim -\frac{4}{3}\frac{\alpha_S}{r} + \kappa r. \tag{4.2}$$

The gluon flux tube connecting the $q\bar{q}$ pair is elongated as they move farther and farther apart, and when the length exceeds some critical value the tube "breaks", whereby it is energetically favourable for the formation of an addition $q\bar{q}$ pair, forming a set of two $q\bar{q}$ pairs [36].

This is however not the full story, as partons can also emit gluons, which themselves can undergo transitions into a pair of gluons or a quark antiquark pair, i.e. the vertices $gq\bar{q}$ and ggg are permitted. As a simple picture, consider the probability of a parton (labelled X) to emit a quark or gluon, given by

$$\mathcal{P}(X \to Xg) \sim \alpha_s \int \frac{dE}{E}\frac{d\theta}{\theta}, \tag{4.3}$$

for an angle θ, and at outgoing energy E. We can see this diverges at low θ, which represents the angle of emission, and so the probability of radiating a gluon at small angle will dominate over large angle emissions. A series of emissions will therefore be in collimated flows - this is the starting point for a jet.

Due to the E integral, there is also a soft singularity as $E \to 0$. The mechanism by which this is remedied is slightly more subtle, and occurs via the cancellation of diagrams corresponding to real and virtual corrections to IRC safe observables [37].

Generalising to various kinds of splittings, one can write down the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation, which encodes the behaviour of partons in hadron collisions via the parton distribution functions (PDF) $f(x,\mu)$, at some energy scale μ

$$\mu\frac{\partial}{\partial\mu} f_j(x,\mu) = \sum_j \int_x^1 \frac{dz}{z}\frac{\alpha_S}{2\pi} P_{ij}(z) f_j\left(\frac{x}{z},\mu\right). \tag{4.4}$$

The sum represents a generalisation to multiple parton splittings, and the $P_{ij}(z)$ represent the splitting functions, for a $j \to ik$ splitting with i taking a fraction z of the total momentum of j. We can write the splitting functions [10]

$$\begin{aligned}
P_{qq}(z) &\sim \frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z) \\
P_{qg}(z) &\sim z^2 + (1-z)^2 \\
P_{gq}(z) &\sim \frac{1+(1-z)^2}{z} \\
P_{gg}(z) &\sim \frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) + \delta(1-z).
\end{aligned} \tag{4.5}$$

Fig. 4.1 provides a schematic picture of what happens after a series of splittings, forming a group of partons at small angle with the incident quark.

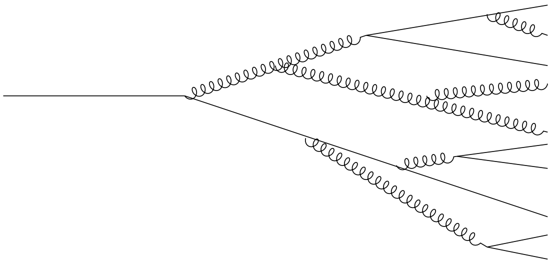


FIGURE 4.1: A demonstration of a quark undergoing multiple collinear splittings, also know as parton showering.

4.1.2 Hadronisation

As we have established, the QCD coupling α runs with energy scale Q. Considering the environment inside a detector, the primary interaction vertex where the accelerated partons collide is at high energy, such that we are in the perturbative regime for QCD where quarks gluons are (almost) free. By the time particles reach the detector however, we are at larger distance scales, and hence lower energy scales, such that quarks and gluons are strongly coupled, and perturbation theory no longer holds. This is where hadronisation (also referred to as fragmentation) occurs, whereby coloured partons form stable colourless hadrons (such as pions or kaons) which are eventually found in detectors.

A key concept for approaching hadronisation models is the local parton-hadron duality (LPHD) [39], which states that there is a similarity between global jet features computed at parton level (that is after showering, but before hadronisation) and those measured (i.e. the ones after hadronisation).

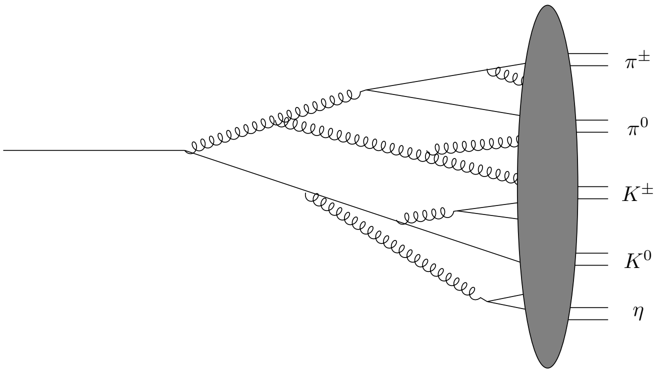


FIGURE 4.2: The showering process in Fig. 4.1, with hadronisation, whereby partons join to form colourless hadrons.

4.1.2.1 Longitudinal Tube Model

A particular, simple example of a hadronisation model is the longitudinal tube model [40], where hadrons arising from a pair of colour-connected partons are confined to a cylinder in (y, p_T). Here y is the rapidity coordinate defined by [41]

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right), \tag{4.6}$$

for a particle with energy E and momentum along the z-axis (taken to be the beam line in colliders) p_z. Defining the hadron density, $\rho(p_T)$, the energy and momentum for a jet is written

$$\begin{aligned} E &= \int_0^{\mathcal{Y}} dy\ d^2p_T\ \rho(p_T)p_T\cosh y = \lambda\sinh\mathcal{Y} \\ p &= \int_0^{\mathcal{Y}} dy\ d^2p_T\ \rho(p_T)p_T\sinh y = \lambda(\cosh\mathcal{Y}-1) \sim E-\lambda, \end{aligned} \tag{4.7}$$

where we are integrating over a tube of length $\mathcal{Y}$. We also have

$$\lambda = \int d^2p_T\ \rho(p_T^2)p_T, \tag{4.8}$$

which sets the scale for hadronisation.

4.1.2.2 Lund String Model

A more realistic model of hadronisation, used in modern fragmentation and hadronisation simulation software, is the Lund string model [42]. As the name suggests, in such a model one treats the interactions between partons as a one dimensional string.

Considering again the simplest example of a $q\bar{q}$ pair, one can construct a Hamiltonian for the string system binding together two quarks (which we label i and j respectively) as

$$\mathcal{H} = E_i + E_j + \kappa \Delta x_{ij}, \tag{4.9}$$

where Δx_{ij} is the $q\bar{q}$ separation $\Delta x_{ij} = |x_i - x_j|$, and κ refers to the string tension. One can image a mode where the quark separation fluctuates periodically between some maximum value $\Delta x'$ and zero, in a bouncing 'yo-yo' like motion. However, as we have seen, one of the driving mechanisms of hadronisation is the ability for the strings to break due to build up of potential energy in the string.

If we label the initial quark antiquark pair as $q_0\bar{q}_0$, then after a splitting we produce a new quark pair $q_1\bar{q}_1$, which are organised into bound states $q_0\bar{q}_1$ and $q_1\bar{q}_0$. The newly produced hadron h $(q_1\bar{q}_1)$, carries a fraction z of momentum modelled by the distribution

$$f(z) \sim \frac{(1-z)^a}{z} e^{-b\frac{m_h^2}{z}}, \tag{4.10}$$

where a and b are parameters tuned via experimental data.

4.2 Jet Clustering Algorithms

We have seen that the large distance behaviour of QCD leads to parton showering and fragmentation, which deposits collimated sprays of colour neutral hadrons in the detectors of large scale particle collider experiments. However, a jet is not a fundamental object of the theory, and so in order to become well defined we are required to implement some kind of jet definition.

In order to define jets, we use jet clustering algorithms, which act a map from the messy detector environment containing sprays of hadrons, back to the original hard interaction that took place. Once defined, the jets can be 'tagged' based on the parton from which they originated, which can then be used to infer which kind of interaction took place.

4.2.1 Higgs Decays into $b\bar{b}$ Pairs

As an example of using jets to infer hard physics, consider the production of a single Higgs boson in the SM, which then decays via its most familiar products, a $b\bar{b}$ pair [43]. Assuming for a moment that the Higgs is produced at rest, the outgoing b-(anti)quarks will be back to back, resulting in two well resolved jets. The jets can then be b-tagged (more on tagging later), and by plotting the invariant mass of the dijet system (that is, the two b-jets combined), we will observe a peak around the mass of the Higgs boson.

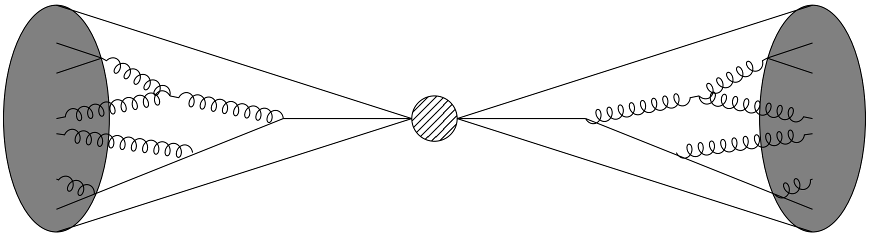


FIGURE 4.3: Schematic of a back to back dijet system at parton level (that is, we do not display the effect of hadronisation), where the blob represents the Higgs production in our example.

We can see the dynamics of a back to back jet system in Fig. 4.3. This is of course a simplistic scenario, and treating the intermediate state (the Higgs in this example) at rest will generally not be the case in a real experiment. We will come back to more complex examples later on in this section.

Assuming that we can indeed correctly identify these jets are originating from b-quarks, and so they are labelled as b-jets, we then construct the invariant mass of the dijet system to see where the $b\bar{b}$ came from.

Generating a small sample of Monte-Carlo events, we can see the kind of mass distribution one might observe for a pair of b-jets from a SM like Higgs boson in Fig. 4.4. Of course in order to define jets here, we have applied a jet algorithm, which we review in the next section.

4.2.2 Sterman-Weinberg Jets

The first jet algorithm was developed in 1977 by Sterman and Weinberg [44], in the context of e^+e^- collisions into hadrons at the SLAC experiment. It was found that at high energies, $e^+e^- \to jj$ processes dominate, which were noted to follow an angular distribution comparable to that of a pair of charged spin-$\frac{1}{2}$ point particles, of $1 + \cos^2\theta$.

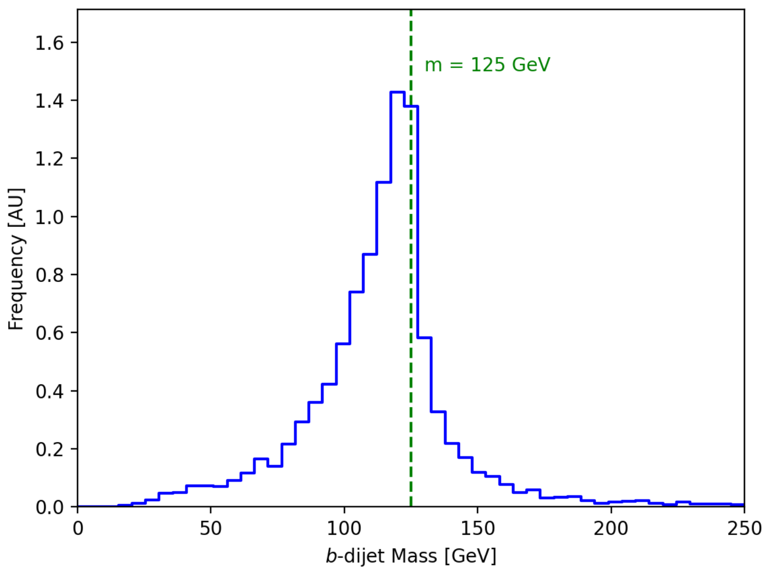


FIGURE 4.4: The $m_{b\bar{b}}$ distribution from the decay of an SM Higgs boson (m_h = 125 GeV).

In the Sterman-Weinberg algorithm, events are classified as containing two jets if a fraction of the entire energy of the event, $1 - \epsilon$, can be contained within two cones of half-angle δ. We can write this for an event with total energy E_{tot}, the final state daughters arising from the showering of two partons, i and j, with combined energy E_i and E_j respectively, are classified as two distinct jets if

$$E_i + E_j > (1 - \epsilon)E_{\text{tot}} \tag{4.11}$$

where the respective products from i and j are contained within an angle

$$\theta_i,\ \theta_j \leq \delta. \tag{4.12}$$

We can see this diagrammatically in Fig. 4.5.

Despite referring to this as the first clustering algorithm, this is perhaps a small misnomer. The Sterman-Weinberg algorithm is actually an example of a cone algorithm, rather than a clustering algorithm. The former aim to directly construct a stable cone out of the final state radiation, where no actual ‘clustering’ takes place. Alternatively, algorithms that explicitly cluster particles together are referred to as sequential recombination algorithms.

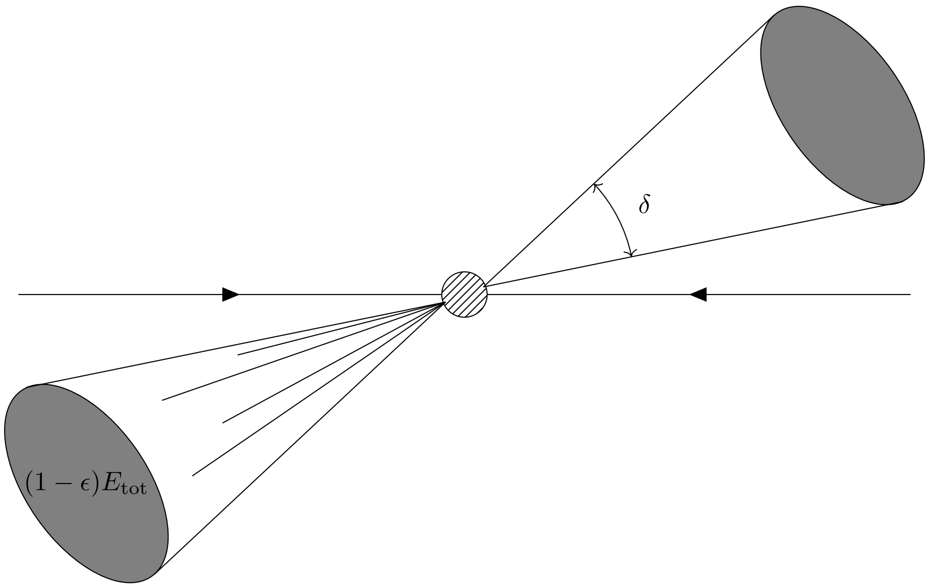


FIGURE 4.5: Demonstration of Sterman-Weinberg jets.

4.2.3 JADE Algorithm

The first such example of a sequential recombination algorithm is the JADE algorithm [45]. As the name suggests, such algorithms work by iteratively combining the final state particles. An intuitive way of thinking about this is that we are reversing the showering process, back to the parton from which the jet originated from.

To implement such an algorithm, we must define a distance measure between a pair of particles i and j. For JADE, we use

$$d_{ij} = \frac{m_{ij}^2}{E_{\text{tot}}^2} \sim \frac{2E_iE_j(1-\cos^2\theta_{ij})}{E_{\text{tot}}^2}, \tag{4.13}$$

where E_{tot} represents the energy in the entire event. The algorithm then proceeds as follows

- → Calculate d_{ij} for all possible pairs ij
- → Find the pair with the minimum d_{ij}, which we refer to as i_0 and j_0
- → Define a cut off distance measure, d_0
- → If $d_{i_0j_0} < d_0$, particles i_0 and j_0 are combined into a single object
- → Repeat the above steps until $d_{i_0j_0} > d_0$, at which point all particles are declared as jets.

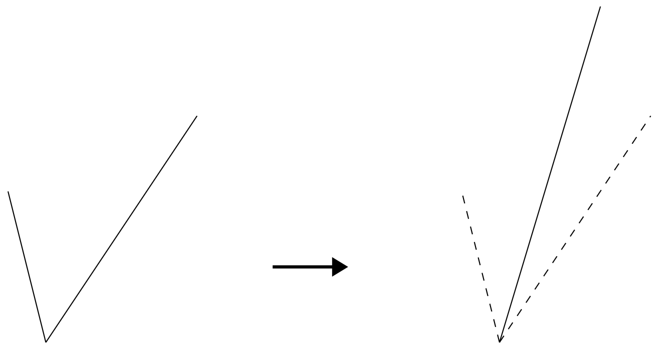


FIGURE 4.6: Visualisation of two particles i and j being combined in a sequential recombination algorithm.

We note that the distance measure vanishes both for collinear pairs, i.e. $\theta_{ij} \to 0$, and also for soft particles $E_{i/j} \to 0$. In practice, as the algorithm works through an event, it will prioritise clustering regions with collinear and soft emissions, however this has the potential to cause issues, where two soft particles $(E_i, E_j \to 0)$ that are widely separated are clustered before some other pair, as the energy of both particles are accounted for in d_{ij}.

Clearly this algorithm is significantly more intensive than cone algorithms, however comes with performative advantages. The flexibility of having a single input parameter d_0, allows the JADE algorithm to behave well to multijet events, which become tricky when using cone algorithms.

4.2.4 $k_\perp$ Algorithm

Shortly after the JADE algorithm was developed, the next sequential was designed, the $k_\perp$ Algorithm [46]. This was also developed primarily in the context of $e^+e^- \to$ hadrons events, and comprised primarily of a small tweak to JADE rather than a rigorous overhaul.

As we mentioned, the behaviour of the the distance measure d_{ij} caused wide angle soft particle pairs to be clustered ahead of perhaps more sensible choices, and so a modified distance measure was introduced,

$$d_{ij} = \frac{2\min(E_i^2, E_j^2)(1 - \cos^2\theta)}{E_{\text{tot}}^2}. \tag{4.14}$$

All we've done here is replace the product $E_i E_j$ with $\min(E_i^2, E_j^2)$. The min function here means that only the energy of the softer of the particle pair i, j is accounted for, and so particularly soft particles will be clustered with nearby neighbours rather than other soft particles at wide angles. The algorithm then proceeds in the same way as JADE, clustering pairs that minimise d_{ij} until all remaining pairs exceed the cut off d_0.

4.2.5 Generalised k_T Algorithm

The generalised k_T algorithm [47] proceeds similarly to above, using a modified distance measure

$$d_{ij} = \frac{\min(p_{Ti}^n, p_{Tj}^n)\Delta R_{ij}^2}{R^2}, \tag{4.15}$$

where the quantity

$$\Delta R_{ij}^2 = (y_i - y_j)^2 - (\phi_i - \phi_j)^2 \tag{4.16}$$

represents the angular separation between particles i and j, where y and ϕ refer to the rapidity azimuthal angle for a given object, and R is the input parameter, analogous to d_0 in the JADE algorithm. To proceed with the algorithm, we also define the so-called beam distance for each single object i, defined as [1]

$$d_{Bi} = p_{Ti}^n. \tag{4.17}$$

The algorithm then proceeds as follows, where we introduce a new naming convention that each intermediate object in the set of particles being clustered is referred to as a pseudojet until it is declared as a jet,

- → Calculate d_{ij} for all possible pairs ij, and d_{Bi} for all particles i
- → The minimum of the set (d_{ij}, d_{Bi}) is taken, which we refer to as d_0
- → If d_0 is a d_{ij}, the pseudojets i and j are combined into a new object
- → Alternatively if d_0 is instead a d_{Bi}, then pseudojet i is declared a jet and removed from the sample
- → The above steps are then repeated until all objects have been combined and declared as jets.

In analogue with the JADE algorithm, the generalised k_T only has a single input parameter, this time R. We can think of R as a kind of jet size (though this is not strictly accurate - however generally a larger value of R will yield jets with a larger radius). We can explicitly see the effect of R by considering the following, for a given pair of objects

---

[1] Note here an alternative notation is possible, whereby R is absorbed into the beam distance, such that $d_{ij} = \min(p_{Ti}^n, p_{Tj}^n)\Delta R_{ij}^2$ and $d_{B,i} = p_{Ti}^n R^2$. In fact we will refer back to this notation in part II.

i and j (where we will take i as the object that minimises $\min(p_{Ti}, p_{Tj})$), such that we have

$$d_{ij} = \frac{p_{Ti}^n \Delta R_{ij}^2}{R^2} = \frac{\Delta R_{ij}^2}{R^2} d_{Bi}. \tag{4.18}$$

In order to preferentially declare i as a jet over combining it with pseudojet j, we therefore require

$$\frac{\Delta R_{ij}^2}{R^2} < 1, \tag{4.19}$$

such that R acts as a cut-off for clustering objects together. As alluded to by the name, there are multiple iterations of the generalised k_T, each of which have their own distinct benefits and drawbacks.

4.2.5.1 The Anti-k_T Algorithm

We begin with the most popular of these flavours, the anti-k_T (AK) algorithm [48]. In this form of the generalised algorithm, we set the variable $n = -2$, such that in the distance measures d_{ij} and d_{Bi} the transverse momentum appears as $1/p_T^2$ - hence the name “anti-k_T”.

$$d_{ij} = \min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right) \frac{\Delta R_{ij}^2}{R^2}, \ d_{Bi} = \frac{1}{p_{Ti}^2}. \tag{4.20}$$

Consequently, harder particles will have a suppression in their respective distance measures d_{ij} and d_{Bi}’s, and hence be preferentially clustered over softer particles. The resulting jets from anti-k_T clustering therefore tend to have a well defined shape around a hard central core of particles.

4.2.5.2 The Cambridge-Aachen Algorithm

The main other version of the generalised k_T algorithm is the Cambridge-Aachen (CA) algorithm [49]. Here we take $n = 0$, producing the modified distance measures

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2}, \ d_{Bi} = 1. \tag{4.21}$$

A stark difference between the CA and the other sequential recombination algorithms is immediately obvious, CA had no dependence on the transverse momentum of the particles being clustered, and is hence purely geometrical.

4.2.5.3 Anti-k_T vs Cambridge-Aachen Jets

We saw that the AK algorithm will preferentially cluster jets around hard particles, the CA however only cares about the proximity from one potential constituent to the next. This behaviour can lead to more erratically shaped jets, with a less well defined shape, which can be advantageous in certain situations, such as resolving fat jets.

In Fig. 4.7 we observe the same event, clustered with the two k_T algorithms, both with parameter $R = 0.4$. We observe the different behaviour of the two algorithms, most notably in the sub-leading jet. When using the AK4 algorithm, the jet catchment area is enhanced by the central hard particles, such that significantly more particles are swept in. In fact this effect is so great, the jet loses enough radiation to drop a position in the p_T ranking. In the CA4 case, we have seen already that clusters are dictated by distance, leading to a more even split of particles in the region between the two overlapping jets.

However, how do we know which of these algorithms is performing ‘correctly’? In the case where we have access to Monte-Carlo information, we can of course manually track the ancestry of each final state hadron, to check which parton it originated from. Of course this is not useful for real experiments, where we do not have access to such information, but there is a modern class of algorithm which potentially solves this issue.

4.2.6 Variable-R Jets

A concept we have not yet covered yet, is selecting an optimal value for the input parameter R. Clearly it is not the case that all jets will fit into a single cone size, and so some thought is required when selecting a particular value.

In fact the angular separation of a given jet can be approximated by the following

$$R \sim \frac{2m}{p_T}, \tag{4.22}$$

in particular note the dependence on jet p_T. For hard (high p_T) jets, one can expect tightly compact narrow cone jets, for softer jets (low p_T), the resulting jet constituents are more spread out. So a jet definition which accounts for this behaviour should be expected to cluster jets of different sizes accurately - enter variable-R [50].

As the name suggests, variable-R (VR) algorithms remove the need the settle on a single fixed cone, which might be not be suitable for multijet events involving boosted objects. In VR algorithms, the distance measure d_{ij} is modified

$$d_{ij} = \frac{\min(p_{Ti}^n, p_{Tj}^n)\Delta R_{ij}^2}{R_{\text{eff}}^2(p_{Ti})}, \tag{4.23}$$

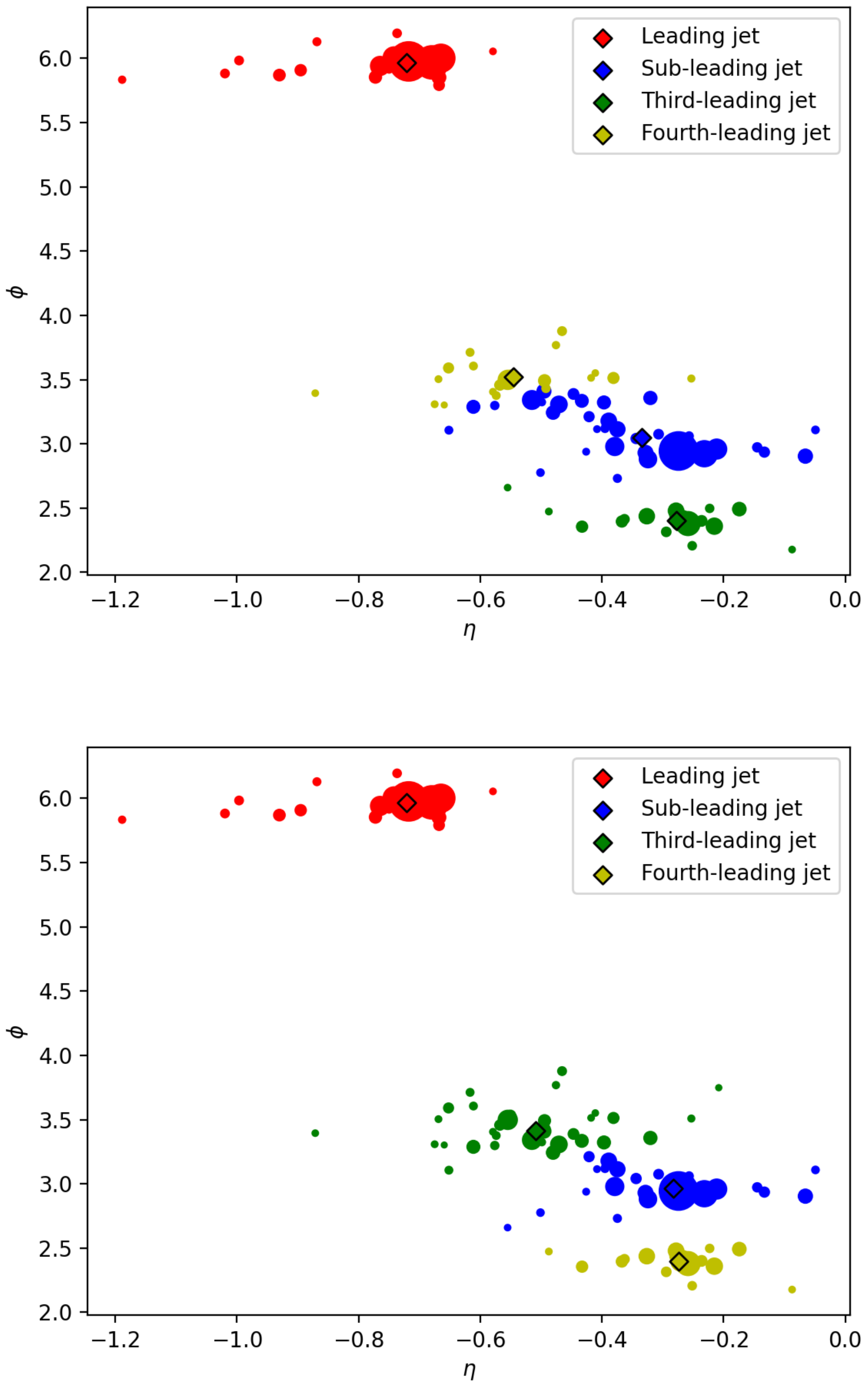


FIGURE 4.7: The same $pp \to jjjj$ event, clustered with (upper) the anti-k_T algorithm, and (lower) the Cambridge-Aachen algorithm, both using $R = 0.4$. Diamonds with black edges refer to the position of the jet, whereas the circles represent each jets constituents, with the size representing the p_T of the particle.

where R_{eff} replaces the fixed input parameter R in traditional algorithms,

$$R_{\text{eff}}(p_{Ti}) = \frac{\rho}{p_T}. \tag{4.24}$$

We immediately notice the new measure R_{eff} encodes Eq. 4.22, where ρ is a dimensionful input parameter, which is generally set to match the p_T scale of the jets in the event being clustered.

The algorithm proceeds in the same way as the generalised k_T, with d_{Bi} taking the same form as before. We also notice the appearance of n in d_{ij}, such that the VR algorithm can modified to mimic the AK or CA algorithms.

We will revisit VR algorithms in the next part of this thesis, where the performance of such algorithms are compared against traditional clustering algorithms on particular 2HDM processes involving high jet multiplicity final states.

4.2.7 Infrared and Collinear Safety

When constructing a jet algorithm, there is a one property that is of particular importance, the concept of infrared and collinear (IRC) safety. For a jet algorithm to be IRC safe, the resultant jets should remain unchanged for a sample of particles with additional soft ($E \to 0$) and collinear ($\theta \to 0$) emissions. As we have demonstrated, soft, collinear splittings are a natural ingredient of the parton showering process, and our jet algorithms indeed ought to be resistant to such emissions.

Generally, cone algorithms, which have not been discussed in detail here, have difficulties with IRC safety. Thankfully however the sequential recombination algorithms we have discussed here are indeed IRC safe. In the next part of this thesis we will review the performance of a new approach to clustering, whereby IRC safety is not a priori guaranteed, and so performative tests are required.

4.3 Jets at the LHC

So far we have focused primarily on jets from a theoretical perspective, and so moving on to our final background discussion, we consider jets as they arise in high energy experiments.

4.3.1 The CMS Detector

There are a number of concurrent experiments at the LHC, including ALICE, ATLAS, CMS and LHCb, each of which are designed with different goals in mind. We limit our discussion to the CMS (Compact Muon Solenoid) detector.

The CMS experiment is designed as a general purpose detector, aiming to address a series of BSM physics endeavours, such as dark matter, Higgs physics (including extended Higgs sectors), and extra dimensions. It is comprised of five layers, the tracker, the electric calorimeter (ECAL), the hadron calorimeter (HCAL), the magnet, and the muon detectors [51].

4.3.2 Detector Geometry

The LHC is the largest particle collider in the world, comprising of a circular tube of circumference 27km. At the point where the protons converge upon one another, the resulting hadrons (after showering and hadronisation of course) are ejected in all directions about the interaction point, sometimes referred to as the primary vertex. The shape of the detector about the beamline is thus shaped like a cylinder, with end caps on the flat sides to cover emissions from as small angles from the beamline as possible.

With this geometry in mind we can define the coordinate conventions used at the detectors. We take the beamline as the z-direction, and define the following angular coordinates [51],

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) = \frac{1}{2}\ln\left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z}\right), \tag{4.25}$$

known as the pseudorapidity. The angle θ is the angle made to the z-axis, i.e. the incoming beamline. We also express η in terms of the three-momentum and z-momentum.

The second coordinate used is the angle ϕ, which rotates about the beamline. In fact we have seen this coordinates in action before, in Fig. 4.7, where we have taken a slice along the detector, and 'unravelled' into a flat two dimensional coordinate system to image the radiation in an event.

4.3.3 Jet Tagging

As we have seen, clustering jets into suitable cones is only half of the battle, in order to extract useful physics we must of course determine which kind of parton the jet came from - this is done via jet tagging. We will limit our discussion to the tagging of jets coming from b-quarks, suitably named b-jets.

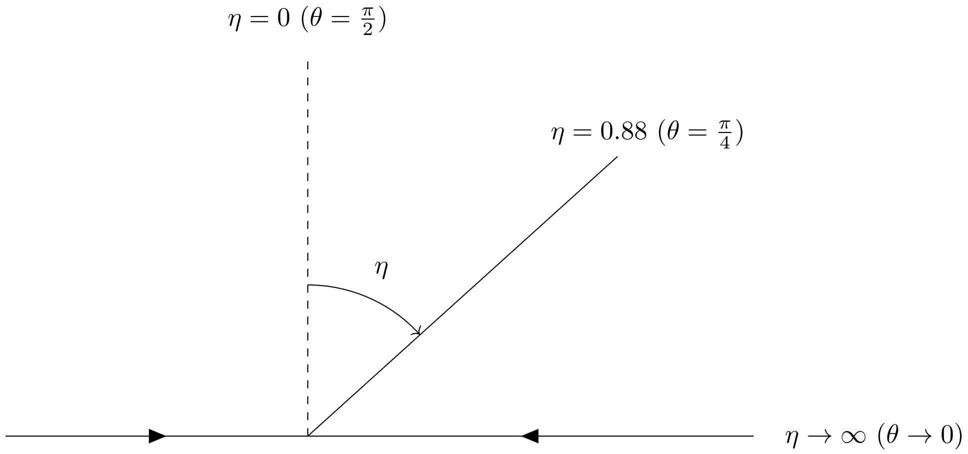


FIGURE 4.8: Demonstration of the psuedorapidity η with respect to the beamline of an event at the LHC.

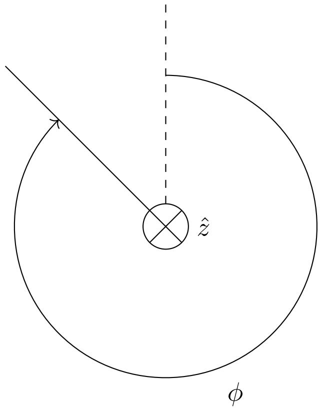


FIGURE 4.9: Demonstration of the azimuthal ϕ angle with respect to the beamline (moving into the page) of an event at the LHC.

4.3.3.1 Impact Parameter

The key feature behind traditional b-tagging is in the lifetime of hadrons containing b-quarks, B-hadrons [53]. In a decay originating from a b-quark, this lifetime is sufficient enough for the B-hadron to travel some measurable distance from the primary interaction point, which is referred to as the impact parameter d_0.

A good indicator on whether a jet came from a b-quark using this information is therefore to trace back tracks in the jet to their point of origin, if they align with some point separated from the hard interaction, the secondary vertex, the jet is more likely to be given a b-tag.

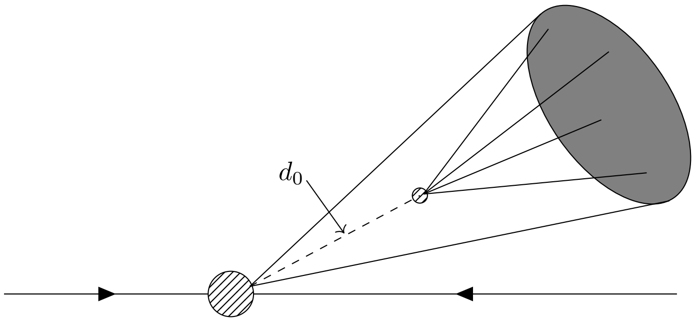


FIGURE 4.10: Diagram of the impact parameter from a B-hadron decay.

4.3.3.2 Origin of the b Lifetime

Before moving on to features within jet tagging, we take a quick detour to describe the origin of the b lifetime, which as we have seen is sufficiently long for the measurement of the impact parameter distance.

Consider the decays of a b-quark. Due to FCNC's, decays into d or s quarks are prohibited, and so the b decays into either u or c quarks, which occurs via the weak interaction. We can see this diagrammatically in Fig. 4.11 for the example of $b \to cW^-$.

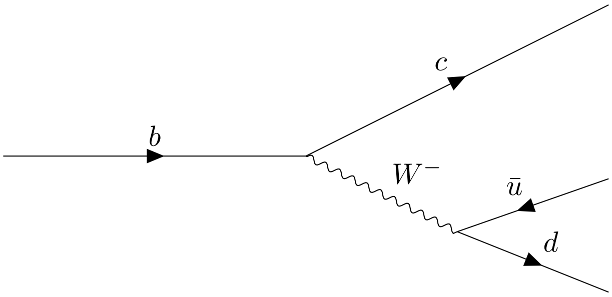


FIGURE 4.11: Example of a b weakly decaying into a c and W^-.

The coupling between pairs of different quark flavours are found in the Cabibbo-Kobayashi-Maskawa (CKM) matrix [54]

$$\begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} = \begin{bmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & (3.82 \pm 0.24) \times 10^{-3} \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & (41.0 \pm 1.4) \times 10^{-3} \\ (8.0 \pm 0.3) \times 10^{-3} & (38.8 \pm 1.1) \times 10^{-3} & 1.013 \pm 0.030 \end{bmatrix}. \tag{4.26}$$

We can see that both V_{cb} and $V_{ub} << 1$, corresponding to rare decays - which consequently gives the b its sufficiently long lifetime to make b-tagging possible.

4.3.3.3 Modern Jet Tagging

In recent times there has been significant work on jet tagging, in particular making use of machine learning methodologies, which we will discuss in the next chapter on machine learning in high energy physics.

4.3.4 Boosted Jets

A prominent feature of jets, in particular jets arising from some intermediate particle of interest, are boosted jets [55]. Consider the example of an SM Higgs boson decaying into a pair of b-jets, simple conservation of momentum tells us that for a Higgs at rest, the decaying $b\bar{b}$ pair will be back to back, and hence in our detectors be collected as two distinct objects.

However, in high energy collisions the Higgs can be created with significant momentum, which we refer to as being ‘boosted’, and so the decay products will be created with a component of their momentum in the direction of the parent Higgs, and become squeezed closer together. For sufficiently boosted Higgses, we can therefore encounter a scenario where the two resulting b-jets are so close together that clustering algorithms struggle to reconstruct the two individual jets, as there may be overlap between radiation.

Instead of trying to reconstruct the two b-jets individually, we can instead construct a single wide cone (large R) object, called a ‘fat jet’, which is comprised of two subjets (namely the jets from the decaying Higgs). We would then expect to be able to extract the Higgs mass from the invariant mass of this single fat b-jet.

4.3.5 Pile-Up and Jet Substructure

So far we have limited our discussion to simple scenarios whereby the only radiation in an event comes from the hard interaction we are interested in, for example from an SM Higgs production. In real life this is not at all the case - a given event will have emissions from a number of interactions, which is known as pile-up (PU).

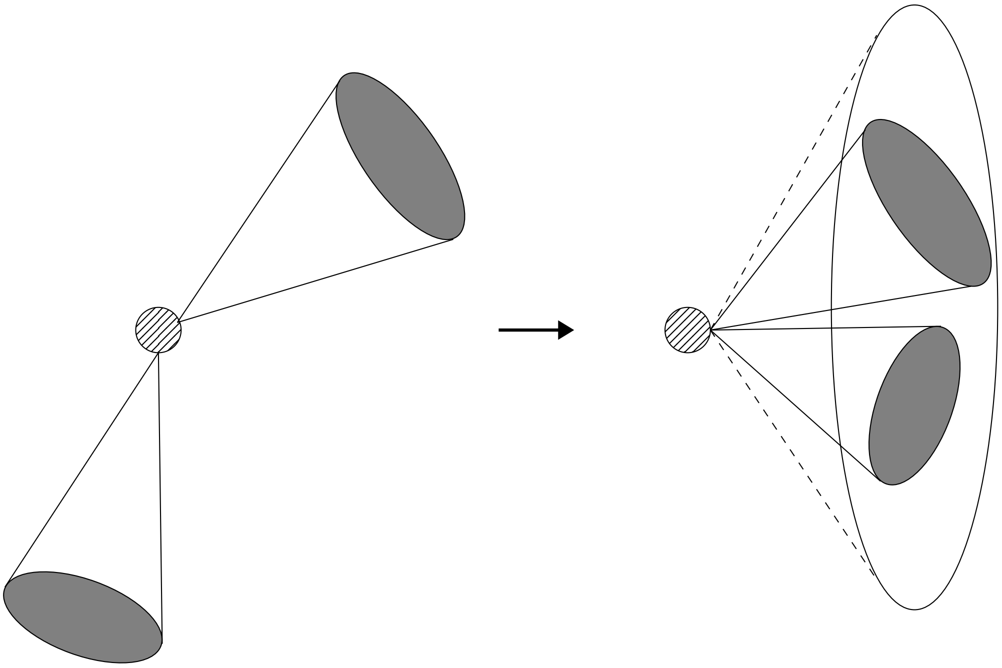


Figure 4.12: The merging of jets due into a single object due to a boosted decay.

At the LHC, instead of accelerating single proton pairs for collision, large bunches comprised of billion of protons are accelerated around the ring, up to 2556 bunches at a time, spaced out by around seven meters, which equates to roughly 25 nanoseconds. Therefore, when collisions occur, multiple protons collide at separate interactions points, up to around 50 interactions per bunch recorded at CMS.

4.3.5.1 Pile-Up Per Particle Identification

A popular method for PU removal is the Pile-Up Per Particle Identification (PUPPI) [56]. PUPPI relies on the fact we can trace back the interaction vertex from which charged tracks originate, such that charged tracks from a vertex displaced from the interaction point of interest can be removed. However charged particles only account for around 60% emissions from a *pp* interaction, so we must still remove the neutral radiation arising from PU events.

In PUPPI, a parameter α is defined, known as the shape, that estimates a probability of a given particle arising from a PU event,

$$\alpha_i = \log \sum_j \frac{P_{Tj}}{\Delta R_{ij}} \Theta(\Delta R_{ij} - R_{\min}) \Theta(R_0 - \Delta R_{ij}). \tag{4.27}$$

Here Θ is the Heaviside step function, and R_0 and R_{min} are input variables, R_0 defines a cone around particle i, such that only particles j within R_0 are included in the sum. R_{min} acts as a minimum cut-off that regulates collinear splittings from i, and only particles j separated from i by at least R_{min} are included in the sum.

One can then plot a distribution of α for all particles, to identify which are believed to have originated from PU. Generally we notice that neutral PU particles follow the same pattern as charged PU particles, such that neutral PU can be identified from the charged PU particles.

4.3.5.2 Jet Grooming

Another methodology of PU removal comes with so called jet grooming. Jet grooming involve methods of tidying up the radiation inside a jet, such that only particles coming from a particular particle of interest are kept (for example a $b\bar{b}$ pair coming from a Higgs boson). Here we briefly review a few methods of jet grooming.

4.3.5.3 Jet Trimming

Trimming is a method used on fat jets [57]. To cluster a fat jet, a large R parameter is used in sequential reconstruction algorithms, however as we have seen, there still exist the underlying subjets associated with the decay products of the boosted particle.

In the trimming procedure, we recluster the constituents of the fat jet with a smaller R, into subjets. For the sample of subjets i, we keep only those that satisfy

$$p_{Ti} = \Lambda f_{\text{cut}}, \tag{4.28}$$

where f_{cut} is a cut-off parameter, and Λ a hard momentum scale. The resulting, trimmed, jet, is the sum of the subjets i which satisfy Eq. 4.28. Radiation arising from the hard partons of interest should be confined to small comes, such that the reclustering should be able to locate the remnants from the areas of interest, while additional uninteresting radiation we expect to be uniformly distributed, and hence will not sneak past the cut.

4.3.5.4 Jet Pruning

Another method of jet grooming is pruning [58]. The basis of jet pruning is that generally the final radiation to be clustered into a jet are soft, wide angled emissions, which can have a severe impact on the resultant jet mass, and hence impact the ability to extract meaningful physics from jet variables.

Given a jet, pruning works by reclustering the constituents using the CA algorithm, and then unravelling the clusters, i.e. forming a series of splittings. For a particular splitting $k \to ij$, the following parameters are computed, firstly

$$z = \frac{\min(p_{Ti}, p_{Tj})}{p_{Tk}}. \tag{4.29}$$

This parameter acts as a measure of the softness of the emissions, and is compared with an input value z_{cut}. We also make a measure of the angular separation of i and j against a cut-off value,

$$\Delta R_{ij} > D_{\text{cut}}. \tag{4.30}$$

For a pair of splittings ij that satisfy $z < z_{\text{cut}}$ and $\Delta R_{ij} > D_{\text{cut}}$, the softer (lower p_T) particle is removed and the process continues, until a sufficiently hard split is found.

4.3.5.5 Softdrop

More recently, the pruning algorithm described above has been modified into the so-called softdrop algorithm [59]. Softdrop proceeds in the same way as pruning, with a different removal criteria. The condition for keeping a particle in softdrop is that a pair ij satisfy

$$\frac{\min(p_{Ti}, p_{Tj})}{p_{Ti} + p_{Tj}} > z_{\text{cut}} \left(\frac{\Delta R_{ij}}{R_0}\right)^{\beta}, \tag{4.31}$$

else the softer of i and j are removed. Here, both z_{cut} and β are input parameters, and hence can be tuned to optimise the softdrop performance.

4.3.5.6 N-Subjettiness

Another method of accessing jet substructure is via the parameter N-subjettiness [60]. The N-subjettiness for a jet k, τ_{Nk}, is defined by the measure

$$\tau_{Nk} = \frac{1}{d_0} \sum_i p_{Ti} \min(\Delta R_{ji}). \tag{4.32}$$

To deconstruct this, the label $j \in (1, N)$ here identifies the candidate subjets inside jet k for which we are computing τ_{Nk} for, and the i iterate over the constituent particles inside the jet k. What we have inside the min(...) function therefore, is that we are choosing the subjet j with smallest angular distance ΔR (defined in Eq. 4.16) to a given constituent i. The leading factor is

$$d_0 = \sum_i p_{Ti} R, \tag{4.33}$$

where R is the value used to cluster the original jet. It is straightforward to interpret τ_{Nk}, the output value simply scales with how likely a jet is to contain at least $N+1$ subjets, whereas for small values a jet is likely to be comprised of N or fewer subjets. For example, if for a jet k, we compute $\tau_{2k} >> 0$, then there is a significant amount of radiation aligned away from the two candidate subjets. We therefore compute for three subjets and find $\tau_{3k} \sim 0$, which implies most of the radiation inside k is aligned well with the three candidate subjets.

Chapter 5

Machine Learning in High Energy Physics

In the final of our background chapters, we will briefly review some of the foundations behind machine learning (ML), with a focus on the application of ML in high energy particle physics experiments.

5.1 An Introduction to Machine Learning

Despite being perceived as a cutting edge, modern area of research, the term machine learning was in fact coined in 1959 by Arthur Samuel, and has been in the realm of active research ever since. The fundamental idea behind machine learning is, in essence, a novel way of approaching problems with data.

Tom Mitchell provided the following quote which succinctly describes ML - “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E”.

5.1.1 A Jet Physics Example

Let’s consider an example in jet physics - the problem of *b*-tagging discussed in the previous chapter. The traditional way of tackling such a problem is to come up with a set of rules, likely derived from the underlying physics, in order to deduce a way to estimate the probability a jet arose from a *b*-quark. There are numerous issues with this kind of approach, for example such methods clearly do not generalise well to changes in the underlying rules. Of course this is unlikely to cause significant issues in physics, but in the unlikely scenario where we find that our understanding of QCD is not quite correct, the rules for our tagger would need to be altered correspondingly.

So how does using ML differ? Instead of coming up with the rules by hand, we collect a large sample of data and feed into an ML algorithm of choice (determined by the exact problem at hand), which we can 'train' to come up with its own rules for assigning *b*-tags. An important feature that is present when using many ML techniques is the ability to look under the hood at the model that has been created, such that we can see what it is the algorithm has learned, and in particular understand why particular decisions have been made.

With the ML approach, as we are not feeding in information by hand, we are less susceptible to issues arising if the 'rules' change. As mentioned this is perhaps unlikely in the context of this specific jet physics example, but in a simpler implementation, such as identifying junk email, it is perhaps more intuitive to imagine a scenario where identifiers of spam email change over time. In such cases, so long as we continue to train the model on new data, the algorithm will continue to learn the features it needs to perform adequately.

5.1.2 Classifying Machine Learning

Modern ML can be applied to a vast array of problems, due to the diverse range of techniques available, which we briefly review here.

5.1.2.1 Supervised Learning

Supervised learning is classified as such due to the fact that the dataset used to train a model contains truth labels of the variable we are trying to predict, which is referred to as the target class. In the example of jet tagging, what this means is that for each jet, we have a dataset, comprised of a series of variables we believe will be suggestive of whether the jet came from a b (in ML these variables are referred to as features), and we also have label telling us whether the jet actually has a *b*-tag or not.

5.1.2.2 Classification

The jet tagging example is a classification problem - we are categorising between a set of discrete target labels, in this case two, i.e. *b*-jet or not a *b*-jet. Classifiers with two target classes are referred to as binary classifiers, but this is not a requirement for classification problems. Cases where we are classifying between more than two different categories are sometimes referred to as multiclass classification problems.

5.1.2.3 Regression

Conversely to classification, but still within the realm of supervised learning, there are regression problems. Regression is useful if we would like to make a prediction on a continuous target variable. One can in fact frame classifications as regression problems, instead of predicting a particular class, we can output the probability that a particular instance belongs to a certain class. In fact this a common strategy with classifiers, as one can therefore experiment with different cut-off's for making a prediction for a given class.

Referring back to the jet tagger, in order to reduce the number of jets we incorrectly assign a b-tag to, we might only want to tag jets we are more confident should be assigned a tag, for example jets which our algorithm outputs a probability of at least 70%.

5.1.2.4 Unsupervised Learning

In contrast to supervised learning, unsupervised learning is used in problems where we do not have access to truth information. In fact we have seen a good example of an application for unsupervised learning in high energy physics (HEP) already, namely in jet clustering algorithms. We have a set of data (four momenta of final state tracks in an event), and we do not know a priori how these should be distributed into jets, i.e. we do not have a truth label to assign to each constituent.

In fact, as we will see in a later chapter, one can indeed apply unsupervised ML techniques to cluster jets in place of the traditional clustering algorithms.

5.1.3 The Role of Data in Machine Learning

As we have demonstrated, the key component of building an ML model is the data we feed in as an input - it is this data (combined with the hyperparameters of your model of choice) which determine how useful the results will actually be.

5.1.3.1 Splitting the Data

A key concept for building a model is the splitting of your data into two subsets - a training set and a testing (or validation) set. The reason for this is that, as we will see, models are built by minimising some function which represents the error of the models predictions. The issue with this is that it is unreliable to evaluate the model performance on the data used to train that very model.

What is done instead, is that a subset of the data (both the series of features x_i and the target class y) are removed from training, such that when we come to evaluate how well a model is performing, the test is done on a set of data that it has not seen before.

5.2 Machine Learning Models

In this section we will review a select few popular ML models - focussing on those useful to the application of jet physics.

5.2.1 Logistic Regression

The most basic ML models are linear models. These are simple in their implementation, but also somewhat limited. As the name suggests, these models are of course only useful on datasets where there are linear relationships between the features and target variable. While we do not explicitly use logistic regression in physics research, the general framework laid out is indicative of other classification methods.

For linear classification problems, we use logistic regression [61] . As the name suggests, despite being a classifier the logistic regression acts as a regressor of sorts, in that it predicts probabilities for the each output class. The output is the logistic of the result, given by [62]

$$\hat{p} = h_\theta(\vec{x}) = \sigma(\vec{x}^T \vec{\theta}), \tag{5.1}$$

where σ is the sigmoid function, $\vec{x}$ is a vector of the features in a given instance of the data set (i.e. the variables one is using to train the model), and $\vec{\theta}$ is a vector of the tuned parameters of the model, which rescale each feature to give the output prediction.

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{5.2}$$

Observing the form of $\sigma(x)$, we can see the symmetry about $x = 0$, so that for the output of a given logistic regressor, we can see that a convenient way to classify between two classes is

$$\hat{y} = \begin{cases} 0, & \hat{p} < 0.5 \\ 1, & \hat{p} \geq 0.5. \end{cases} \tag{5.3}$$

When comparing with the formula for the sigmoid function, we can see that this decision boundary is equivalent to

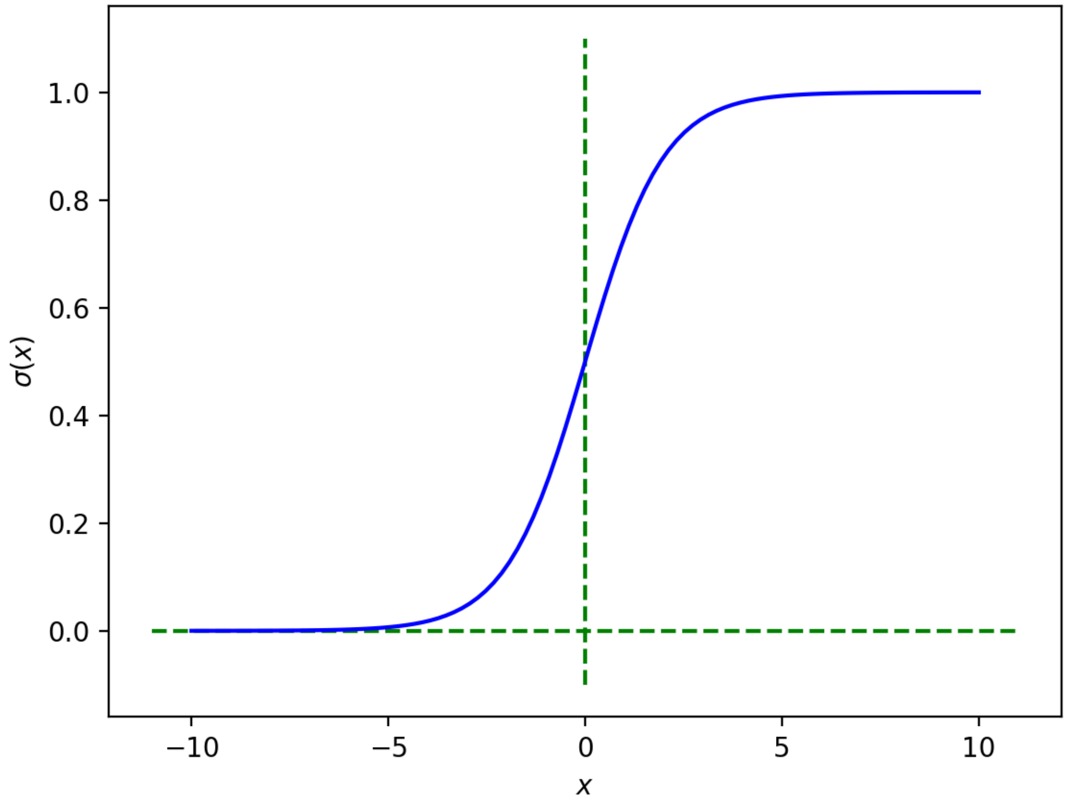


FIGURE 5.1: The sigmoid function $\sigma(x)$.

$$\hat{y} = \begin{cases} 0, & \vec{x}^T\vec{\theta} < 0 \\ 1, & \vec{x}^T\vec{\theta} \geq 0. \end{cases} \tag{5.4}$$

The goal of training a logistic regression algorithm is therefore to output large values of $\hat{p}$ for instances with $y = 1$, and low values of $\hat{p}$ for instances labelled with $y = 0$, the behavior of which is encoded in the cost function

$$c(\theta) = \begin{cases} -\log(\hat{p}) \text{ if } y = 1 \\ -\log(1-\hat{p}) \text{ if } y = 0. \end{cases} \tag{5.5}$$

Looking at the behaviour of the cost function, we can see that $-\log(x)$ is large as $x \to 0$, such that an instance with truth value $y = 0$ has a large cost associated with being incorrectly predicted $\hat{y} = 1$, and likewise for incorrect predictions the other way around. Therefore in order to find an optimal model, we minimise the following overall cost function

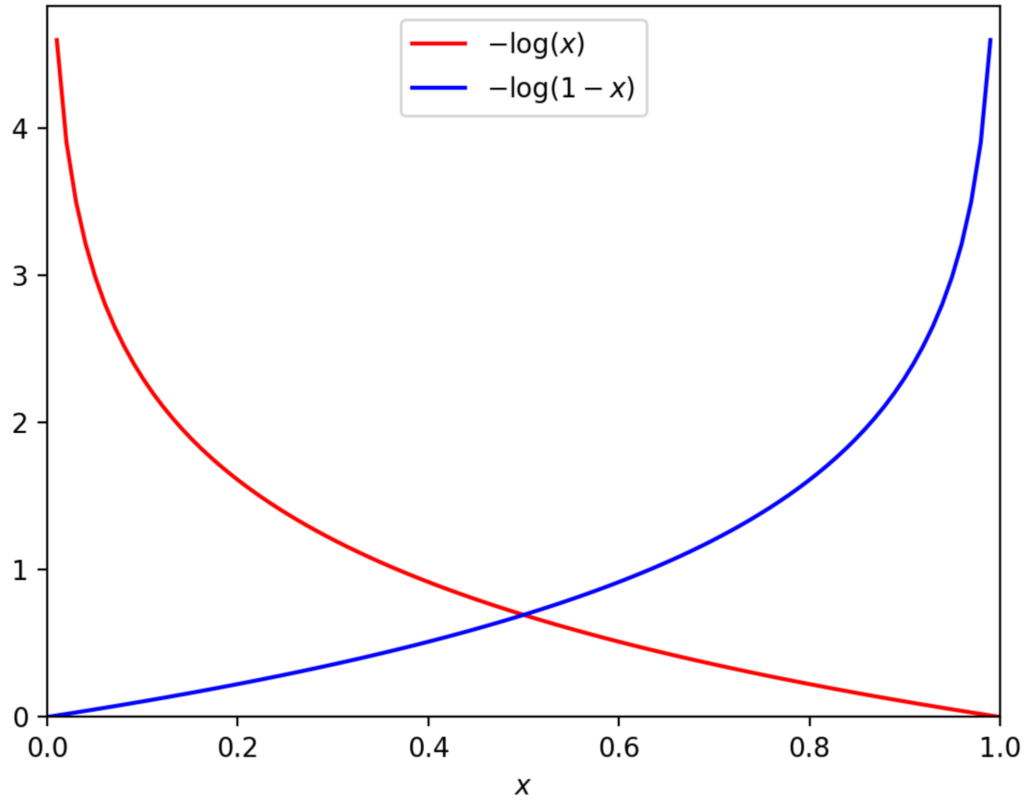


Figure 5.2: The cost function for logistic regression, as defined in Eq. 5.5.

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left[y^{(i)}\log(\hat{p}^{(i)}) + (1-y^{(i)})\log(1-\hat{p}^{(i)})\right], \tag{5.6}$$

i.e. the mean average loss over all instances in the training data set.

Unlike in the case of linear regression (which is not covered here), there does not exist a closed form solution for the cost function, and so minimisation algorithms are used, such as gradient descent, which is beyond the scope of this chapter.

5.2.2 Deep Learning

The final family of ML models we will discuss here are deep learning (DL) models, which are the most cutting edge techniques at the forefront of artificial intelligence (AI) research. DL can be applied to all sorts of problems, a notable example being the AlphaZero [63] algorithm which can outperform the top human players in both chess and go, and we will see later the potential for deep learning when applied to particle physics.

5.2.2.1 Neurons

The basic structure of deep learning neural networks (NNs) are loosely based on biological neurons. Neurons receive and send electrical impulses, and are arranged in vast arrays to complete complicated tasks despite their rather simple basic function.

A trivial example of neuron system is for the logic computation $\alpha = \beta$,

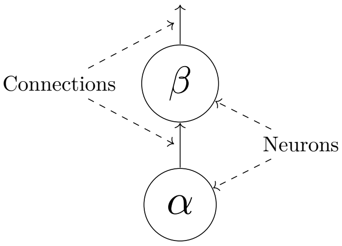


FIGURE 5.3: Demonstration of neurons for the trivial logic computation $\alpha = \beta$, demonstrated by the fact that whenever α fires, β will also fire.

5.2.2.2 Perceptron

Using the a series of these neurons, one can construct a network, where a series of inputs x_i $(i = 1, .., n)$ are fed into a series of n neurons, which can be connected to further neurons before reaching an output z.

Each of the neurons in a perceptron will come with a weight w_i, such that output is

$$z = \sum_{i}^{n} w_i x_i = \vec{x}^T \vec{w}, \tag{5.7}$$

before outputting however, an activation function is applied to the output in the so-called threshold logic unit (TLU), such as the Heaviside step function $\mathcal{H}$, such that the output is

$$h_w(\vec{x}) = \mathcal{H}(z). \tag{5.8}$$

The structure in Fig. 5.3, if we take α as an input layer and β as a TLU, would be an example of a single-layer perceptron (SLP), however these can be generalised to multi-layer perceptrons (MLP), which can have variable depths.

5.2.2.3 Multi-Layer Perceptron

MLPs are perhaps the most common example of a NN, consisting of an input and output layer - as with the SLP - as well as a number of 'hidden' layers. The multi-layered structure allows the MLP to capture highly complex patterns in data [64].

MLPs are 'fully connected', which means that each neuron in a given layer is connected to all neurons in the previous and next layer.

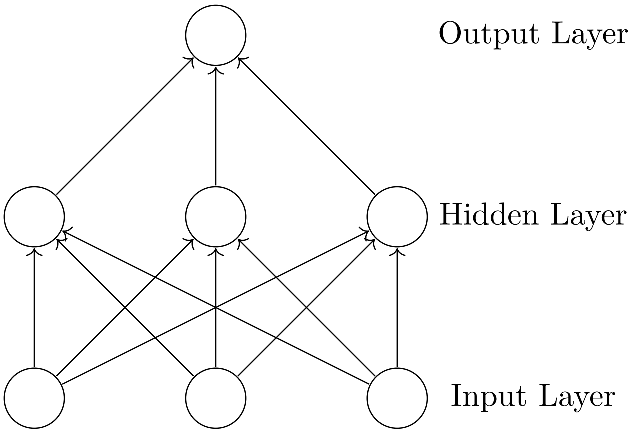


FIGURE 5.4: Visualisation of simple MLP with three features, comprising of a single hidden layer.

Another difference in MLPs is in the activation function used. In order to be able to model data that is not linearly separable, nonlinear activation functions are used, of which there are several choices - including the hyperbolic tan, sigmoid and rectifier linear unit (ReLU).

In order to to train an MLP NN, a common algorithm used is backpropagation, where the weights w_{ij} for a given feature i and layer j are sequentially updated in order to fit the data and optimise some cost function. The mathematical details of backpropagation are beyond the scope of this chapter.

When training a NN, we can actually see the model learning by looking at the evolution of the loss function over time - this is possible as NNs are trained over 'epochs'. An epoch is a training iteration, which includes a full sweep of the training data set. At each epoch of training, the model will make its set of predictions on the validation data set and compute a score - generally the loss value - so that over a series of epochs we can see the loss decreasing over time until our model reaches an optimal performance.

We can see in Fig.5.6 the evolution of the accuracy and loss, on both training and validation datasets. In fact by plotting for the result on both of these datasets we unveil another perk of tracking the scores by epoch - namely that we can compare the scores

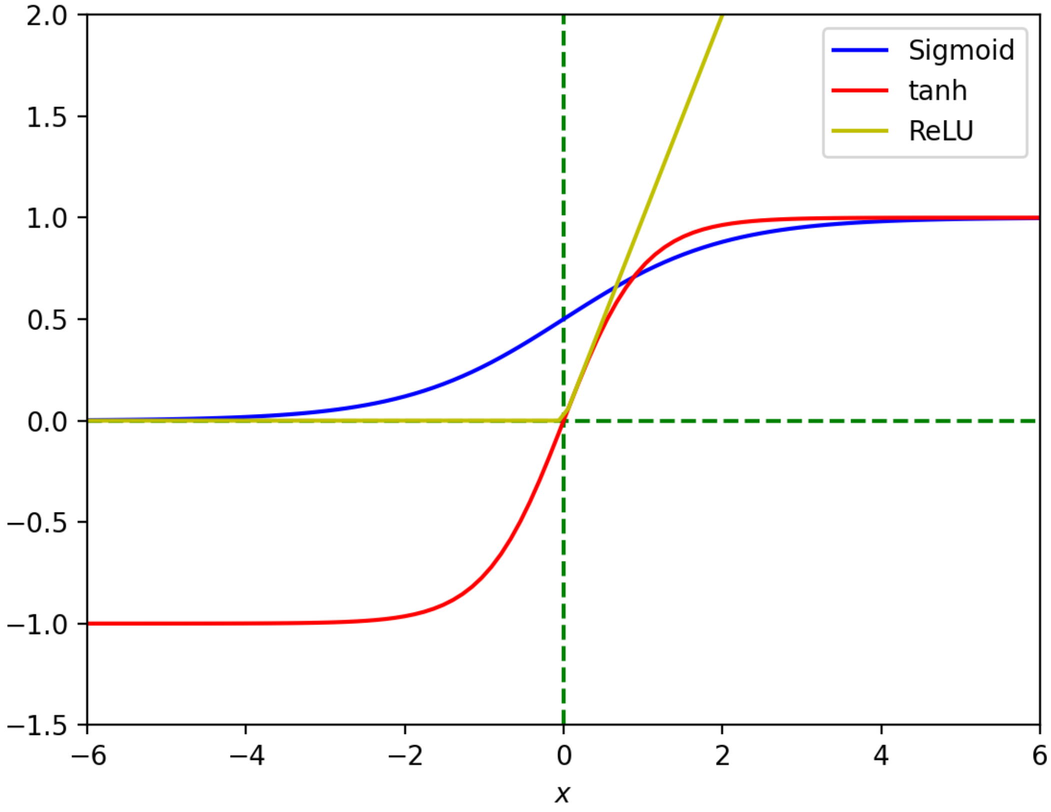


FIGURE 5.5: Three activation functions, the sigmoid, hyperbolic tan tanh, and ReLU.

on the two sets to check the model isn't being overfit to the training data (more on this later).

5.2.2.4 Convolutional Neural Networks

The final deep learning models we will discuss in this section are convolutional neural networks (CNNs) [65]. A CNN is a specialised flavour of NN which can take images (or rather matrices) as their inputs rather than one dimensional vectors - and hence are particularly useful for image recognition problems, such as the well known example of classifying handwritten digits.

Images can be very simply used as inputs by mapping each pixel into a numerical scale representing colour (in fact one can even represent RGB colours as a (3×1) vector at each pixel). Traditionally, to use a standard NN for such tasks, one would have to perform a transformation on the input data to flatten the pixels, i.e. compress a (3×3) matrix into a (9×1) vector. It is found however that the performance in these cases for more complex images is extremely limited.

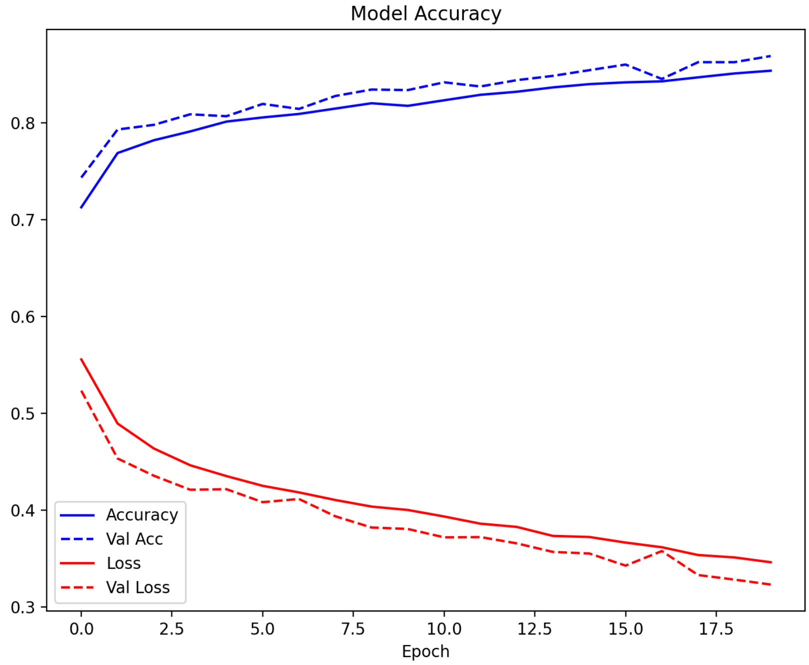


FIGURE 5.6: An example of the evolution of both model accuracy (i.e. proportion of correctly classified instances in the test set) and loss over a number of epochs.

With this improvement however comes potential issues with performance - for high resolution images the amount of data will be exceedingly large - a 4K image is comprised of ($3840 \times 2160 \simeq 8$ million) pixels. Thankfully part of the CNN structure is a so-called convolution layer, which applies a filter to subgroups of connected pixels in the image and represents them as a single value.

$$\overset{\text{Image}}{\begin{array}{|c|c|c|}\hline 1 & 3 & 4 \\ \hline 2 & 2 & 1 \\ \hline 3 & 5 & 1 \\ \hline \end{array}} \times \overset{\text{Kernel}}{\begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline \end{array}} \longrightarrow \boxed{11}$$

FIGURE 5.7: A (3×3) subset of an image being reduced to a single number by the convolution kernel layer.

This action has two benefits, firstly we are reducing the amount of data being fed into the model training, which will lead to an increase in computational performance, but also secondly this acts as a way of extracting high level features from the image. The precise features learned from the convolution layer will vary depending on the exact kernel used, which is also a positive as one can tune exactly which kernel to use depending on what we want our model to learn from the images.

Similarly to convolutions, there are also pooling layers in CNNs, which also act as dimensionality reductions, but in a slightly more elementary way. Pooling layers will also iterate sub-matrices across the image, but instead of applying a complex kernel, they will simply extract either the average or maximum value of the sub matrix, named average pooling and max pooling respectively.

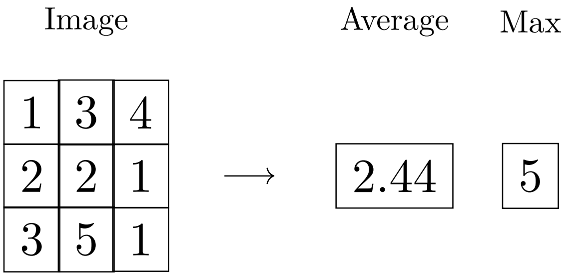


FIGURE 5.8: A (3×3) subset of an image being reduced to a single number by a pooling layer, common choices for which are either taking the max or the mean average of the pixel values of the input matrix.

A CNN architecture can be comprised of multiple layers of various different transformations on the input images, including pooling and convolutions, and one can even create more complex multi-stream networks which take multiple input images.

5.2.3 Unsupervised Models

So far we have primarily focused on supervised problems, however the applications of ML in physics are not limited to problems with truth information. We will introduce a few unsupervised ML techniques in this section.

5.2.3.1 *k*-Means Clustering

Perhaps the simplest, and most intuitive, unsupervised learning method is k-means clustering [66]. The set of inputs will be a series of instances $i = 1, ..., n$ with a value for each feature $j = 1, ..., m$, i.e.

$$(\vec{x}_1,\ \vec{x}_2, ...,\ \vec{x}_n), \tag{5.9}$$

where each $\vec{x}_i$ is an m dimensional vector. We note this is the same input for a supervised problem, except of course of the truth labels. Given this set of inputs, and an input parameter k, k-means works by computing k centroids in the data, and clustering instances into groups defined by each centroid. We can therefore see that the input k defines the number of groups we want to cluster the data into - and so k-means might be limited in its application to jet physics for example, as we would need to know how many final state jets we expect before clustering.

Initially centroids $\mathcal{C}_i$ (i=1,...,k) are placed at random [1], and instances are partitioned into groups $\mathcal{G}_i$ based on these centroids. At the next time step, the locations of $\mathcal{C}_i$ are moved based on the quantity

$$\mathcal{C}_i^{(t+1)} = \frac{1}{|\mathcal{G}_i^{(t)}|} \sum_{\vec{x}_j \in \mathcal{G}_j^{(t)}} \vec{x}_j. \tag{5.10}$$

The algorithm simply runs by updating and procedurally moving the centroids, and converges when the centroids no longer move - such that a minimum is found.

The k-means is a fairly simple example of an ML clustering algorithm, and one that comes with its limitations. Firstly it requires k as an input, although it is possible to run over multiple values of k and find an optimal value, this can become computational intensive. Furthermore, k-means does not generalise well to where the groups $\mathcal{G}_i$ are different sizes.

5.2.3.2 Spectral Clustering

We move on to a more robust ML clustering method - spectral clustering [67]. Rather than taking the raw instances (in feature space) as an input as with other models, we perform a transformation on this data, and map the points into the so-called embedding space - which is generally optimised to cluster the points into groups. The data used for the clustering step is represented as a graph - a series of nodes (points), connected by edges, which can have an associated weight called an affinity which measures how likely two connected nodes are to belong to the same final group.

For a set of n input points in the feature space, the first step is to construct an $(n \times n)$ similarity matrix $\mathcal{A}$, which is a symmetric matrix where an entry $\mathcal{A}_{ij}$ represents the similarity between points i and j, defined by some custom measure. The matrix $\mathcal{A}$ is then used to construct a Laplacian matrix $\mathcal{L}$ as we will see below.

The general procedure for a given set of coordinates is to represent the points as nodes in a graph, with weighted edges denoting the probability of two joining nodes belonging to the same group $\mathcal{G}$. In order to obtain the final clusters, the graph is split into S

---

[1] In fact, if the initial placement is completely random, then k-means is not guaranteed to converge to an optimal solution. There are methods of selecting an informed inital placement which can assist in finding the optimal solution.

subgraphs $\mathcal{G}_k$ $(k = 1, ..., s)$, where the groupings aim to minimise the affinities split to create them, mathematically we are minimising the parameter N_{cut}

$$N_{\mathrm{cut}} = \frac{1}{2}\sum_k \frac{W(\mathcal{G}_k, \bar{\mathcal{G}}_k)}{\mathrm{vol}(\mathcal{G}_k)}. \tag{5.11}$$

Here the quantity $W(\mathcal{G}_k, \bar{\mathcal{G}}_k)$ represents the cumulative sum of edges one must cut in order to detach subgroup $\mathcal{G}_k$ from the remainder of the group $\bar{\mathcal{G}}_k$, i.e.

$$W(\mathcal{G}_k, \bar{\mathcal{G}}_k) = \sum_{i\in\mathcal{G}_k}\sum_{j\in\bar{\mathcal{G}}_k} a_{ij}. \tag{5.12}$$

The term in the denominator is a measure of the affinities throughout the entire group to each point inside a subgroup $\mathcal{G}_k$, given by

$$\mathrm{vol}(\mathcal{G}_k) = \sum_i \sum_{j\in\mathcal{G}_k} a_{ij}. \tag{5.13}$$

The affinity matrix $\mathcal{A}$ can then be defined as

$$\mathcal{A}_{ij} = (1 - \delta_{ij})a_{ij}, \tag{5.14}$$

which in turn can be used to construct the Laplacian L. If we define the diagonal matrix D as

$$D_{ij} = \delta_{ij}\sum_k a_{ik}, \tag{5.15}$$

one can define the Laplacian as

$$L = D^{-\frac{1}{2}}(D - \mathcal{A})D^{-\frac{1}{2}}, \tag{5.16}$$

the eigenvectors of which are used to determine the transformed coordinates of the points in the new embedding space. In addition to L, we define indicator vectors in order to label the group in which a particular point is assigned,

$$\vec{h}_{ik} = \begin{cases} \mathrm{vol}(\mathcal{G}_k)^{-\frac{1}{2}} \text{ for } i \in \mathcal{G}_k \\ 0 \text{ otherwise.} \end{cases} \tag{5.17}$$

It turns out proceeding with vectors of this form yields a Non-deterministic Polynomial (NP-hard) problem - so we use a relaxing assumption that the elements of h can have arbitrary values, which we denote as h_{relax}, such that the eigenvalues can written

$$\lambda_{\text{min}} = \min_{x \neq 0} \frac{h^H_{\text{relax}}\ L\ h_{\text{relax}}}{h^H_{\text{relax}}\ h_{\text{relax}}}. \tag{5.18}$$

There will be a zero eigenvalue corresponding to an eigenvector for $(D - \mathcal{A})$ with all equal values - this is the trivial solution where all points are clustered into a single object. The remaining k eigenvectors $\vec{h}_{\text{relax}}$ are used to construct the $(n \times k)$ matrix U, where each column in U corresponds to a $\vec{h}_{\text{relax}}$. We read off each row of U as a length-k vector, labelled by $\vec{y}_i$, which are each representative of one particle. The points $(\vec{y}_i)_{i=1,..,n}$ are clustered into groups using k-means clustering.

In a later chapter we will see how spectral clustering can be used to perform the task of jet clustering, compared with the traditional algorithms discussed in the previous chapter.

5.3 Evaluating Machine Learning Models

We have seen that generally models will be optimised by minimising some loss function, but how do we interpret how 'good' the model performance actually is? Due to the the nature of applying different loss functions, these can be non-trivial to interpret outside of comparing different models (with the same loss function) against one another. It turns out there are many possible ways of evaluating performance, which go beyond the more intuitive and obvious examples.

5.3.1 Evaluation Metrics

Before discussing explicit examples of some metrics, we will introduce some simple notation used to define them here, and note that much of this section applies mostly to classification rather than regression or unsupervised problems.

Given a set of predictions, a simple way of visualising the performance is to print out the confusion matrix. This takes the following form

$$\mathcal{M} = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix}, \tag{5.19}$$

where TP and FP refer to true positives and false positives, and FN and TN false negatives and true negatives. Consider a simple binary classifier which outputs 0 or 1,

TPs are instances with truth value 1, for which 1 was predicted, while FPs have truth value 0 but are predicted a 1, and vice versa for the FNs and TNs.

5.3.1.1 Accuracy

Perhaps the simplest evaluation metric is the model accuracy $\mathcal{A}$ - simply the number of correctly predicted instances divided by the total number of instances, expressed as number between 0 and 1. One can express the accuracy in terms of the positive and negative rates as

$$\mathcal{A} = \frac{TP + TN}{TP + FP + TN + FN}. \tag{5.20}$$

As alluded to, the accuracy is in fact a rather limited metric for evaluating a model. For example, in order to be reliable, clearly the dataset needs to be split 50/50 between positive and negative classes, otherwise models will become biased to predicting the more common class. In fact if we are predicting an occurance that occurs in around 10% of the data, a model that always outputs 0 will have a 90% accuracy score!

5.3.1.2 Precision and Recall

As the limitations in using accuracy perhaps imply, it is not always sensible to apply a ‘one size fits all’ evaluation metric, indeed the context of the problem might suit a particular metric. In the physics case of a jet tagger - this for example could depend on if we care more about a) identifying as many b-jets as possible, at the cost of predicting more false positives, or b) wanting to reduce the background as much as possible, sacrificing some true b-tags such that the final sample has fewer non-b’s incorrectly classified as b’s.

In fact the two scenarios described above can be accounted for by choosing different metrics, namely precision and recall. The precision $\mathcal{P}$ is defined as

$$\mathcal{P} = \frac{TP}{TP + FP}, \tag{5.21}$$

and gives a better score where the model reduces the number of FPs, such that $\mathcal{P}(FP \to 0) \to 1$. In practise, for the jet example, this means that the model will sacrifice identifying all of the b-jets, in order to reduce incorrectly applying b-tags.

Alternatively, there is the recall $\mathcal{R}$,

$$\mathcal{R} = \frac{TP}{TP + FN}. \tag{5.22}$$

In analogue with the precision, we have $\mathcal{R}(FN \to 0) \to 1$, such that a b-tagger will aim to tag as many of the truth b-jets as possible, while also sweeping up some additional non b-jets.

There is in fact a balancing act between $\mathcal{P}$ and $\mathcal{R}$ known as the precision recall trade-off. As one of these is increased, the other will decrease, such that there is a trade-off.

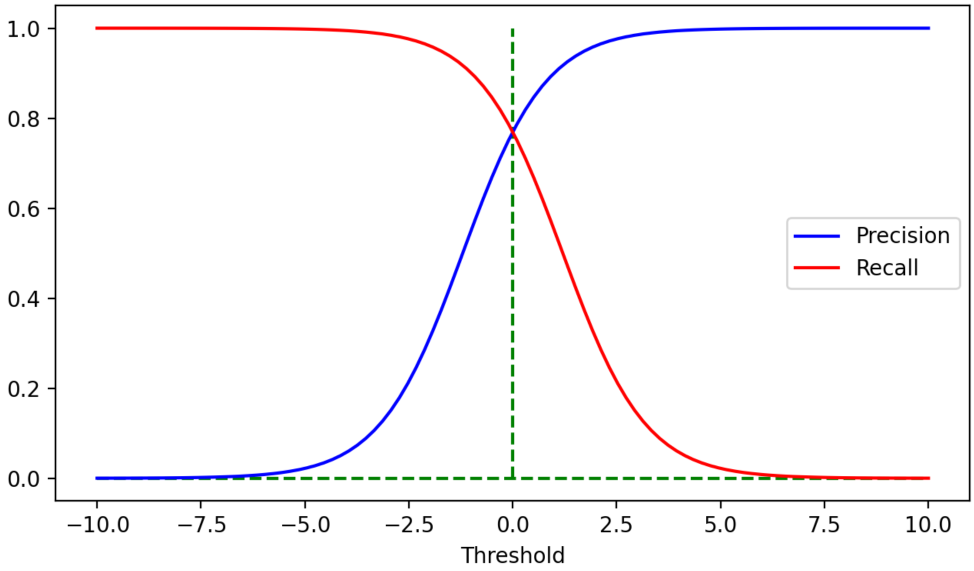


FIGURE 5.9: A visualisation of the precision recall trade off, as recall decreases, the precision increases, and vice versa.

In Fig. 5.9, we can think of each point along the x-axis (labelled "Threshold"), as a single model, and hence a value of precision and recall evaluated on the test set. As one moves along the x-axis, we can think of us moving the decision boundary for classifying an instance as a 0 or 1 - generally one might think to select 0.5, but by increasing (decreasing) this boundary we are, in effect, artificially changing the precision and recall.

5.3.1.3 F1-Score

In addition to precision and recall, there exists a more generalised metric, which acts as a balance between the two, known as the F1-score

$$\mathcal{F}_1 = \frac{2}{\mathcal{P}^{-1} + \mathcal{R}^{-1}} = \frac{2\mathcal{P}\mathcal{R}}{\mathcal{P} + \mathcal{R}}. \tag{5.23}$$

The $\mathcal{F}_1$ acts as a convenient middle ground between precision and recall, and in analogue with accuracy is easily interpreted as a number between 0 and 1, with 1 representing a perfect score.

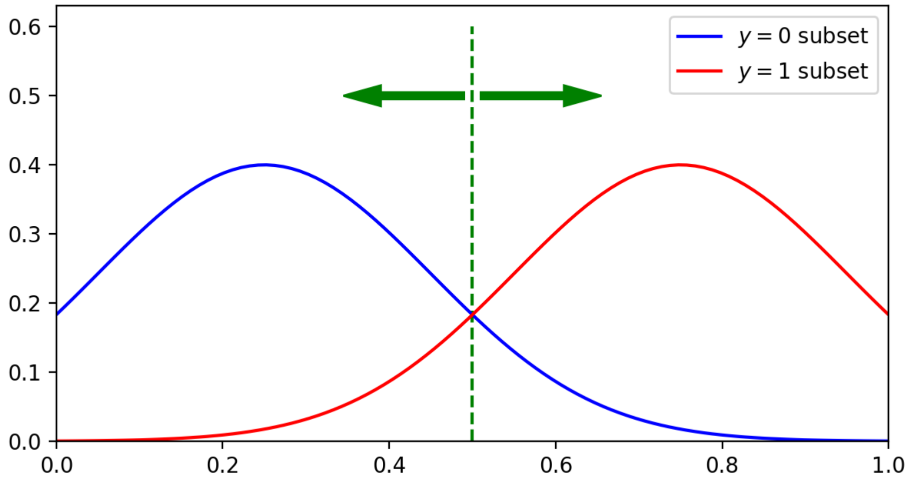


FIGURE 5.10: The shifting of the decision boundary, where the two curves represent the two subsets of data split by truth value. Instances to the left of the boundary will be predicted a 0, and to the right a 1.

5.3.1.4 ROC Curves

Another good test for model performance also builds off of the idea of moving a decision boundary, like with precision and recall. This is the receiver operating characteristic (ROC) score, derived from the ROC curve.

The ROC is very simple a plot of the true positive rate

$$TPR = \frac{TP}{P} \tag{5.24}$$

against the false positive rate

$$FPR = \frac{FP}{N}, \tag{5.25}$$

where P and N refer to the number of true positives and negatives in the dataset. The ROC score is then defined as the area under this curve, which will be 1 for a perfect classifier, and 0.5 for a random classifier.

5.4 Machine Learning in High Energy Physics

Now that we have been through a whirlwind tour of the fundamentals of machine learning, we will now review some of the examples of its uses in high energy physics research. For some overall reviews on the subject, we refer the reader to [68, 69, 70].

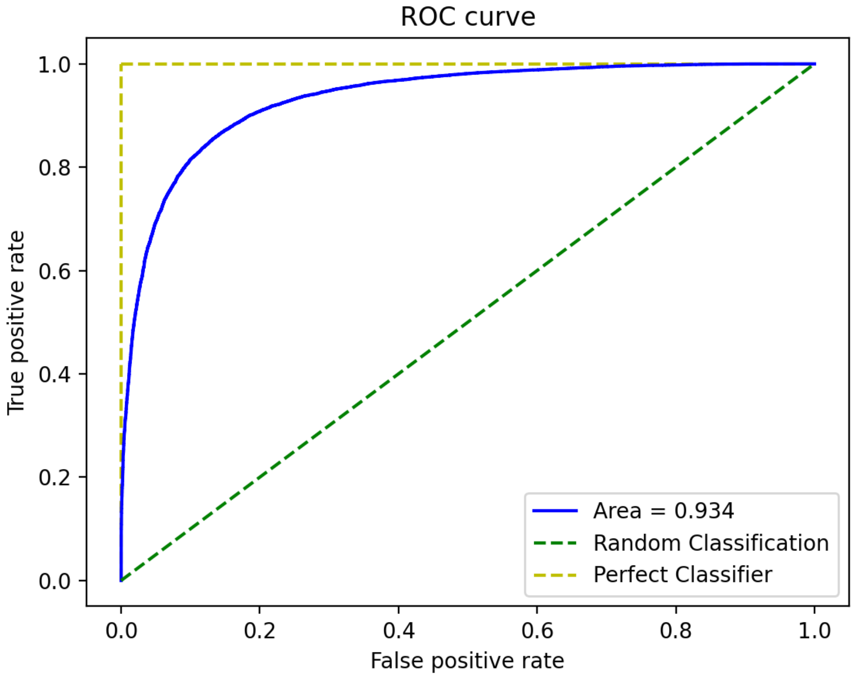


FIGURE 5.11: Demonstration of the ROC curve for a well performing classifier, with the ROC curve for a random classifier and perfect classifier.

5.4.1 Jet Tagging

As we have seen, perhaps one of the more intuitive examples of an ML application for particle physics is in jet tagging - it is a well defined supervised classification problem, and it is fairly straightforward to generate labelled training data via Monte-Carlo methods.

5.4.1.1 *b*-Tagging

The physics example we have referred to in this section is of tagging jets originating from *b*-quarks, which is in fact an active area of research in ML and particle physics. As seen previously, information on the secondary vertex can be indicative of the chance a given jet came from a *b*. It therefore follows that one could feed some information on the jet vertices (among any other things that might be useful) into an ML model to perform tagging.

Recent endeavors at the CMS collaboration can boast up to a 15% increase in tagging efficiency over other algorithms [71]. Other studies include the ATLAS detector [72] using recurrent neural networks (RNNs), as well as [73, 74].

5.4.1.2 Top Tagging

While b-jets are particularly useful for accessing the Higgs sector, jets arising from t-quarks are also useful in a variety of contexts, such as accessing dark matter candidates in certain BSM models. A good review on the variety of ML methodologies applied to top tagging can be found at [75].

Owing to their large mass, t-quarks actually decay before hadronisation, such that the final state from a top will be comprised of a series of objects which all came from the t. An example would be in the decay $t \to bW$, $W \to jj$, such that the final state is comprised of a b-jet (which itself requires tagging) and a pair of jets which reconstruct the W mass - although the W can decay in other ways, such as leptonically i.e. $W \to \mu^- \bar{\nu}$.

In scenarios with boosted tops, instead of trying to reconstruct each object individually, we can form a fat jet with a substructure comprising of these decay products - as seen in the section on jet physics. Boosted top taggers therefore rely on information of jet substructure, which can be represented as a series of variables (recall N-subjettinness for example), see [76, 77].

Alternatively one can actually ‘look’ inside the jet by visualising the region of the detector containing the jet constituents as pixels in an image, if which there are a number of studies [78, 79, 80].

5.4.2 Detector Physics

While not the focus of this work in particular, it is interesting to note the additional areas in which ML can be utilised in particle physics. Another area where ML can useful is workings of the detectors themselves.

5.4.2.1 Track Reconstruction

One such example is in reconstructing charged particle tracks as in [81, 82]. There are a large number of outgoing particles in a given event, and the detectors work in such a way that a series of spacetime points can be recorded when a particle ‘hits’ that layer - however from this series of points it is useful to trace out the trajectory a given particle follows for identification purposes.

5.4.2.2 Triggers

Another experimental problem at particle colliders are in the triggers, which are kinematic constraints used to decide whether or not to store a particular particle. It turns out the number of particles in a given LHC event is so large that it is impossible to record everything. Generally the stuff we are interested in is of a higher energy, and hence transverse momentum, and so low p_T particles can be discarded before they even have a chance to contaminate the event by not being stored in the first place.

Traditionally one would use kinematic constraints to determine this, however studies exploring the use of ML algorithms have been show to perform quickly and efficiently, such as using BDTs [83], or deep learning [84].

5.4.2.3 Pile-Up

ML has also recently been applied to the problem of pile-up mitigation, which we have seen before. For example in [85, 86, 87].

5.4.3 ML in BSM Physics

A final emerging use of ML in particle physics is moving away from experiments into the theoretical realms of BSM model building and exploration. Even limiting to 2HDMs, to explore the possible parameter space in the physical mass basis there are nine parameters,

$$m_h,\ m_H,\ m_A,\ m_{H^\pm},\ \sin(\beta-\alpha),\ \lambda_6,\ \lambda_7,\ m_{12}^2,\ \tan\beta. \tag{5.26}$$

To scan over these points over a sufficiently fine grid the number of points to check can become incredibly large - and hence it is in the interest of computational efficiency to be able to perform a targeted exploration of the parameter space. Examples of applying ML to such problems include [88, 89, 90].

Part II

Research, Results and Beyond

Chapter 6

Revisiting Jet Clustering Algorithms for New Higgs Boson Searches in Hadronic Final States

6.1 Introduction

Before starting we note the research here is published as a preprint [1] of which I am the lead author. The results presented represent an updated version of the paper in question.

The ultimate motivation of our study is to address the incomplete nature of the Standard Model (SM) of particle physics, by looking for signs of physics Beyond the Standard Model (BSM). In particular, we pose the question of whether different jet clustering techniques might be more or less suited to particular final states of interest, coming from topologies involving an extended Higgs sector from some 2-Higgs Doublet Model (2HDM). In such scenarios, as we will explain in more detail, high b-jet multiplicity final states are expected and a point worth addressing is which current experimental jet reconstruction is in fact optimal for these types of searches.

Several BSM scenarios with an enlarged Higgs sector allow for the existence of additional neutral Higgs boson states, CP-even or CP-odd. These resonances can be both lighter, or heavier, than the SM-like Higgs boson discovered at the LHC in 2012, which has a mass of approximately 125 GeV [91]. These new physics frameworks are ubiquitous in non-minimal models of Supersymmetry (SUSY) [92], in particular, but not only, in the Next-to-Minimal Supersymmetric Standard Model (NMSSM) [93]. If one departs from SUSY and remains with low-energy models, a BSM framework including these states in its particle spectrum is the 2HDM [94, 95, 21].

In such a 2HDM, two complex Higgs doublet fields undergo Electro-Weak Symmetry Breaking (EWSB), yielding five physical Higgs states, labelled as h, H (which are CP-even with, conventionally, $m_h < m_H$), A (which is CP-odd) and a pair of charged states with mixed CP properties, $H^{\pm}$. It is currently possible for the observed 125 GeV Higgs boson to be identified as either h or H in 2HDM scenarios. In both scenarios, when $m_h < m_H/2$ or $m_A < m_H/2$, the decays $H \to hh$ and/or $H \to AA$ (respectively) may occur. Taking $h(H)$ as the SM-like 125 GeV Higgs boson, for a $H(h)$ state with a mass of order 250(60) GeV or more(less), the dominant decay mode in a 2HDM is bottom-antibottom quark pairs [96, 97], i.e., $h \to b\bar{b}$, so that the final state emerging from the hard scattering $pp \to H \to hh$ is made up, at the partonic level, of four (anti)quarks[1], see Fig. 6.1. However, due to the confinement properties of Quantum Chromo-Dynamics (QCD), the partonic stage is not accessible by experiment, only the hadronic "jets" emerging at the end of the parton shower and hadronisation phase are seen.

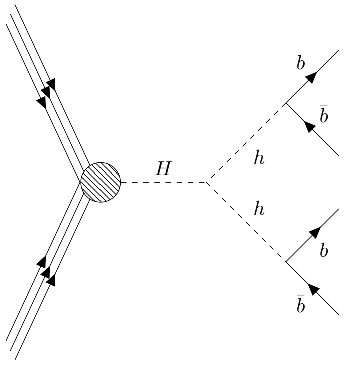


FIGURE 6.1: The 2HDM process of interest, where the SM-like Higgs state ($m_H = 125$ GeV) produced from gluon-gluon fusion decays into a pair of lighter scalar Higgs states, hh, each in turn decaying into $b\bar{b}$ pairs giving a four-b final state.

In order to decipher the source of these showers of hadrons, "jet clustering algorithms" are currently used. Jet clustering algorithms reduce the complexity of such final states by attempting to rewinding the showering, back to the parton it originated from, such that we move from a large sample of particles to a smaller number, each of which represents a state emerging from the hard interaction of interest in a given event. In other words, we characterise the sample of particles originating from a single parton as an object in itself - a jet. Needless to say, there is a variety of jet clustering algorithms available and we will dwell at length on these in a forthcoming section.

The purpose of this paper is to determine whether alternative jet reconstruction tools, in particular a modification to traditional sequential combinations alogrithms employing a variable inter-jet distance measure [50] (so-called 'variable-R' algorithms, where R represents a typical cone size characterising the jet), might be better suited to the

---

[1] Notice that the same argument can be made for the case of $pp \to H \to AA \to b\bar{b}b\bar{b}$ when $m_A < m_H/2$.

four-b final states coming from 2HDMs. We approach this problem from a simplified, theoretical perspective, in order test a range of different combinations, in order to inform whether a detailed experimental analysis might be worth pursuing. Furthermore, the four-b final state that we are invoking here is an ubiquitous signal of BSM Higgs boson pairs which are lighter than the SM one so that they can be produced from it[2], crucially giving access (through the extraction of the h state properties) to key features of the underlying BSM scenario, e.g., in the form of the shape of the Higgs potential, hence, of the vacuum stability and perturbative phases of it.

While the above outlines that the problem of optimal jet reconstruction is clearly an experimental endeavour, we stress that this study is undertaken at a theoretical level. The aims of this paper are to employ a simplified analysis in order to compare the relative performance of traditional fixed-R jet clustering against a variable-R method. A comprehensive, more realistic, experimental investigation is left to a future study. For example, another key feature of the hadronic final state initiated by b-quarks that we intend to study is that the emerging jets can be "tagged" as such, unlike the case of lighter (anti)quarks and gluons, which are largely indistinguishable from each other. Here, we implement a simplified method of tagging using Monte Carlo (MC) truth information on the b-partons, along with a probabilistic implementation of inefficiencies. For a more detailed discussion on b-tagging at detectors, we refer the reader to [98].

The layout of the paper is as follows. In the next section, we describe how we performed jet reconstruction and b-tagging as well as discuss the tools used for our simulations. In the following one, we present our results for both signal and background. Then, we conclude.

6.2 Methodology

6.2.1 Jet Clustering Algorithms

In order to extract proper physics from hadronic sprays found in particle detectors, algorithms are used to characterise the detected radiation into distinguishable objects, the aforementioned jets. There is indeed a rich history associated with the development and evolution of algorithms for jet definition, beginning in 1977 with Sterman and Weinberg [44], initially deployed in the context of $e^+e^- \to$ hadron scatterings. For a detailed look into the evolution of jet clustering, we refer the reader to [99]. Here we focus on the modern implementations.

---

[2] Here, ubiquitous refers to the fact that this signal is very typical of a variety of BSM scenarios, so that we effectively use the 2HDM for illustration purposes. Our results can therefore be applied to the case of other new physics models.

The type of algorithms currently utilised are known as sequential recombination algorithms, or 'jet clustering algorithms' [99]. Rather than categorising the entire event at once, as in previous approaches, each particle in the event is considered and all are iteratively combined together based on some inter-particle distance measure, until all particles are gathered into stable jets.

The type of algortihms currently deployed at the LHC all take a similar form, descending from the generalised k_T algorithm used initially in e^+e^- colliders. This uses an inter-particle distance measure which can be written in the form

$$d_{ij} = \min(p_{Ti}^n, p_{Tj}^n)\Delta R_{ij}^2, \tag{6.1}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 - (\phi_i - \phi_j)^2$ is the angular separation between particle i and j, with y and ϕ being the rapidity and azimuth of the associated final state hadron. They also make use of the 'beam distance', which is the separation between object i and the beam B,

$$d_{Bi} = p_{Ti}^n R^2. \tag{6.2}$$

Note that we use the same notation as in [50], where R^2 is included in the definition of d_{Bi}. (An alternative convention is to embed R^2 into the definition of d_{ij} such that $d_{ij} = \min(p_{Ti}^n, p_{Tj}^n)\frac{\Delta R_{ij}^2}{R^2}$, leaving $d_{Bi} = p_{Ti}^n$, like in [100].) For a set of particles, all possible d_{ij}'s and d_{Bi}'s are calculated and the minimum is taken. If the minimum is a d_{ij}, objects i and j are combined and the process is repeated. If, instead, a d_{Bi} is the minimum, then i is declared a jet and removed form the sample. This procedure is then repeated until all objects are classified into jets.

In d_{Bi}, R is a fixed input variable which dictates the size of the jet and acts as the cut-off for any particle pairing. If we consider some pair of particles i and j, with i having lower p_T (and hence being selected in d_{ij}), we can write (for $n \geq 0$)

$$d_{ij} = \Delta R_{ij}^2 p_{Ti}^n = \frac{\Delta R_{ij}^2}{R^2} d_{Bi}. \tag{6.3}$$

Since we require the ratio $\frac{\Delta R_{ij}^2}{R^2} < 1$ to avoid declaring i a jet over merging i with j, we can see that R acts as an effective cut off for the maximum separation of two pseudojets to be combined and, hence, it is proportional to the final jet size. From this general formulation, the main two jet clustering algorithms currently in use at the LHC are the Cambridge-Aachen (CA) [101, 49] one and the anti-k_T[48] one, which use the above expressions with $n = 0$ and -2, respectively [102].

6.2.2 Jet Clustering with Variable-R

There has in fact been a more recent development to these techniques. One notices that the above algorithms require as input a fixed parameter, R, which in the case of the anti-k_T algorithm effectively acts as the jet radius, hence a cone parameter when tracing the jet back to the interaction point[3]. Recall this acts as a cut off for combining hadrons and can therefore be interpreted as implementing a size limit on the jets depending on the particle separation.

The angular spread of the final jet constituents has a dependence on the initial partons p_T. For higher p_T objects the decay products will be more tightly packed into a more collimated cone whereas for low p_T objects one would expect the resulting jet constituents to be spread over some wider angle. One can therefore imagine the need for carefully selecting the R value used for clustering depending on the p_T of the final state jets, but what about in a multi-jet scenario where the final state partons have a wide range of p_T's?

The aforementioned variable-R jet clustering algorithm [50] alters the above scheme so as to adapt to events with jets of varying cone size. A modification to the distance measure d_{ij} is made, by replacing the fixed input parameter R with a p_T dependent $R_{\text{eff}}(p_T) = \frac{\rho}{p_T}$, where ρ is a chosen dimensionful constant (taken to be $\mathcal{O}(\text{jet } p_T)$). With this replacement, the beam distance measure becomes

$$d_{Bi} = p_{Ti}^{n} R_{\text{eff}}(p_{Ti})^2. \tag{6.4}$$

When the distance measures are calculated, d_{Bi} will therefore be suppressed for objects with larger p_T and hence these objects become more likely to be classified as jets. For low p_T objects, d_{Bi} is enhanced and so these are more likely to be combined with a near neighbour, thus increasing the spread of constituents in the eventual jet.

We hypothesise therefore that, in multijet signal events where one might expect signal b-jets with a wide spread of different p_T's, a variable-R reconstruction procedure could improve upon the performance of traditional fixed-R routines. In particular, using a variable-R alleviates the balancing act of finding a single fixed cone size that suitably engulfs all of the radiation inside a jet, without sweeping up too much outside 'junk'.

As a brief visualisation, we can map the constituents of b-tagged jets (herafter, b-jets for short) in the same event, which have been clustered using both a variable-R and fixed $R = 0.4$ scheme, as seen in Fig. 6.2. We notice that, for the leading and sub-leading

---

[3] It is an important distinction to notice that R only looks like a cone parameter for the anti-k_T algorithm ($n = -2$). In this case, the 'min' in Eq. (6.1) picks out the larger p_T pseudojet. If we take i as the higher p_T object, then, if j is further from i than R, i and j will not be combined. In other words, if i is close to the eventual jet axis, low momenta j's will not be included, so that in the end R effectively ends up as the jet radius. While this point is important, we also note that we mostly use the anti-k_T algorithm in this study and so referring to R as an effective cone size is largely justified.

b-jets, the jet content is roughly the same. For the lower p_T jets, however, the variable-R jets gather a wider cone of constituents. The loss of constituents will harm our ability to accurately reconstruct Higgs masses when analysing b-jets. In Fig. 6.3 we see a case where using a larger fixed cone ($R = 0.8$), to try and gather all of the constituents, only resolves three b-jets. Variable-R however 'finds' all four b-jets expected from the signal. We can see that fixed-R sweeps radiation from a nearby jet into the leading b-jet, whereas variable-R is able to resolve both due to the larger p_T (and hence smaller R_{eff}) of the leading b-jet, while also having a large enough cone to suitably reconstruct the lower p_T jets.

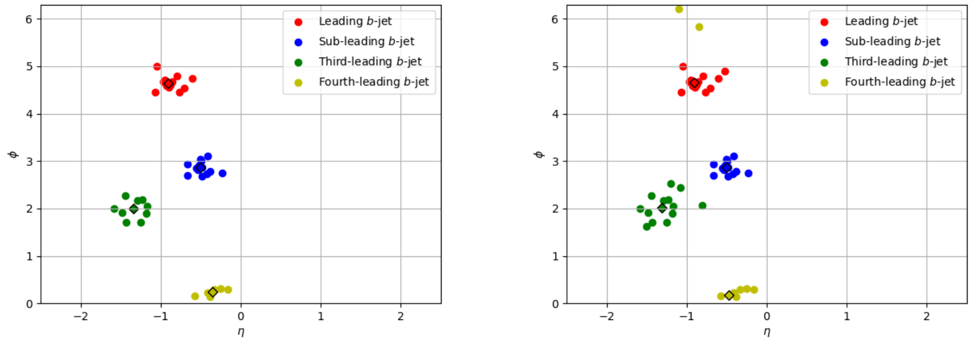


FIGURE 6.2: The same MC event in (η, ϕ) space. Tracks have been clustered with (left) a fixed $R = 0.4$ and (right) variable-R algorithm. The coloured points are the constituents of the corresponding b-jet in the legend and black outlined diamonds are at the overall (η, ϕ) coordinates of the formed b-jet. The anti-k_T algorithm is used in both cases.

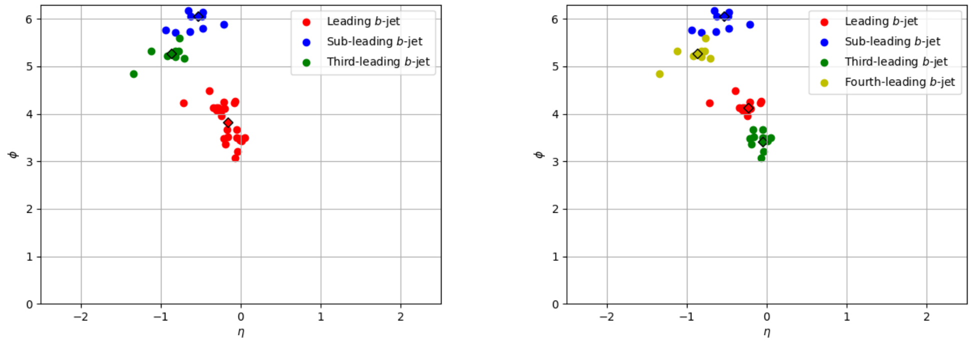


FIGURE 6.3: Same plot as in Fig. 6.2, however, here, the given event is clustered into three b-jets when a fixed $R = 0.8$ is used (left) and four b-jets when we use a variable-R approach (right).

6.2.3 Implementation of b-Tagging

In this paper we implement a simplified MC informed b-tagger. For events clustered using a fixed-R cone size, jets within angular distance R from each parton level b-(anti)quark are searched for and tagged as appropriate. For scenarios where multiple jets are found, the closest is taken as the assignee for the b-tag. When the variable-R approach is used, the size of the tagging cone is taken as the effective size of the jet, R_{eff}, defined above.

In addition, we account for the finite efficiency of identifying a b-jet as well as the non-zero probability that c-jets and light-flavour plus gluon jets are mistagged as b-jets. We apply the variable mistag rates and tagging efficiencies from the `Delphes` CMS card.

6.2.4 Simulation Details

We consider two sample benchmark points, one of which where we take h as the SM-like 125 GeV Higgs Boson, and one with m_H. In the former case (which we will refer to as BP1), we use $m_H = 700$ GeV, and in the latter (BP2 henceforth) $m_h = 60$ GeV. Both benchmarks are in a 2HDM Type-II (2HDM-II henceforth), which have been tested (and pass as not currently excluded) against theoretical and experimental constraints by using 2HDMC [23], HiggsBounds [107], HiggsSignals [108] as well as checking flavour contraints with SuperISO [109]. We generate samples of $\mathcal{O}(10^5)$ events, with $\sqrt{s} = 13$ TeV. In SuperISO, we test against the following flavour constraints on b-meson decay Branching Ratios (BRs) and mixings, all to a 2σ level: $\text{BR}(b \to s\gamma)$, $\text{BR}(B_s \to \mu\mu)$, $\text{BR}(D_s \to \tau\nu)$, $\text{BR}(D_s \to \mu\nu)$, $\text{BR}(B_u \to \tau\nu)$, $\frac{\text{BR}(K\to\mu\nu)}{\text{BR}(\pi\to\mu\nu)}$, $\text{BR}(B \to D_0\tau\nu)$ and $\Delta_0(B \to K^*\gamma)$.

The production and decay rates for the subprocesses $gg, q\bar{q} \to H \to hh \to b\bar{b}b\bar{b}$ are presented in Tab. 7.1, alongside the 2HDM-II input parameters. (Notice that the H and h decay widths are of order MeV, hence much smaller than the detector resolutions in two- and four-jet invariant masses, respectively, so that the Higgs states can essentially be treated as on-shell.) In the calculation of the overall cross section, the renormalisation and factorisation scales were both set to be $H_T/2$, where H_T is the sum of the transverse energy of each parton. The Parton Distribution Function (PDF) set used was `NNPDF23_lo_as_0130_qed` [110]. Finally, in order to carry out a realistic MC simulation, the toolbox described in Fig. 6.4 was used to generate and analyse events [111, 112, 113, 114][4].

---

[4] Note that we use the Leading Order (LO) normalisation for the signal cross sections here, for consistency with the fact that most of the background ones in our forthcoming analysis are only implemented at LO. While this affects our final results on event rates and significances, we re-instate here that the main purpose of our paper is to assess the jet clustering performance, rather then the exact values of signal and backgrounds rates.

Label	m_h (GeV)	m_H (GeV)	$\tan\beta$	$\sin(\beta-\alpha)$	m_{12}^2	BR($H \to hh$)	BR($h \to b\bar{b}$)	$\sigma(gg \to H \to hh \to 4b)$ (pb)
BP1	125	700.668	2.355	-0.999	1.46×10^5	6.218×10^{-1}	6.164×10^{-1}	1.870×10^{-2}
BP2	60	125	1.6	0.1	4×10^3	6.764×10^{-1}	8.610×10^{-1}	6.688

TABLE 6.1: The 2HDM-II parameters and cross sections of the process in Fig. 6.1 for each benchmark point.

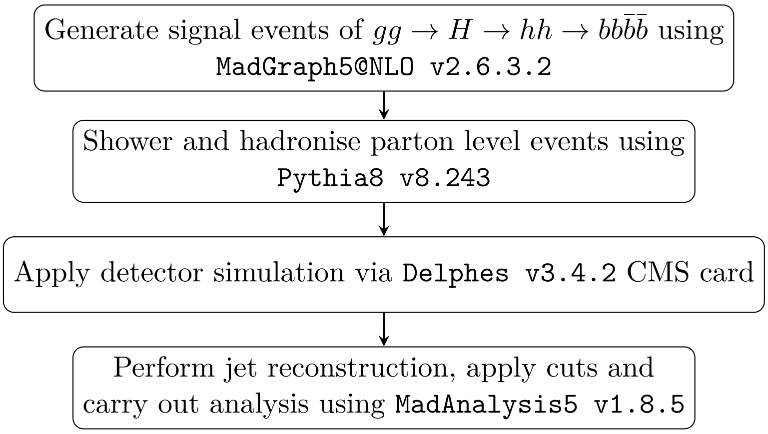


FIGURE 6.4: Description of the procedure used to generate and analyse MC events.

6.3 Cutflow

Before introducing the full sequences of cuts that we have adopted here, some discussions are in order on their possible choice, as intimated in the previous section. In existing four b-jet analyses by the ATLAS and CMS collaborations, seeking to extract chain decays of Higgs bosons like the ones considered here from the background, rather restrictive cuts have been used for the ensuing fully hadronic signature. Taking CMS as an example, upon enforcing the same p_T cuts on b-jets as in Ref. [115] on BP2, we noticed that the signal selection efficiency was too low (in various respects, as described later) to enable one to create a MC sample suitable for experimental analysis assuming Run 2 and 3 luminosities. For BP1, cuts informed by [116] of all four b-jets satisfying $p_T >$50 GeV are used. In the case of BP2, we use a flat cut on all four b-tagged jets of 20 GeV. It remains to be seen if this is viable at the LHC, but for the purpose of this study we present results in this regime as a demonstration of the effect of using a variable-R jet reconstruction algorithm on low-p_T jets from 2HDM-II decays into $bb\bar{b}\bar{b}$ final states.

6.4 Results

In this section we present the results for our signal at both the parton and detector level. In the latter case, we also discuss the dominant backgrounds, due to QCD $4b$

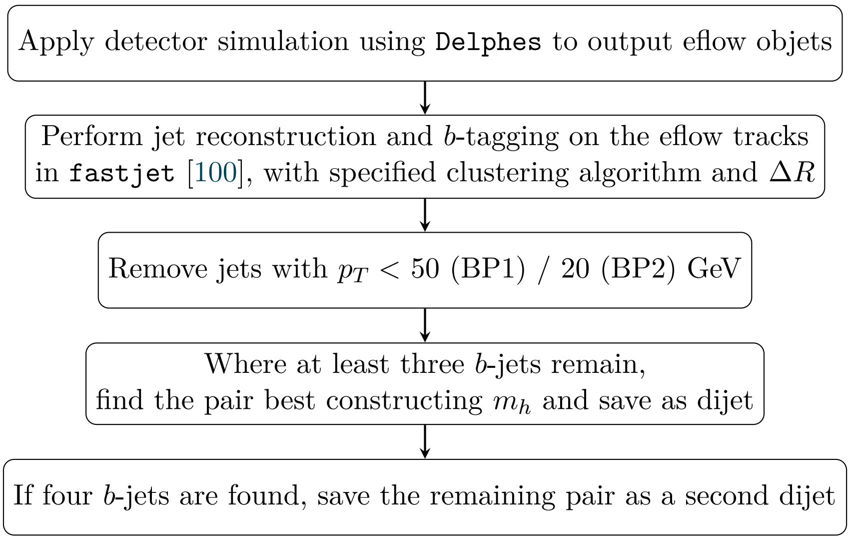


FIGURE 6.5: Description of our initial procedure for jet clustering, b-tagging and selection of jets. Notice that the starred cut $(*)$ will eventually be modified in our optimised b-jet selection. Also note that the bulk of our analysis is performed at particle rather than detector detector, so MC truth information is used for cuts on jet constituents.

production, $gg, q\bar{q} \to Zb\bar{b}$ and $gg, q\bar{q} \to t\bar{t}$[5].

6.4.1 Parton Level Analysis

At the Matrix Element (ME) level, all the events have four b-quarks originating from the decay of the two light Higgs bosons (h). We plot the R separation between the b-quarks coming from the same light Higgs state (see upper panel of Fig. 6.6). The two distributions corresponding to BP1 and BP2 are markedly different. This can be understood as follows. In general, the angular separation between the decay products a and b in the resonant process $X \to ab$ can be approximated as $\Delta R(a,b) \sim \frac{2m_X}{p_T^X}$. Hence, we plot in the middle panel of Fig. 6.6 the transverse momentum of each of the h bosons.

For $m_h = 60$ GeV, the light Higgs boson has less p_T than for lower values (owing to the smaller $m_H - m_h$ mass difference), therefore, the b-quarks are more widely separated in this case, compared to $m_H = 700$ GeV. In the light of this, we can already conclude that there is a strong correlation between the mass difference $m_H - m_h$ and the cone size of the jet clustering algorithm that ought to be used. In particular, we can say that, in order to maximise the number of jets[6] for different choices of the light Higgs boson mass, we need to vary the jet radius parameter. That is, a fixed jet radius parameter may not be suitable here for all mass configurations. In the lower panel of Fig. 6.6, we finally plot the ΔR separation between the two light Higgs states. For the configuration

---

[5] In fact, we have checked that the additional noise due to $t\bar{t}b\bar{b}$ events as well as hadronic final states emerging from $W^+W^-, W^\pm Z$ and ZZ production and decay are negligible, once mass reconstruction around m_h and m_H is enforced.

[6] This is done also with a view at background rejection.

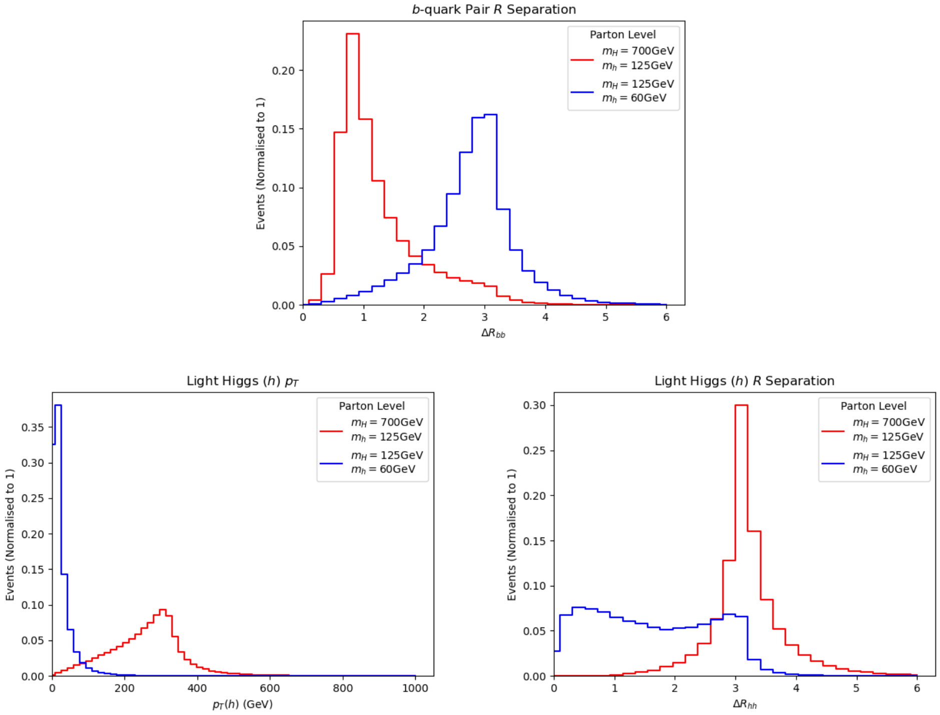


Figure 6.6: Upper panel: the ΔR distribution between the two b-partons originating from the same h. Lower left panel: the p_T distribution of the light Higgs boson h originating from H decay; lower right panel: the ΔR distribution between the two h states originating from the H decay. No (parton level) cuts have been enforced here.

in BP1, it is clear (since $\Delta R \approx \pi$) that the $H \to hh$ decay is dominantly back-to-back (in the laboratory frame). However, for $m_h = 60$ GeV, there is a double peak structure. This occurs due to a recoil effect from Initial State Radiation (ISR), which only becomes apparent at the mass boundary where $m_H \simeq 2m_h$. The inability of the two emerging h states to fly apart implies some overlapping of the b-quark momenta. Hence, we expect that the signal, upon enforcing a jet clustering algorithm, will have a rather high b-jet multiplicity, so long that the two b-jets stemming from h decays are resolved, unless detector acceptance and signal selection cuts reduce it, which is quite possible given the light masses considered for the h state in relations to typical jet p_T thresholds used in applying b-tagging. We will investigate this later.

As a final study, in fact, the p_T of the b-quarks is plotted. This is done in Fig. 6.7. From the top histogram we can see that, while the range of p_T's in BP1 is much larger, in both mass configurations the b-quarks have a range of p_T's, and hence one would expect the resulting jets to have a similar spread. In particular, we also plot the highest and lowest p_T's amongst the b-quarks in a given event (lower left and right frames, respectively), and notice a stark difference in both cases. Further to the discussion in Sec. 6.2.2, one would therefore expect the resulting spread of radiation from each signal b-quark to vary in solid angle and hence the resulting jets be of differing sizes. This thus motivates the need for a jet reconstruction sequence that behaves sensibly for jets of various cone sizes. Therefore, in the next section, we firstly test how jet clustering with fixed-R input behaves and then introduce the variable-R algorithm.

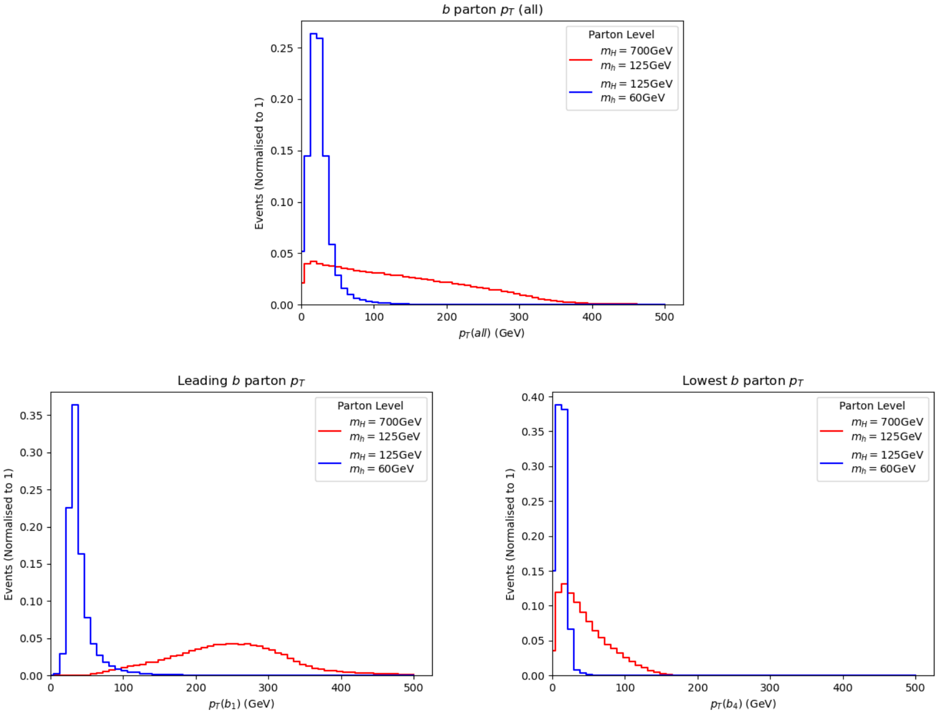


FIGURE 6.7: Upper panel: the p_T distribution for all b-quarks. Lower left panel: highest p_T amongst the b-quarks; lower right panel: lowest p_T amongst the b-quarks. No (parton level) cuts have been enforced here.

6.4.2 Jet Level Analysis

In this section we consider a jet level analysis, using hadronised parton showers that have been run through detector simulation and clustered into jets. We will compare the kinematic distributions of final state b-jets, when clustered with a fixed cone, and with variable-R, for both mass configurations BP1 and BP2. In particular we will be interested in the b-jet multiplicity, that is the number of b-tagged jets in a given event. This is of course indicative of how 'well' our clustering is performing, in that we know the final state has four b-quarks, and so a good algorithm should reconstruct all four of these frequently. We will also investigate the mass distributions of b dijets and four b-jet masses, which indicate our ability to observe the signals containing BSM Higgs Bosons.

We first consider the effect of a variable vs fixed cone strategy by observing kinematic variables from signal events for each benchmark point. We choose a value of $R = 0.4$, and use the anti-k_T algorithm throughout (The results for the CA scheme are very similar, so we refrain from presenting them). For variable-R, we use different values for each of the benchmarks as follows, for BP1 we use $\rho = 100$GeV, for BP2 we use

$\rho = 20\text{GeV}$. These values are inferred from the p_T scale of the fixed cone b-jets. Finally, we use $R_{\text{min}} = 0.4$ and $R_{\text{max}} = 2.0$ throughout wherever variable-R is used.

We begin by observing the b-jet multiplicity, for each of the benchmarks/algorithms.

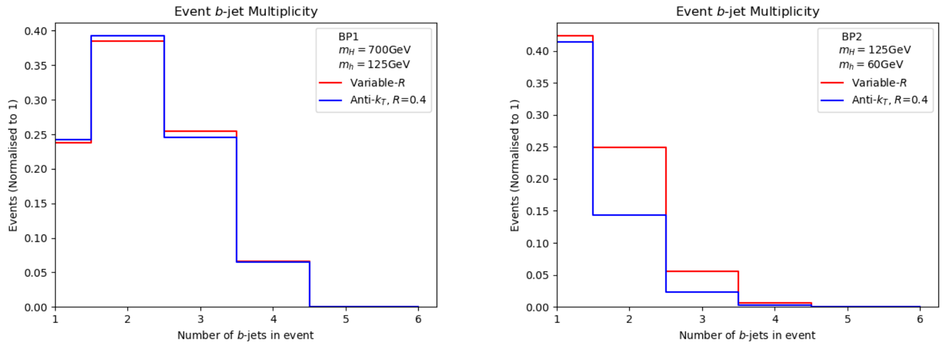


FIGURE 6.8: Left panel: The b-jet multiplicity distributions for BP1. Right panel: For BP2.

The stark difference between the two plots is due to the relative kinematics of the final state b-jets. Due to the different mass configurations, b-jets from BP2 have significantly lower p_T than those from BP1, and so significantly more are lost to the trigger, as well as from the (p_T dependent) b-tagging efficiencies.

As for the effect of variable-R, in the BP1 case there is very small increase toward events with higher b-jet multiplicity. The shift is significantly clearer in BP2.

In order to extract evidence of new physics from b-jet signals, we look at the invariant mass of dijet systems, in order to reconstruct a mass that resembles the resonance from which the pair of jets originated.

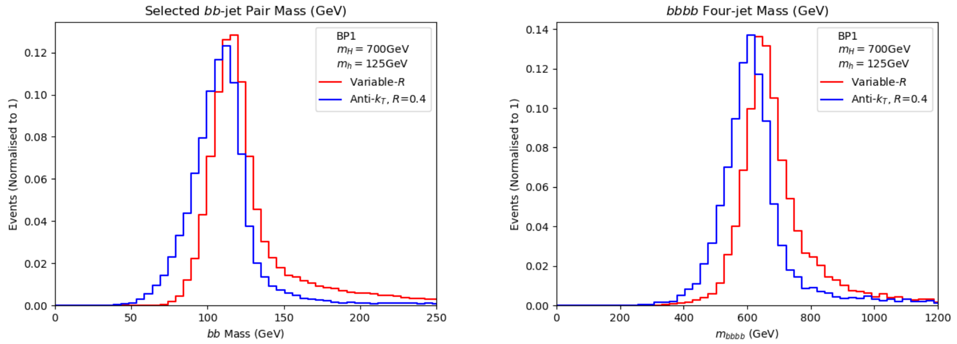


FIGURE 6.9: Left panel: The b-dijet invariant masses for BP1. Right panel: Four-b invariant mass.

From Fig.6.9, we can see more definitively the benefits of using a variable-R jet clustering algorithm. In both cases, the peak of the variable-R plot is shifted right of the fixed cone, closer to the expected mass of the of respective decaying Higgs. Finally, for the invariant mass of all four b-tagged jets we have

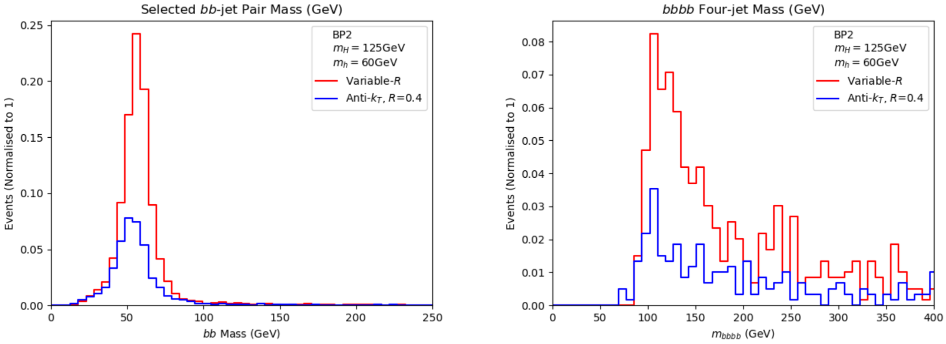


FIGURE 6.10: Left panel: The four b-jet invariant masses for BP2. Right panel: Four-b invariant mass.

The same behaviour for four b-jets masses (Fig. 6.10) can be seen as in the dijet plot, events clustered with variable -R have the four b invariant mass more closely aligned with the expected positions at m_H.

6.4.3 Signal-to-Background Analysis

Clearly a good algorithm should not just amplify the signal, but also avoid sculpting the backgrounds. As a final exercise, we perform a calculation of the signal-to-background rates, so as to compare the various jet reconstruction procedures mentioned in this paper also in connection with their performance in dealing with events not coming from our BSM process. In order to do so, we perform the selection procedure described in Fig. 6.14. We use the anti-k_T measure throughout but conclusions would not change in case of the CA one.

6.4.3.1 Jet Quality Cuts

Before proceeding on with the analysis, there is one additional cut we consider in order to compare signal over background significance of the two approaches. As per [50], we employ the use of jet quality cuts, which themselves are well motivated. In particular, we compute the distance δ between the energy and p_T centre of the jets, which are defined as follows:

$$\vec{P}_E = \sum_i E_i \hat{\vec{p}}_i, \;\; \vec{P}_{p_T} = \sum_i p_{Ti} \hat{\vec{p}}_i, \tag{6.5}$$

where i labels each constituent of a given jet, $\hat{\vec{p}_i}$ is the four-momenta of the i^{th} constituent, normalised to unity, and E_i and p_{Ti} are the energy and transverse momentum of the i^{th} jet constituent respectively. In applying the jet quality cut, we are imposing that, for a given jet i,

$$\Delta R(\vec{P}_E, \vec{P}_{p_T}) < \delta, \tag{6.6}$$

where δ is a user defined cut-off. In order to get an idea of the usefulness of jet quality cuts, we can look at the b dijet and four jet invariant mass peaks corresponding to m_h and m_H, on samples of events with and without quality cuts.

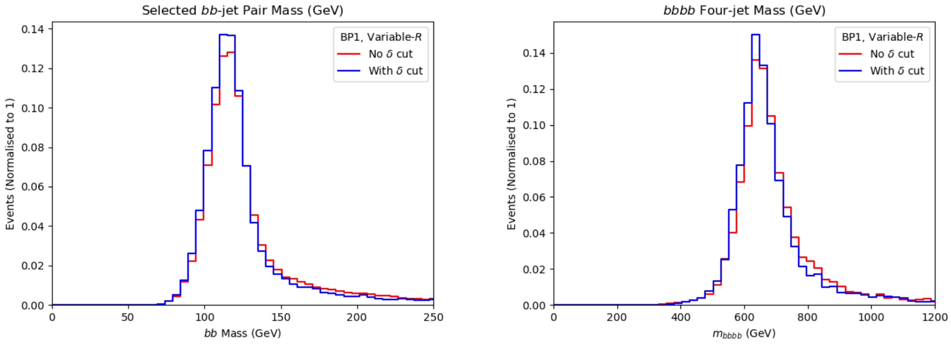


FIGURE 6.11: Left panel: The b-dijet invariant masses for BP1, with and without the addition of jet quality cuts as defined in Eq. 6.5 and 6.6. Right panel: The four b-jet invariant mass. Here we have used a value of $\delta = 0.05$ for BP1.

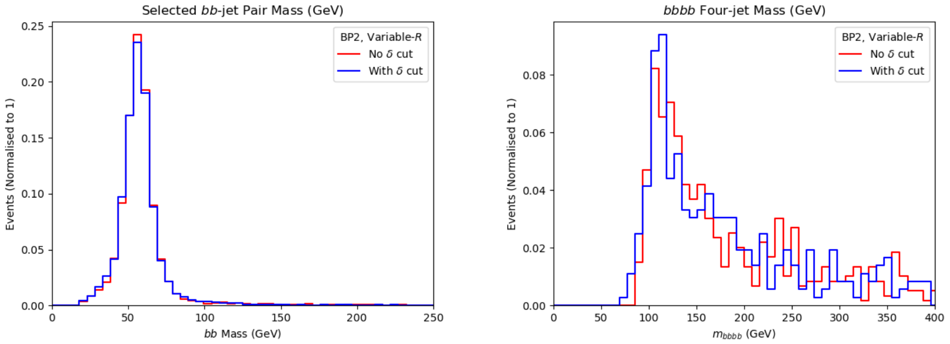


FIGURE 6.12: Left panel: The b-dijet invariant masses for BP2, with and without the addition of jet quality cuts as defined in Eq. 6.5 and 6.6. Right panel: The four b-jet invariant mass. Here we have used a value of $\delta = 0.1$ for BP2.

We can see in Fig. 6.11 the increased height of the mass peaks that comes with applying jet quality cuts on the benchmark point with $m_H = 700$GeV. The effect is also present for the $m_h = 60$GeV benchmark in Fig. 6.12, in particular in the four b-jet mass distribution.

We further note that, while there is a hint of signal yield improvement with jet quality cuts, the main gains come from the reduction of backgrounds, which we see in the following section.

6.4.3.2 Signal Selection

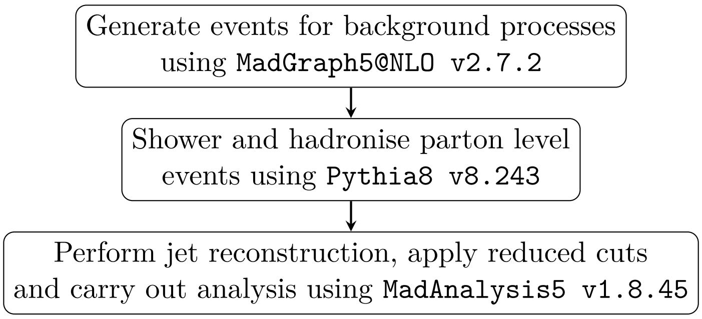


FIGURE 6.13: Description of the procedure used to generate and analyse MC events for background processes.

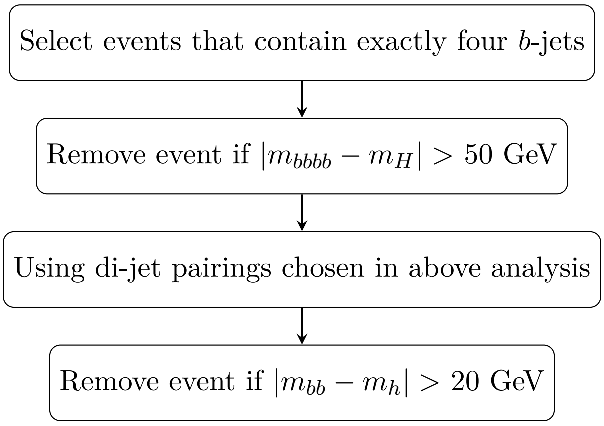


FIGURE 6.14: Event selection used to compute the signal-to-background rates.

To carry out this exercise, we generate and analyse $pp \to b\bar{b}b\bar{b}$, $pp \to Zb\bar{b}$ and $pp \to t\bar{t}$ background processes using the toolbox described in Fig. 6.13 [111, 112, 113, 114]. Tab. 6.2 contains the cross sections in pb for signal and the various background processes upon applying the aforementioned cuts and mass selections, including the jet quality cuts.

It is clear from the data obtained that the QCD-induced $pp \to b\bar{b}b\bar{b}$ process is the dominant background channel[7], followed by $pp \to Zb\bar{b}$ and $pp \to t\bar{t}$. Our next step is then to calculate the event rates in order to get the significances for two values of (integrated) luminosity, e.g., $\mathcal{L} = 140$ and 300 fb^{-1}, corresponding to full Run 2 and 3

---

[7] In fact, we have computed the full four-jet sample produced by QCD, i.e., including all four-body partonic final states, yet, in presence of the described kinematical selections and b-tagging performances, the number of non-$b\bar{b}b\bar{b}$ events surviving is negligible [106, 117, 118].

data samples, respectively. The event rate (N) for the various processes is given by:

$$N \ = \sigma \times \mathcal{L}. \tag{6.7}$$

After the event rates have been calculated, we simply evaluate the significance, Σ, which is given by (as a function of signal S and respective background B rates)

$$\Sigma = \frac{N(S)}{\sqrt{N(B_{b\bar{b}b\bar{b}}) + N(B_{Zb\bar{b}}) + N(B_{t\bar{t}})}}. \tag{6.8}$$

It is then clear from Tabs. 6.3–6.4 that the variable-R approach works better than fixed-R one also in providing the best significances, no matter the choices of R for the latter. The improvement in the final significances is indeed very significant. This should not be surprising, given the ability of the former in outperforming the latter from the point of view of kinematics. Again, while the signal-to-background analysis has been performed for the anti-k_T algorithm, the same conclusions are reached for the CA case.

Process	variable-R		$R = 0.4$	
	BP1	BP2	BP1	BP2
$pp \to H \to hh \to b\bar{b}b\bar{b}$	2.077×10^{-4}	8.962×10^{-3}	1.254×10^{-4}	3.210×10^{-3}
$pp \to b\bar{b}b\bar{b}$	3.798×10^{-3}	2.131	1.651×10^{-3}	9.470×10^{-1}
$pp \to t\bar{t}$	7.973×10^{-4}	2.850×10^{-2}	1.595×10^{-3}	2.217×10^{-2}
$pp \to Zb\bar{b}$	9.689×10^{-6}	2.627×10^{-2}	3.876×10^{-6}	9.695×10^{-3}

Table 6.2: Cross sections (in pb) of signal and background processes upon enforcing the reduced cuts plus the mass selection criteria $|m_{bbbb} - m_H| < 50$ GeV and $|m_{bb} - m_h| < 20$ GeV for the various jet reconstruction procedures.

	variable-R	$R = 0.4$
BP1	1.145	0.823
BP2	2.268	1.214

Table 6.3: Final Σ values calculated for signal and backgrounds for $\mathcal{L} = 140$ fb^{-1} upon enforcing the reduced cuts plus the mass selection criteria $|m_{bbbb} - m_H| < 50$ GeV and $|m_{bb} - m_h| < 20$ GeV for the various jet reconstruction procedures.

	variable-R	$R = 0.4$
BP1	1.676	1.205
BP2	3.320	1.777

Table 6.4: Final Σ values calculated for signal and backgrounds for $\mathcal{L} = 300$ fb^{-1} upon enforcing the reduced cuts plus the mass selection criteria $|m_{bbbb} - m_H| < 50$ GeV and $|m_{bb} - m_h| < 20$ GeV for the various jet reconstruction procedures.

6.4.4 Variable-R and Pile-Up

It has been noted that the nature of variable-R, combined with our reduced p_T restrictions, allow for wider cone signal b-jets. We therefore perform a quick analysis of events with pile-up and multiple parton interactions (MPI), using variable-R. As briefly mentioned, in order to perform such a study a proper detector simulation is required. We therefore now employ the use of `DELPHES`, passing our hadronised events (and pile-up, simulated in `Pythia8`) through the CMS card (with the same b-tagging efficiencies and c/light-jet mistag rates as before). We also perform the same exercise with a fixed cone of $R = 0.4$ to compare.

We present in Figs. 6.15-6.16 the m_{bb} and m_{bbbb} spectra, as described earlier, for the signal with pile-up and MPI, comparing $R = 0.4$ and variable-R jet clustering. We note that, with the addition of pile-up, we use a different value for the variable-R parameter ρ. Furthermore we also do not use any jet quality cuts here, as we have discussed the main benefit of such techniques are in background reduction, which we do not consider for pile-up. We see that, with pile-up simulation added on top of our signal events, using a variable-R jet reconstruction and tagger passes significantly more events in its selection procedure.

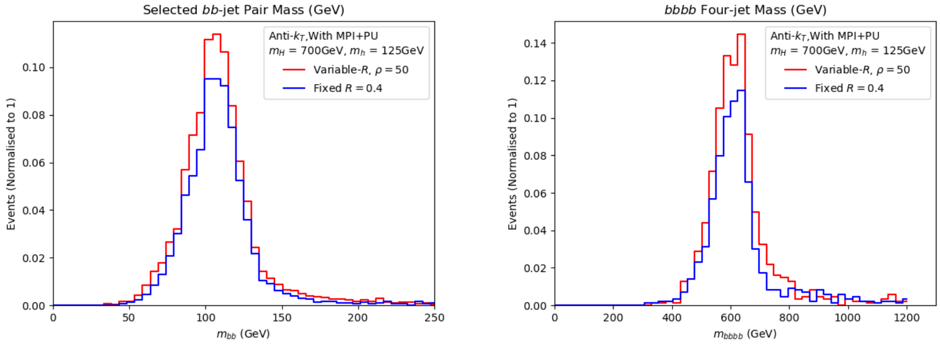


FIGURE 6.15: Left panel: The b-dijet invariant masses for BP1, using variable-R and fixed-R clustering, when considering the effect of pile-up and MPI. Right panel: The same for the four-b jet invariant mass.

As a final point, we note that a further pile-up mitigation technique is possible in variable-R, which is in the values chosen for the $R_{\text{min/max}}$ variables. Clearly if, for some particular process, one discovers that using a variable-R sweeps in too much extra ‘junk’ into the jets, a simple reduction of R_{max} is always possible.

6.4.5 Other Variable-R Studies

Before concluding, we review here some other studies from the literature utilising a variable-R reconstruction procedure.

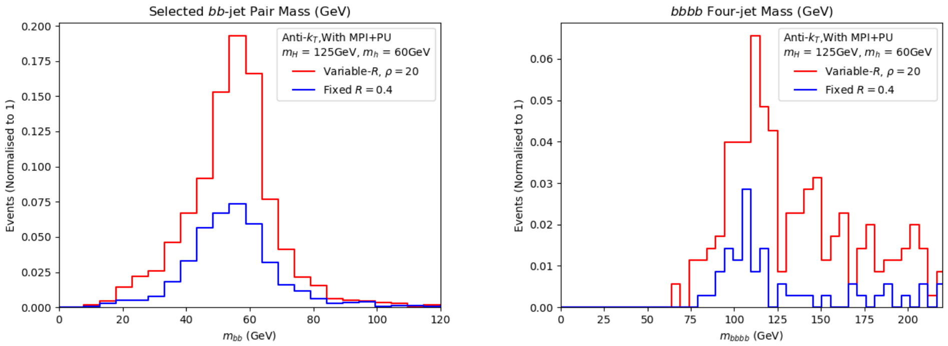


FIGURE 6.16: Left panel: The b-dijet invariant masses for BP2, using variable-R and fixed-R clustering, when considering the effect of pile-up and MPI. Right panel: The same for the four-b jet invariant mass.

We note that, while the leading b-jet has an R_{eff} roughly in line with expected values ($R_{\text{eff}} \simeq 0.5$), the lowest p_T b-jets have large cone sizes ($R_{\text{eff}} > 1.0$), risking potential contamination from additional radiation. This effect is discussed in [119]. We do not implement any vetoes to remedy this effect, yet, despite this, our results still suggest that the variable-R approach displays an improvement over traditional methods.

There have been other studies utilising variable-R methods for physics searches. For example, in highly boosted object tagging of $hh \to b\bar{b}b\bar{b}$ decays in [120]. Futhermore, in [119] mentioned above, a variable-R algorithm is deployed in the context of heavy particle decays. In both examples, an improvement over current fixed-R methods is present when using variable-R, which is in line with our findings.

As a final word on using variable-R jet reconstruction in experiments, we discuss its use in relation to b-tagging performance. In particular, the studies of Refs. [121, 122] explore the possibility of Higgs to b-jet tagging at ATLAS using variable-R techniques. Specifically, these studies deal with boosted topologies, focusing on fat b-jet substructure, so the validity of applying these techniques in a non-boosted regime is to be determined.

6.5 Conclusions

In this paper, we have assessed the potential scope of the LHC experiments (from mainly a theoretical perspective) in accessing BSM Higgs signals induced by cascade decays of the 125 GeV SM-like Higgs state discovered in July 2012, through the following prototypical production and decay channel: $gg, q\bar{q} \to H \to hh$, where H is the SM-like Higgs state and h is a lighter BSM Higgs state, with mass less than $m_H/2$, so as to induce resonant production and decay, thereby enhancing the overall rate. Any such a h boson, largely independently of the BSM construct hosting it, would decay to $b\bar{b}$ pairs, eventually leading to a four b-jet signature. The latter is extremely difficult to establish at the

LHC, owing to the substantial hadronic background. Therefore, b-tagging techniques are to be exploited in order to make such a signal visible. However, this poses the problem that the latter are most efficient at large transverse momentum of the b-jets, say at least 20 GeV, which in turn corresponds to a significant loss of signal events if the BSM Higgs mass is in the sub-60 GeV range. Hence, if one intends to maximise sensitivity to this benchmark signature of BSM physics, a thorough reassessment of the current Run 2 approaches is mandated for and especially so in view of the upcoming Run 3.

The first message we deliver is that, with current p_T cuts on final state b-jets, using a fixed-R jet reconstruction and tagging procedure, will lead to a poor signal visibility, with a majority of signal b-jets being lost. We instead presented a reduced cut-flow, based on existing $b\bar{b}\mu^+\mu^-$ analyses, and showed that this indeed provides a window onto $gg \to H \to hh \to bb\bar{b}\bar{b}$ signals with $m_H = 125$ GeV and $m_h < \frac{m_H}{2}$.

Additionally, and perhaps more remarkably, we also tested a variable-R reconstruction approach on events with this reduced cut-flow and showed a significant improvement in signal yield as well as signal-to-background rates. We notice that in final states of this kind, the signal b-jets have a wide range of p_T and hence varied spread of constituents. Using a fixed cone of a standard size ($R = 0.4$) constructs well higher p_T jets in an event but does not capture much of the wider angle radiation from lower p_T jets. This leads to two issues. Firstly, it will prove difficult to accurately construct m_h and m_H in the two- and four-jet invariant masses. Secondly, these jets will more often be lost due to kinematic cuts. A larger cone ($R = 0.8$), conversely, will gather up too much ‘junk’ in the higher p_T jets, which again will contaminate the signal.

We have obtained all of the above in presence of a sophisticated MC event simulation, based on exact scattering MEs, state-of-the-art parton shower, hadronisation and B-hadron decays as well as a simplified detector simulation. Given the results of our analysis, we recommend a more thorough detector level analysis is undertaken, for a variety of different high b-jet multiplicity scenarios, to explore whether a shift to variable-R jet clustering, on the one hand, could be implemented and, on the other hand, would improve upon current signal significance limitations using fixed-R jet reconstruction. In fact, while we have quantitatively based our case on the example of the 2HDM-II (wherein h was the lightest CP-even Higgs state, yet it could equally have been the CP-odd one, A), our procedure can identically be used in other BSM constructs featurings light (pseudo)scalar states emerging from decays of the SM-like Higgs state and in turn decaying into $b\bar{b}$ pairs.

Chapter 7

Jet Visualisation for New Higgs Boson Searches at the LHC

7.1 Introduction

The work presented in this chapter is yet unpublished, instead establishing a proof-of-concept for further research into an academic paper once more investigation has been completed. I am the lead author of the current progress presented here.

In recent years, the evolution in the field of machine learning has allowed physicists to test and explore novel approaches to problems pertaining to high energy physics. The areas of application for ML are plentiful, ranging from the exploration of parameter space for models beyond the standard model [88, 89, 90], improving jet tagging efficiencies [73, 74, 76, 77, 78, 79, 80], alternative jet clustering methodologies [2] and in the function of the detector itself [81, 82, 83, 84]. In the case of physics searches at the LHC involving jets, the application of more advanced deep learning techniques have been and continue to be explored, in particular the use of methods involving the mapping of the detector into an image.

The use of imaging is particularly attractive proposition - while a given model itself will generally be a rather complex multi-layered network, with a large number of inputs corresponding to the pixels of the image - implementing them to make classifications requires little to no preparation beyond the detector registering particle hits and potentially jet clustering and tagging. Simpler methods, such as BDTs, will generally be trained on variables extracted from a given jet (such as N-subjettinness), which can involve summations over all of its constituents.

Many of these studies - such as those tasked with jet identification and tagging - are largely agnostic to the specific physics from which they came, and as such can be applied to a variety of specific LHC searches. In fact a particular advantage of ML generally is the remarkable adaptability to adjacent problems. A particular model architecture can easily be trained on a new dataset (for example coming from a different physics process), without any changes to the model itself.

In this study, we investigate the potential of using jet substructure visualisation techniques to identify events coming from 2HDM-II models, by designing a convolutional neural network. It has been the long standing aim of the LHC to find evidence of new physics beyond the current standard model, and while a variety precision measurements have manifested tensions with standard model predictions, the golden target of a new particle discovery remains elusive. There are of course well established LHC searches for numerous processes of potential interest using more traditional techniques [32, 33, 34], as well as exploratory studies investigating alternative clustering algorithms for 2HDM-II final states [1]. We build upon the remarkable ability of image recognition informed jet taggers, to map information representative of an entire event to an image that can be used for training.

A particular area of interest occur for events containing boosted Higgses, such that the resulting decay products are enclosed at small angles into "fat jets" - which we introduced in the previous chapter on jet physics. For such boosted topologies, traditional jet clustering techniques can struggle to resolve the individual decay products from the boosted mother. It can in fact be more convenient to instead employ a large cone clustering to gather everything that came from the original Higgs into a single jet. This is a potentially convenient way of observing $H \to b\bar{b}$ decays, and have invariant masses directly matching the Higgs - however traditional jet tagging would require finding two b-tags for such jets, in order to verify it is indeed a combination of two slim b-jets. Double b-tagging can involve the use of complex, high-level features [123]. By imaging jets we present a method of peering into the structure of fat jets using low level information (i.e. detector hits), without the need to compute additional variables such as N-subjettiness.

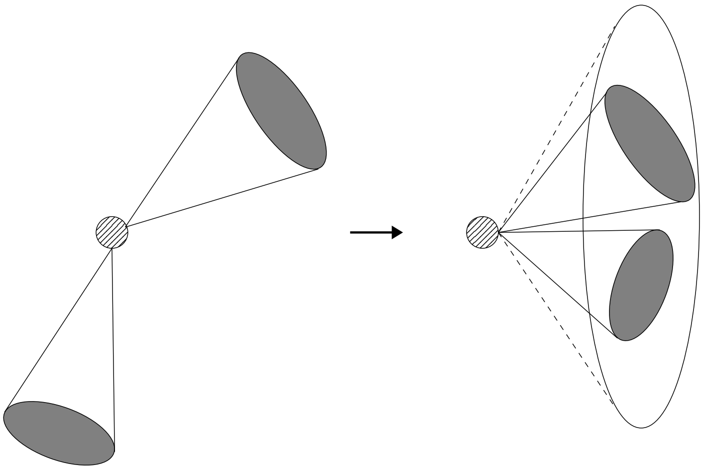


FIGURE 7.1: The merging of jets due into a single object due to a boosted decay.

For this study, we focus on 2HDM-II decays of heavy Higgses, into SM-like Higgs pairs, and subsequently into final states of four b-jets, i.e. $gg \to H \to hh \to bb\bar{b}\bar{b}$.

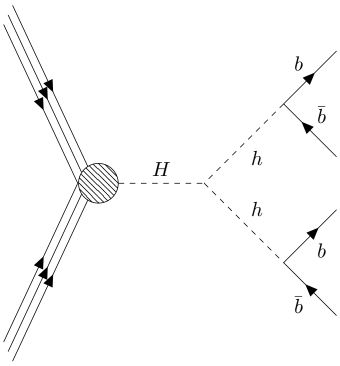


FIGURE 7.2: The 2HDM-II signal process of interest for this study.

7.2 Methodology

The approach to our study can be broken into three main stages - the generation of the physics events to train our classifier, mapping this data into a sample of images, and the training and tuning of the machine learning classifier itself.

7.2.1 Event Simulation

Firstly we select a particular 2HDM-II benchmark point and simulate signal (and background) events using a Monte-Carlo event generation tool kit comprised of `MadGraph v2.6.3.2` [111] and `Pythia v8.243` [112], with an analysis cutflow applied to the jets via `MadAnalysis v1.8.5`, with detector simulation in `Delphes v3.4.2` [105] interfaced with `FastJet` [100] for jet reconstruction. In order to work with a benchmark point that is not experimentally excluded[1], we test the benchmark against `2HDMC` [23], `HiggsSignals` [108], `HiggsBounds` [107] and `SuperISO` [109].

m_h (GeV)	m_H (GeV)	$\tan\beta$	$\sin(\beta-\alpha)$	m_{12}^2	BR($H \to hh$)	BR($h \to b\bar{b}$)	σ(pb)
125	700	1.5	0.999	1.8×10^{5}	3.231×10^{-2}	9.066×10^{-1}	3.542×10^{-1}

TABLE 7.1: The 2HDM-II parameters and cross sections of the $gg \to H \to hh \to bb\bar{b}\bar{b}$ process for the selected benchmark point for this study.

The selected benchmark model we use is shown in Tab.7.1. We are identifying the currently discovered Higgs boson as the lighter scalar in the 2HDM framework, h. We

---

[1] Although we recall here that the malleability of ML methodologies mean the results presented would be easily testable on both 2HDM-II models with different parameters, but also different models entirely (that result in the same final states).

generate a sample of $\mathcal{O}(10^5)$ events of the process $gg \to H \to hh \to b b \bar{b} \bar{b}$, with the tools described above.

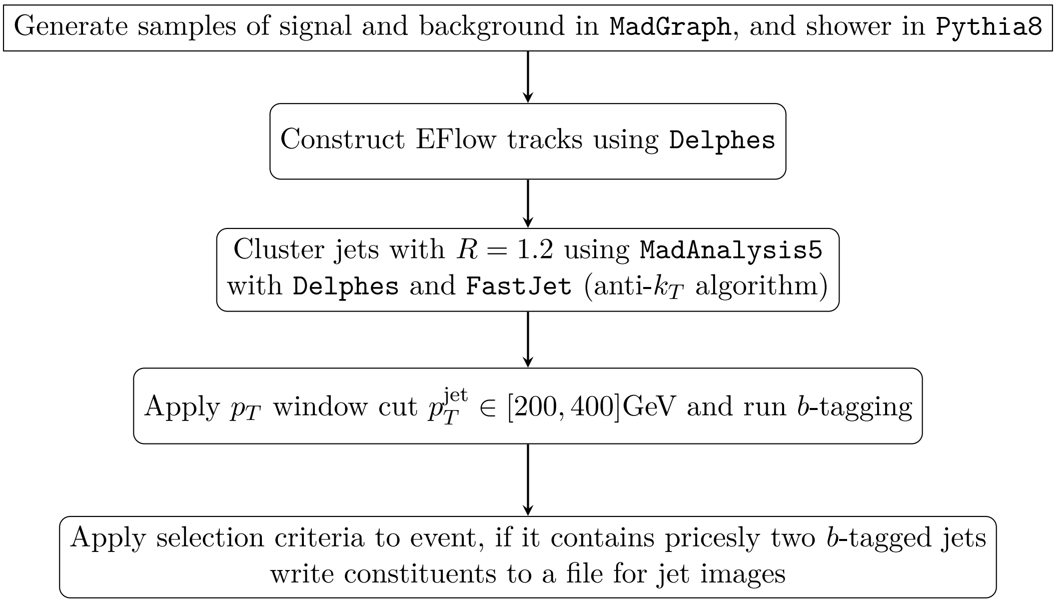
The background processes we consider are the following,

$$\begin{aligned} &pp \to t\bar{t},\ (t \to W^+ b,\ W^+ \to jj),\ (\bar{t} \to W^- \bar{b},\ W^- \to \text{all, all}) \\ &pp \to b b \bar{b} \bar{b} \\ &pp \to Z b \bar{b},\ Z \to b\bar{b}. \end{aligned} \tag{7.1}$$

Due to the kinematic differences between the signal process and leading backgrounds, we will limit the training of our CNN to jets within a specific p_T window of $200\text{GeV} < p_T^{\text{jet}} < 400\text{GeV}$. In order to achieve this, we therefore apply generation level cuts within `MadGraph`, given by

$$\begin{aligned} &pp \to t\bar{t};\ p_T^{\text{gen}}(t) > 250\text{GeV} \\ &pp \to b b \bar{b} \bar{b};\ p_T^{\text{gen}}(b) > 100\text{GeV} \\ &pp \to Z b \bar{b},\ Z \to b\bar{b};\ p_T^{\text{gen}}(Z) > 250\text{GeV}, p_T^{\text{gen}}(b) > 200\text{GeV}. \end{aligned} \tag{7.2}$$

in order to improve the selection efficiency at jet level. The showering and hadronisation is performed as usual via `Pythia8` interfaced with `MadGraph`. The basis constituents used in jet clustering are then EFlow objects from the detector simulation in `Delphes`, which are stored to form the jet images, where finally, an analysis cutflow selection is applied to all events before passing to the next phase.



Generate samples of signal and background in `MadGraph`, and shower in `Pythia8`

Construct EFlow tracks using `Delphes`

Cluster jets with $R = 1.2$ using `MadAnalysis5` with `Delphes` and `FastJet` (anti-k_T algorithm)

Apply p_T window cut $p_T^{\text{jet}} \in [200, 400]$GeV and run b-tagging

Apply selection criteria to event, if it contains pricesly two b-tagged jets write constituents to a file for jet images

Figure 7.3: Description of the procedure used to generate and analyse MC events in preparation for ML training.

7.2.2 Jet Image Generation

The next step is to then map the jets into images. For each of the processes, we have a sample of events containing precisely two b-jets, along with information on their EFlow constituents. Before fitting the constituents in to pixels, there are several preprocessing steps we perform in order to enhance the key substructure features.

7.2.2.1 Data Input

The input data used from the event generation are EFlow tracks, which source from charged and neutral hadrons, and photons in the detector simulated with `Delphes`. These contain kinematic information of each constituent track of each jet in the event, as well as the same information for the jet the constituent belongs, where in the previous step we have filtered events such that only those that contain two b-tagged jets. In particular, we have

$$n^{\text{event}},\ n^{\text{jet}},\ p_T^{\text{jet}},\ m^{\text{jet}},\ \eta^{\text{jet}},\ \phi^{\text{jet}},\ n^{\text{const}},\ p_T^{\text{const}},\ \eta^{\text{const}},\ \phi^{\text{const}}. \tag{7.3}$$

With the $(\eta, \phi)^{\text{const}}$ coordinates, we can map out the position of each constituent in detector space, and the jet information $(\eta, \phi)^{\text{jet}}$ enables us perform transformations on all constituents of jet (described below) wile preserving the relative structure of the jet.

By clustering the boosted Higgses into jets of large radius $R = 1.2$, we expect to gather all of the information of the event into two such jets, each with an invariant mass close to $m_h = 125\text{GeV}$. The key feature of these jets is therefore the presence of two subjets inside the $R = 1.2$ cone. As a method or preprocessing, we perform a translation and rotation on the constituents, such that the two subjets are found in the same regions across all images. This processing acts as a dimensionality reductions of sorts, we are removing the geometric variation in the distribution of the two subjets within the fat jet - the CNN should therefore in principle more easily ‘pick out’ the subjets and learn the substructure to classify Higgs jets.

7.2.2.2 Coordinate Transformations

The first step is the translation, which is done by locating the hardest constituent (c_1) in a given fat jet, ranked by p_T, and defining new coordinates (η', ϕ') such that the track c_1 is at the origin, and all other constituents of the jet are translated such that their relative positions with respect to c_1 are unchanged.

Now that the jet is centred on the origin in (η', ϕ'), we next look for the second hardest constituent $\bar{c_2}$ of the jet, if it is sufficiently separated from c_1 such that

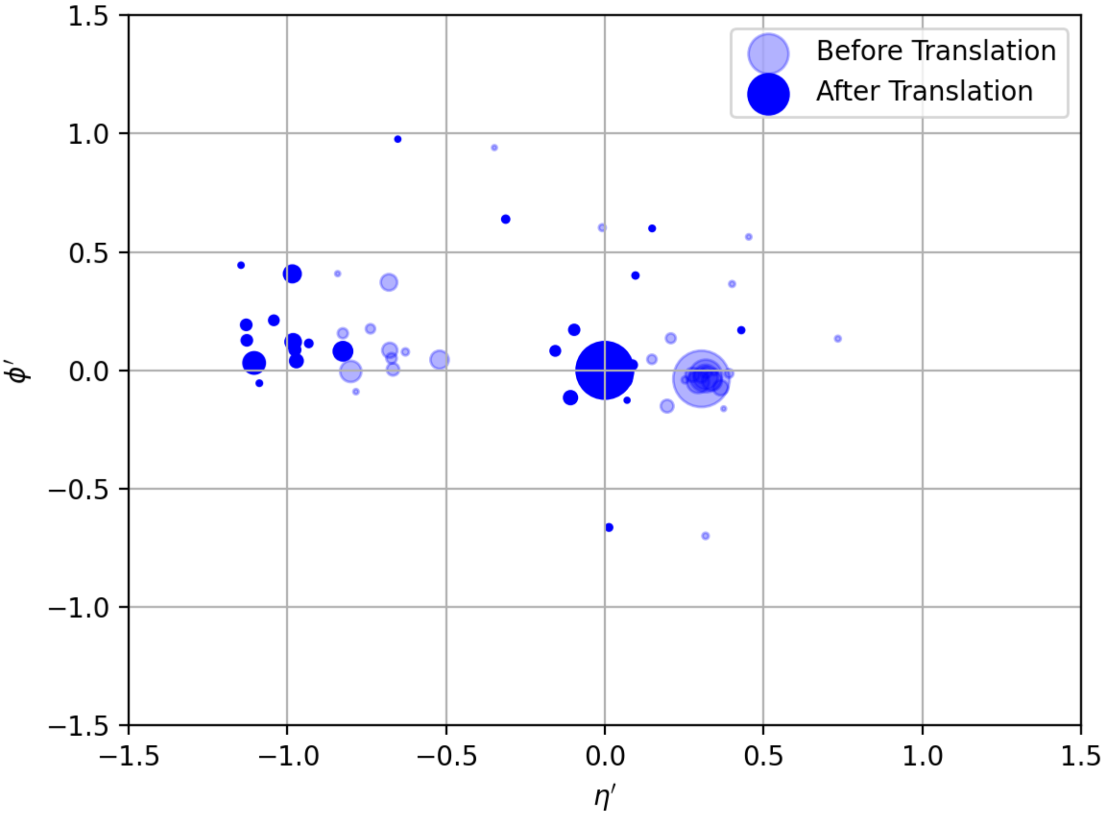


FIGURE 7.4: An example jet before and after the translation preprocessing step. The point sizes are weighted by p_T, to demonstrate that the highest p_T particle is translated to the origin in (η', ϕ').

$$\Delta R(c_1, \bar{c}_2) > 0.2, \tag{7.4}$$

then we label $\bar{c}_2$ as c_2, if it is separated from c_1 by an angular distance smaller than 0.2 we look for the next hardest constituent and compute the distance to c_1, until Eq. 7.4 is satisfied. Once c_2 has been located, we compute the polar coordinates of c_2 from (η', ϕ'), i.e. (r', θ'), and compute the angle required to rotate c_2 to $\theta' = \frac{3\pi}{2}$, denoted here by $\Delta\theta$. All of the jet constituents are then rotated by $\Delta\theta$ around the hard jet centre constituent at the origin.

We note here that our method in transforming the jets differs from other studies - which tend to rely on reclustering a fat jet with a smaller cone size in order to define subjet candidates, and translate/rotate based on the leading subjets. We propose a simpler, less time consuming method which we show can still lead to images that can be classified to a high level. Additionally, we open the door to the potential for deep learning classification methods that do not require clustering at any stage, even in to the initial fat jets which are put into images. We will discuss the prospect of such methods in the final section.

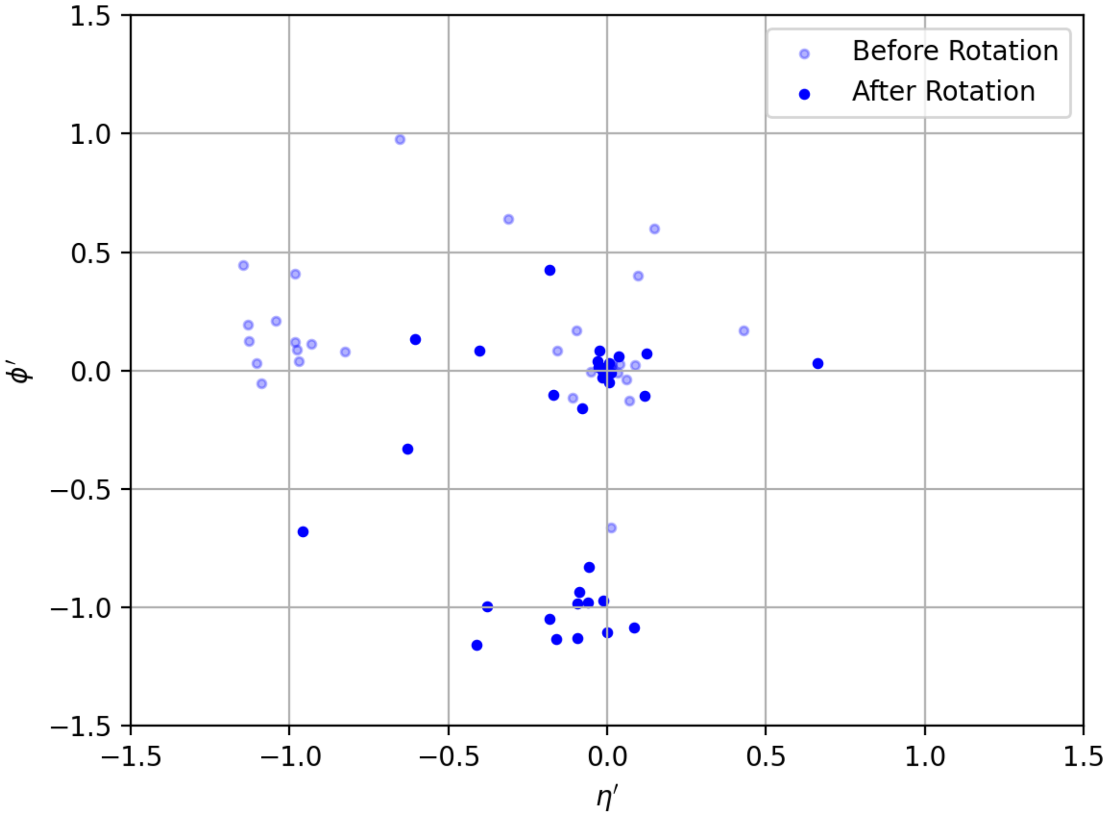


FIGURE 7.5: An example jet before and after the rotation preprocessing step.

7.2.2.3 Flipping The Image

Finally, we perform a simple flip about the $\eta' = 0$ axis, such that the half of the image containing the highest total p_T is always in the region $\eta' < 0$. We will see that certain processes contain a larger spread of radiation away from the centre of each subjet, and so by gathering everything to one side we are highlighting this wide angle radiation. This is particularly useful for visualising the images (see section. 7.2.3), as the wide cone radiation is soft compared with the subjet centres, so gathering onto the same region of the image will increase the cumulative p_T for generating images over a large sample of jets.

7.2.2.4 Binning and Stacking

With the constituents translated, rotated and flipped into place, we can finally apply the last step of preprocessing, which is to group the jet constituents together into 'pixels'. We use a (0.1×0.1) grid in (η, ϕ) space, and bin constituents weighted by their p_T into two dimensional bins, plotting over the range $(\eta' \in [-1.5, 1.5], \phi' \in [-1.5, 1.5])$. The resulting image for a given jet is therefore a (30×30) pixel image, with an intensity scaling with the cumulative sum of p_Ts of all constituents in a given (0.1×0.1) pixel.

However, the we aim to classify these images at an event level, rather than jet level, and we recall that the events in our selection contain two b-jets. To build a model that takes an input representative of the entire event, we stack the jet images for the two fat jets in an event, i.e. assign two values to each pixel in the image, corresponding to each of the two jets. The input for the CNN therefore are $(30 \times 30 \times 2)$ pixel images, represented as arrays.

7.2.3 Average Jet Images

Having laid out the procedure for forming jet images, we now move on to looking at the features of the signal and background images, which we hope to train our network on. It is not so useful to flick through a gallery of $\mathcal{O}(10^4)$ images, and so instead we present an average jet image of a large sample of N jets. The values for each pixel in the N jet average image is simply the cumulative sum of that pixel over all N jets, divided by N.

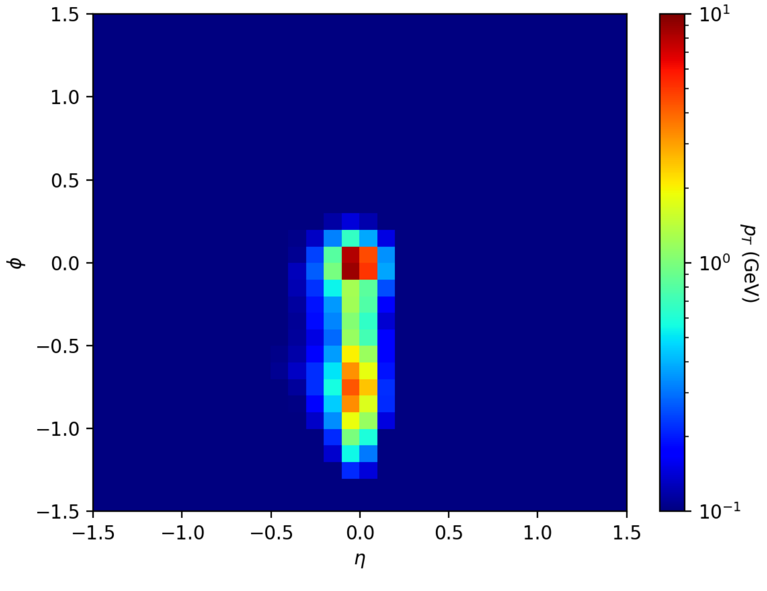


FIGURE 7.6: The average image for a sample of 10,000 fat b-jets coming from the signal process $H \to hh \to b\bar{b}b\bar{b}$.

We can see in Figs. 7.6 - 7.9 the general substructure in the wide cone jets from each process. For the signal, we observe a clear two-subjet structure, with little radiation 'spilling' outside of the hard central cores of each subjet. Visually, the $t\bar{t}$ background appears to be the background most closely resembling the signal, again displaying two distinct subjets, however due to the nature of the $t/\bar{t}$ decays, there is a more complex substructure associated with the combination of b-jets and jets from the $W^{\pm}$ bosons.

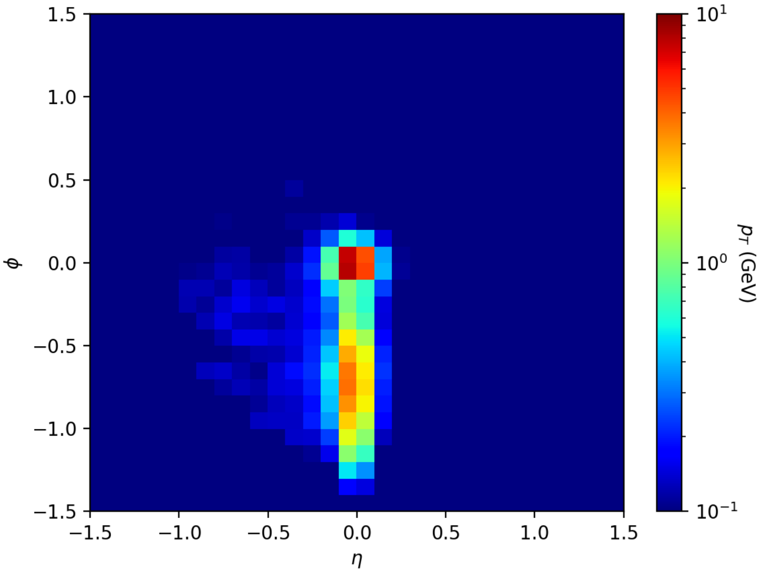


FIGURE 7.7: The average image for a sample of $10,000$ fat b-jets coming from the background process $pp \to t\bar{t}$.

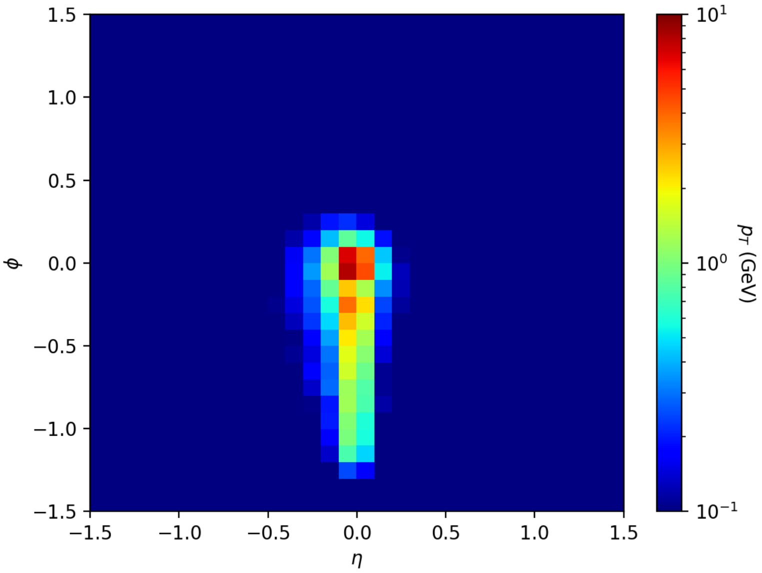


FIGURE 7.8: The average image for a sample of $10,000$ fat b-jets coming from the background process $pp \to bb\bar{b}\bar{b}$.

This is where the motivation of the flip stage of the preprocessing becomes apparent, we are gathering all of the additional radiation from nearby truth jets (i.e. not directly from the b's) into one half of the image.

We can see in both the $pp \to bb\bar{b}\bar{b}$ and $pp \to Zb\bar{b}$ backgrounds that there is not as strong a substructure in the fat jets. For the former, as the b-jets are coming from QCD

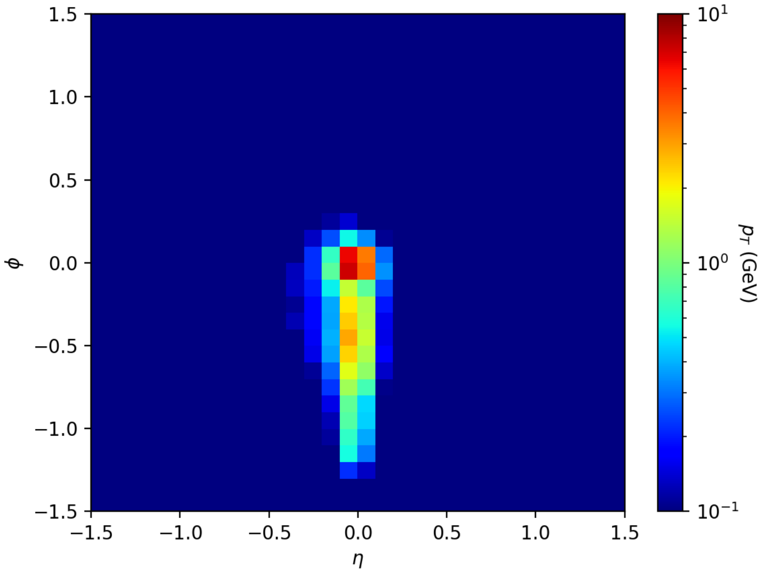


FIGURE 7.9: The average image for a sample of 10,000 fat b-jets coming from the background process $pp \to Zb\bar{b},\ Z \to b\bar{b}$.

processes, there is no boosting and hence it is we do not generally observe two b-jets collinear with one another. For the $Zb\bar{b}$ case, there is a hint of a second subjet, however this is softer than in the signal and $t\bar{t}$ cases, owing to the imbalance in the two types of fat jets from this process. Clearly we will expect some fat boosted objetcs from the Z, however in our images where fat jets are read in pairs from a given event, this will be countered by the remaining $b\bar{b}$ pair produced with the Z, which we do not expect to be boosted (and hence merged into a single object with clear substructure).

7.2.4 Machine Learning Training

Having taken a qualitative look at the jets we are training on, we now move on to describing the CNN architecture used to train the classifier model. What we present here is to be considered an early prototype for a CNN model - used to establish the feasibility of classifying the signal from a mixed sample of background images.

The toy model we implement here is comprised of the following;

- A (3×3) 2D convolutional layer
- A (4×4) max pooling layer
- A (5×5) 2D convolutional layer
- A (2×2) max pooling layer

- A 0.5 dropout layer
- A flattening layer to length 128
- A 0.4 dropout layer
- A length 32 dense layer
- A 0.25 dropout layer
- A length 10 dense layer
- A 0.25 dropout layer
- A 2 node output layer.

We have seen the role of pooling and convolution layers in extracting information from the images. The flattening allows us to feed image data into a fully connected MLP like structure, and dropouts represent a random sampling where certain nodes are dropped - which prevents the model from overfitting too much. Finally the two node output layer represents the probability for classifying an image as signal (1) or background (0).

7.3 Preliminary Analysis and Results

Here we present early stage performance of the model described in section 7.2.4. To visualise the training for the model, we plot the evolution of the loss and accuracy, both on the training and validation data sets.

We then check the output score for all of the images in the validation set, which we recall is a score between zero and unity, and interpreted as a probability for the output of a given instance coming from the signal process. If this is close to one the model is sure it came from the signal process, close to zero and the model is confident it is instead a background.

Finally, we plot a receiving operator characteristic (ROC) plot, and compute the area under the curve (AUC), which gives a convenient numerical scoring between zero and unity.

In Fig. 7.10 we can see the model learning, and slowly plateauing off at a peak performance. We observe that the accuracy and loss on the validation sets tracks the training sets fairly closely, such that we can be confident the model is not overfitting on the training set - and therefore should generalise well to new, unseen images.

As mentioned, the model output for a particular event is a measure of the model's estimated probability for that event belonging to the signal set. To see how well the

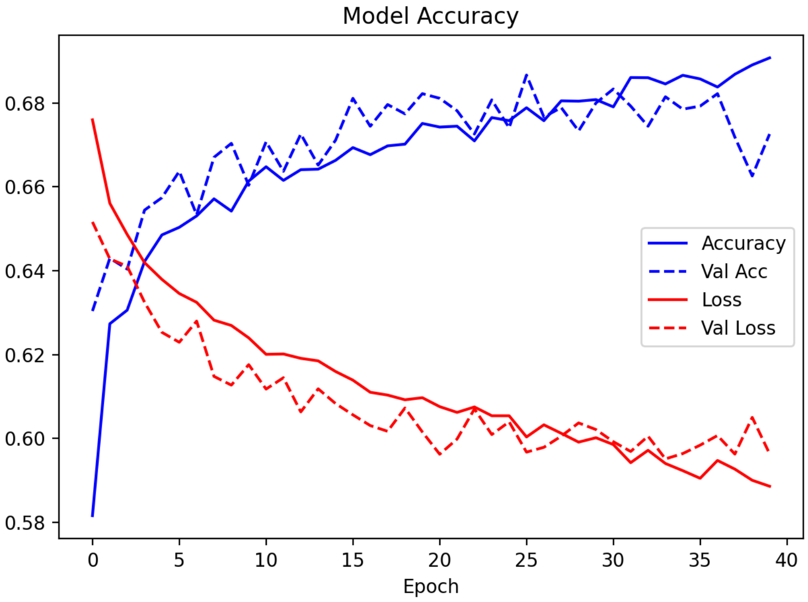


FIGURE 7.10: The evolution of the loss and accuracy, across 40 training epochs.

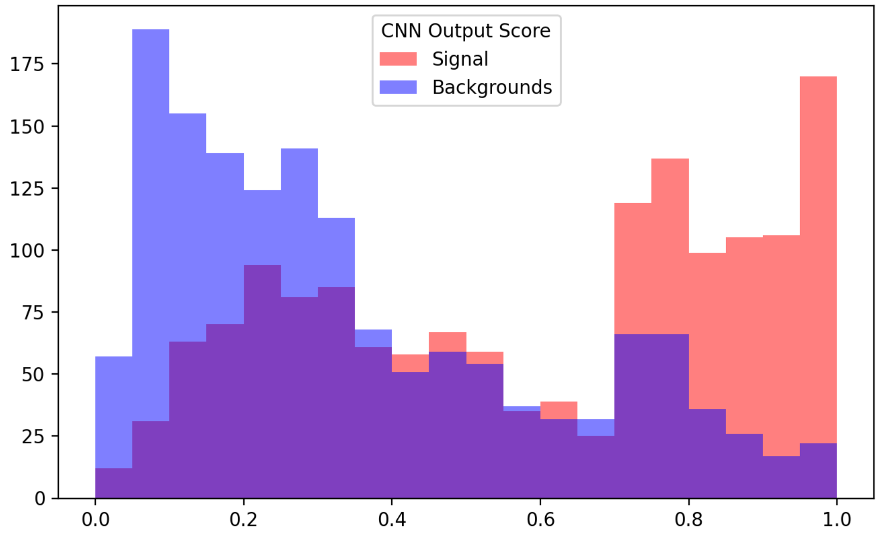


FIGURE 7.11: The model output for the unseen images in the validation set.

model is doing, we separate out the signal and background datasets to see how closely the prediction matches the truth information. We can see that generally background events are predicted values closer to zero, with a large peak at around 0.1, and likewise the signal generally predicts higher values, indeed with a spike at 1.0. There is however some significant overlap, suggesting there are a large number of events the network struggles to classify - and hence that further improvements are desirable.

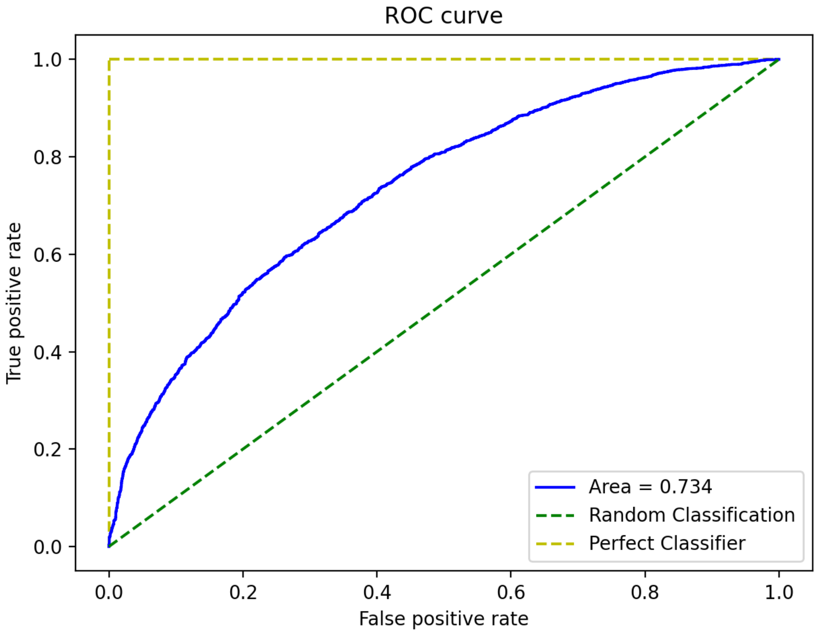


FIGURE 7.12: The ROC curve, with ROC AUC plotted, to demonstrate the performance of the final iteration of the model in training.

Finally, we present the ROC curve for samples in the validation set. This plots the true positive rate, given by

$$TPR = \frac{TP}{P} \tag{7.5}$$

against the false positive rate

$$FPR = \frac{FP}{N}, \tag{7.6}$$

where each point on the curve represents the model with a different decision boundary. We can see that this model, with an appropriately placed decision boundary, can achieve a TPR of around 0.7, with a background rejection of under 0.4.

7.4 Comments and Future Work

We have presented a prototype for a jet visualisation CNN algorithm capable of classifying events from 2HDM-II $H \to hh \to bb\bar{b}\bar{b}$ processes from the leading backgrounds, $pp \to t\bar{t}$, $pp \to bb\bar{b}\bar{b}$ and $pp \to Zb\bar{b}$. In contrast with other studies on jet image machine learning, we make use of a stacking method in order to display information representative of the entire event in a single image.

We demonstrate a simple toy model CNN that can achieve a ROC AUC of up to 0.74 when trained for the signal against an equal mixture of the leading backgrounds - which motivates further investigation to implement a more robust network.

We now outline potential improvements to the model which will for the basis of future work in designing a final iteration of the network with optimised performance. As mentioned, for the basis of this study we aimed to limit the inputs to low level information in the form of detector deposits. It is possible however to incorporate additional information into the model like in [124]. We note in our model, we are losing global information on the distribution of the jets within the detector, such as the distance ΔR between the two fat jets. This could be included in the model via a series of additional features, or in keeping with the goal of using only low level information, a second stream could be added to the neural network with an event image covering more of the detector, which would capture the distribution of radiation at a global event level.

In our treatment of backgrounds, we have used a simplified 50/50 split of data between signal and background, with each background comprising 16.6% of the total data (i.e. on third of the background data). This split is made with the machine learning aspect of the problem in mind - that is a binary model being trained on data that is not evenly split will be biased to the class with most samples. This is not the reality of the underlying physics, indeed each process has an associated cross section and hence different event rates. Despite this, once could proceed training a model in the way we present, and evaluate on a realistic spread of the different processes weighted by the corresponding cross sections, or alternatively depending on the exact rates apply a slightly different ML approach, such as anomaly detection, if the signal process is sufficiently rare when compared with the backgrounds.

Again related to the physics, the precise implementation of the classifier can also be adjusted. Here we apply a short cutflow in the name of simplicity, by filtering events that contain two b-tagged jets with a $p_T \in [200, 400]$GeV. It would also be possible create a version of our model that can be applied after a more intensive selection process, for example incorporating masses of the fat jets to reduce the background rates - which would also go some way in addressing the points in the previous paragraph. Even if the original background rates are larger than the signal, a sensible cutflow might go some way in reducing the backgrounds to a similar level to the signal - and there active efforts investigating this, informed by experimental cutflows.

Another point relating to the grouping of this background data is that we have combined all backgrounds into a single class - despite the fact that we have demonstrated each background has different features that would distinguish them from the signal. This choice is again made with the ML implementation in mind, it is more straight forward to train and evaluate a binary classifier. While this is a convenient simplification, more investigation could be done to investigate the separation power between the signal and each background individually.

We have seen that the architecture implemented here has been designed by hand as a toy model. Once issues pertaining to the input and physics have been addressed, a final model would of course be designed by explicitly scanning over the tunable hyperparameters in order to maximise the classifying performance.

Chapter 8

Spectral Clustering for Jet Physics

8.1 Introduction

We note that the work presented here is based on [2], of which I am a co-author. My role in this research was in qualitative support of the physics and algorithm, as well as informing on methods of evaluating the performance of the algorithm.

We have investigated the various uses of machine learning in particle physics, and in presented a study using supervised deep learning to classify between events arising from different processes. In this chapter, we present research on a more fundamental area of high energy physics - that is in the jet clustering algorithm itself. Current choices for algorithms are the anti-k_T [48] and Cambridge-Aachen [49] algorithms, which are examples of sequential recombination algorithms.

These algorithms have been wildly successful for a variety of reasons, for example they are resistant to clustering events with a different number of jets (unlike cone algorithms), are IR safe, and are fast. Despite this, as computational resources increase, we can look toward alternative methods, and compare these with the traditional ones. The process of jet clustering is actually quite simple, the input is a series of particle coordinates, defined in some space on (η, ϕ) in a detector, and the output jets are essentially just groupings of these initial coordinates. It is quite clear that this can be recast as an unsupervised ML problem, we do not a priori know how many jets we should be clustering into, nor is there any information on the “correct” ways to organise the inputs into groups.

In this study we consider the use of spectral clustering (SC) machine learning on the problem of jet reconstruction. As we have seen, spectral clustering is powerful unsupervised ML technique, that can significantly outperform more basic methods such as k-means clustering. In transforming input data into a new space, one has control over the definition of similarity used - it can therefore be seen how spectral clustering can go beyond simple clustering based on Euclidean distances in parameter space. This freedom is highly suggestive that spectral clustering might be suited to jet reconstruction - one can incorporate information beyond just the detector geometry, such as p_T or object type (charged track, neutral track, photon etc), which could be useful in jet reconstruction.

We begin by reviewing the methodology behind the formation and training of our SC algorithm, in particular highlighting the modifications required to apply SC to jet physics. We then review the toolkit used to simulate physics events, which we test our algorithm on, as well as describing the physics processes we are testing. Following this, the precise steps for forming our algorithm are outlined, and finally we present the results of SC compared with tradition jet algorithms.

8.2 Methodology

We have seen the procedure for finding the Laplacian L which forms the basis of SC Here we will outline how we go about adapting this technique to cluster particles into jets.

8.2.1 Modifications to Spectral Clustering

We have that, as with other unsupervised ML techniques, spectral clustering requires the user to input the expected number of groupings before any clustering has taken place - this is a clear drawback when applying to jet physics - however we propose a modified algorithm which does not require an input for the number of final state jets.

Spectral clustering works by computing a Laplacian matrix L from the graph of input points to be clustered. We can then compute the minimum eigenvalues λ_{min}, each of which corresponds to a grouping of points. In the generalised method, the number of these eigenvalues we chose determines the number of clusterings to partition the data into - which is also the dimensionality of the embedding space c. To circumvent the issue of not knowing a priori the number of expected final state jets, we do not fix the number of eigenvalues we compute, and instead compute all eigenvalues smaller than a cut-off λ_{cut}.

Another modification we must make to the usual spectral clustering methodology is in the implementation of a stopping condition. In traditional jet algorithms this is well defined in that points are combined and sequentially removed until no more remain, at which point all particles have been clustered. An intuitive choice might be the distance between the two points most recently combined, however it turns out this varies significantly between events, and that taking the mean distance between points for all combinations is a more stable choice.

Finally, as the dimensions of the embedding space are not of equal importance in a clustering (in contrast with the intuition for geometrically clustering), we normalise the eigenvectors of the Laplacian by the corresponding eigenvalue to some power β. For a particular eigenvector satisfying $\lambda < \lambda_{\mathrm{cut}}$, which can be written as

$$\sum_j L_{ij} h_j = \lambda h_i, \tag{8.1}$$

where j iterates over points in the sample. The coordinate for a point j in the embedding space will therefore be

$$m_j = \left(\lambda_1^{-\beta} h_{1j}, ..., \lambda_c^{-\beta} h_{cj} \right), \tag{8.2}$$

i.e. the component of a vector in the n^{th} dimension are suppressed by a factor of λ_n^{β}.

8.2.2 Coordinate System

The coordinate system we use differs slightly to that used in the previous chapter. We keep the same definitions of the beam line z direction, and angular azimuthal coordinate ϕ, however in place of the pseudorapidity η we have seen before, we use the rapidity defined by

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{8.3}$$

We see in Fig. 8.1 a demonstration of the same event, plotting in the physical coordinate system in a physics detector, and in the embedding space from our implementation of spectral clustering.

8.2.3 Physics Simulation

Of course in order to test our algorithm, we require some physics data to cluster in to jets. To generate event samples, we use a simple tool kit of `MadGraph` interfaced with `Pythia` to produces showered and hadronised particles. Instead of using a full detector simulation, we apply an approximation in the form of a cutflow, which are tailored to each process.

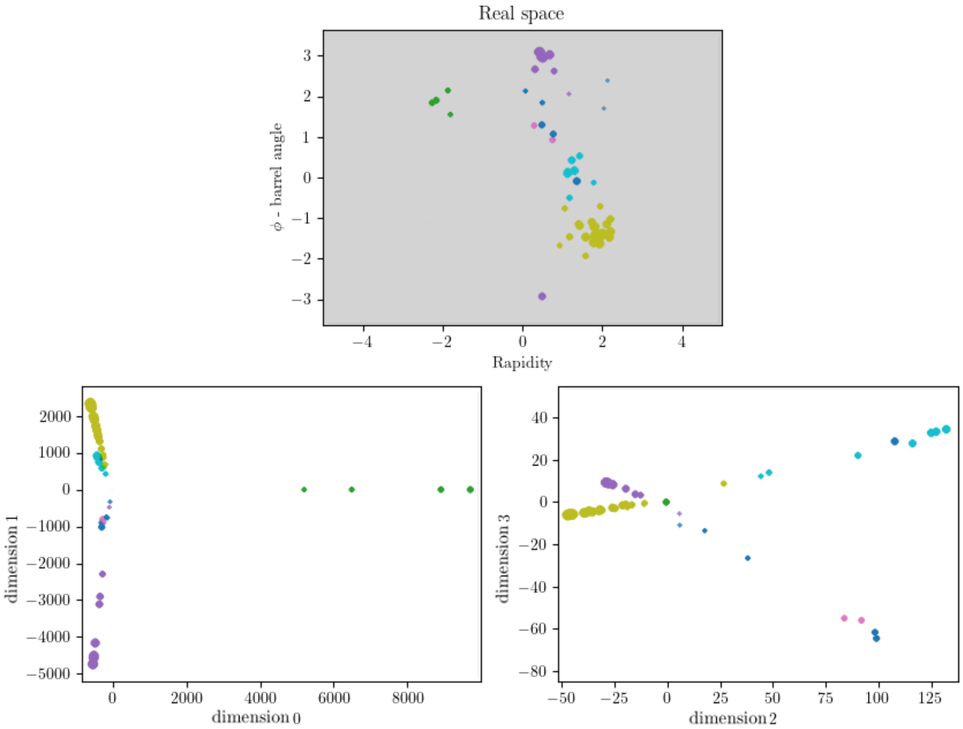


FIGURE 8.1: An example event, plotted in (upper) the detector space (y, ϕ), and (lower) the first four dimensions in the embedding space.

We consider four different scenarios, two of which arise from 2HDM-II models, where we simulate the decay process

$$pp \to H \to hh \to bb\bar{b}\bar{b}, \tag{8.4}$$

as seen in [1]. In the first of these benchmarks, we identify H as the SM-like Higgs boson with $m_H = 125\text{GeV}$, with the lighter scalar Higgs mass $m_h = 40\text{GeV}$ - this point is referred to as the light Higgs (LH) process. In the second benchmark, which we refer to as heavy Higgs (HH), we identify $m_h = 125\text{GeV}$, and select a heavier $m_H = 500\text{GeV}$ scalar partner.

The next two we consider are SM processes, firstly a top quark process (which we refer to as TT)

$$pp \to t\bar{t}, (t \to W^+b, W^+ \to jj), (\bar{t} \to W^-\bar{b}, W^- \to l\nu_l), \tag{8.5}$$

where $l = e, \mu$. Finally, we simulate three jet events when testing the IR safety of the algorithm,

$$pp \to jjj, \tag{8.6}$$

referred to as 3J. The cutflow used to emulate a simple detector are outlined below in Tab. 8.1.

Process	p_T^{particle} (GeV)	η^{particle}	p_T^{jet}	η^{jet}	p_T^{miss} (GeV)	p_T^{lepton} (GeV)	η^{lepton}	ΔR(lepton,jet)
LH	>0.5	<2.5	>15	-	-	-	-	-
HH	>0.5	<2.5	>30	-	-	-	-	-
TT	>0.5	<2.5	>30	<2.4	>50	>55 or >120*	<2.4	>0.4$^\triangle$
3J	-	<2.5	-	-	-	-	-	-

TABLE 8.1: The cutflow applied to each of the processes considered in this paper, where cuts for the TT process are taken from [125].

In the cutflow for the TT process, we note the following, ($\star$) in the lepton p_T cuts, if the lepton is a μ then a p_T of 55GeV is required, if it is an electron then either p_T>55GeV is required if the electron is isolated [126], or p_T>120GeV if it is not isolated. Furthermore ($\triangle$), the cut on separation between the jet and lepton is interchangeable with a cut on the p_T difference ($|p_T^{\text{jet}} - p_T^{\text{lepton}}| > 40$GeV).

In the LH and HH processes, to analyse how well our clustering algorithm is performing we test the invariant b dijet mass and four b-jet masses against the respective model values for m_h and m_H. For the TT events, there are three masses that can be reconstructed, the W mass from the $W^+ \to jj$ decays, the hadronic top mass can be reconstructed from the b-jet along with the W^+ remnant jets, and finally the leptonic top can be reconstructed from the combination of the resulting b-jet along with the missing momentum.

As a simplified method of jet tagging, we use MC truth information. For example for b-jets, we take each truth b-quark in the event sample, and match it to the closest jet, i.e. the one that minimises

$$\Delta R(b, \text{jet}) = \sqrt{(y_b - y_{\text{jet}})^2 + (\phi_b - \phi_{\text{jet}})^2}, \tag{8.7}$$

if $\Delta R(b, \text{jet}) < 0.8$, the tag is kept, otherwise no tag is assigned. The same procedure is followed for other jet flavours, replacing the b-quark with the relevant MC particle.

8.2.4 Spectral Clustering For Jets

We now outline the spectral clustering algorithm designed for reconstructing jets. The input for our algorithm are a series of final state particles with the cutflow applied defined in Tab. 8.1. The particles in a given event are read as pseudojets, randomly ordered. The algorithm proceeds in time steps, recursively combining pseudojet pairs via the so-called E scheme, where a new combined pseodojet is formed by summing the momenta, i.e. for a joining of pseudojets i and j into k

$$p_k(t+1) = p_i(t) + p_j(t). \tag{8.8}$$

In order to select two psuedojets i and j to join, we compute affinities between the nodes (particles), which form a measure of the likelihood two particles belong in the same jet. A simple basis for this is of course geometric distance, which in our coordinate system (y, ϕ) is defined as

$$d_{ij} = \sqrt{(y_i - y_j)^2 - (\phi_i - \phi_j)^2}. \tag{8.9}$$

Clearly one would expect the affinity to be larger for pseudojets that are close together, the affinity measure we use is given by

$$a_{ij} = e^{-d_{ij}^{\alpha}/\sigma_v}. \tag{8.10}$$

We note that affinity measure will change for each time step t, however this dependence is suppressed in the above. Points separated by a large distance are unlikely to be good candidates for joining, in order to reduce noise, for each pseudojet, we therefore only keep a select number of neighbours, k_{NN}, setting sufficiently low affinity pairings to zero. These affinities are used to compute the symmetric normalised Laplacian L, the entries of which will be

$$L_{ij} = \begin{cases} \propto -a_{ij} \text{ for } i \neq j \\ 1 \text{ for } i = j. \end{cases} \tag{8.11}$$

We define the matrices

$$D_{ij} = \delta_{ij} \sum_k a_{ik}; \; A_{ij} = (1 - \delta_{ij}) a_{ij}, \tag{8.12}$$

such that the Laplacian can be initially written

$$L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}}, \tag{8.13}$$

where we again note that everything here has a time dependence, for each step in the clustering process the Laplacian will change as we combine particles. We introduce the quantities $W_{ij}(t) = \delta_{ij} w(t)_i$ and $Z_{ij}(t) = \delta_{ij}(t)$, where $z_i(t)$ is the sum of all affinities connected to a particle i, and $w_i(t)$ measures the weighting of the contribution of pseudojet i to a cluster. We can then write the Laplacian as

$$L(t) = W(t)^{-\frac{1}{2}}(Z(t) - A(t))W(t)^{-\frac{1}{2}}, \tag{8.14}$$

where the time dependence is shown explicitly. At each time step, when two pseudojets are combined, L will reduce in size by a row and a column. The eigenvectors of L are then

$$Lh_k = \lambda_k h_k, \tag{8.15}$$

where $k = 1, ..., c$ labels the number of eigenvectors, which are used to define the pseudojets in the embedding space. As discussed, to capture the flexibility for desired number of jets, we compute all non trivial eigenvectors whose eigenvalues satisfy

$$\lambda_k < \lambda_{\mathrm{cut}}. \tag{8.16}$$

Eigenvectors are rescaled by their corresponding eigenvalue raised to the power of β, and to circumvent potential issues arising from zero division errors (for small eigenvalues), we apply the following function to the eignenvalues

$$\lambda_k = \min(\lambda_k, 0.001). \tag{8.17}$$

With the eigenvectors we can construct the points in the embedding space, the coordinates of which for the j^{th} particle will be

$$m_j(t) = \left(\lambda_1(t)^{-\beta}h_{1j}(t), ..., \lambda_c(t)^{-\beta}h_{cj}(t)\right), \tag{8.18}$$

where here the time dependence is shown explicitly. In order to cluster the points, we define a distance measure in the embedding space,

$$\sqrt{d_{ij}(t)} = \cos^{-1}\left(\frac{m_i(t) \cdot m_j(t)}{||m_i(t)||\ ||m_j(t)||}\right), \tag{8.19}$$

where $||m_i(t)||$ is the Euclidean length of the vector m_i at a time t. Finally, as discussed in section 8.2.1, we implement a stopping condition, inspired by traditional clustering algorithms, based on the input distance parameter R [1] and mean distance between each pseudojet pairing. While the condition

---

[1] Unlike in traditional clustering, R is in fact optimised as part of the model training rather than being chosen by the user.

$$\frac{2}{c(c-1)} \sum_{i \neq j} \sqrt{d_{ij}(t)} < R \tag{8.20}$$

is satisfied, we continue clustering pseudojet pairs with the minimum $d_{ij}(t)$. Once this is broken, i.e.

$$\frac{2}{c(c-1)} \sum_{i \neq j} \sqrt{d_{ij}(t)} > R, \tag{8.21}$$

all pseudojets are relabelled as jets, at which point jet level cuts can be applied. In our implementation, we apply the cut that, if a given jet has fewer than two constituents, it is removed and the particles clustered into the jet are considered noise.

8.2.5 Model Parameters

In the procedure for our spectral jet clustering algorithm, we defined a number of parameters. As we are now in the realm of machine learning, these are actually parameters which are tuned in the training of the model. We note that in many ML methodologies, the model parameters can be rather abstract and hence difficult to interpret physically - in the case of our spectral clustering algorithm this is not the case - we outline the parameters and their interpretation below.

In the affinity measure defined in the previous section,

$$a_{ij} = e^{-d_{ij}^{\alpha}/\sigma_v}, \tag{8.22}$$

we introduced the quantities α and σ_v. The first of these parameters α defines the shape of the distribution that describes nearby particles around a given point, with $\alpha = 2$ corresponding to a Gaussian kernel. We also have σ_v, which estimates the distance scale we expect between particles arising from the same shower (and hence belonging to the same jet). It is analogous to the input parameter in the generalised k_T jet algorithm, here referred to as R_{k_T}, and will therefore generally take on similar values. We note that α and σ_v are related, in that an increase to the parameter α leads to a reduction of the probability of joining particles with a separation exceeding σ_v.

When computing affinities between pseudojets, recall that we replace sufficiently small affinities with zero, such that only a fixed number of neighbours, k_{NN}, are kept. Reducing the number of neighbours kept acts as a way of reducing noise, however values that are too small will lead to a loss of signal.

To create a variety of spectral clustering algorithm that does not require prior knowledge of the desired final number of clusters, we introduce a way of computing a variable number by considering those with corresponding eigenvalues lower than some cut off λ_{cut}. Increasing λ_{cut} will therefore increase the number of dimensions of the embedding space - although not by any measurable amount - as the effect will vary event by event.

In order to properly weight each dimension in the embedding space, we normalise each eigenvector with its eigenvalue raised to the power of the parameter β. This rescaling suppresses eigenvectors with larger eigenvalues - which contain lower quality information.

Finally, in our addition of a stopping condition we implement the parameter R, that acts as a cut-off mean distance between pseudojets in the sample, that when exceeded halts the clustering and declares jets.

8.2.6 Parameter Optimisation

To determine a good set of the above parameters, we perform scans on a sample of 2000 MC events. We test the quality of a given set of hyperparameters by comparing the constituents of a clustered jet against the MC truth particles that came from the decaying quark. For a given parton, such as a b-quark from which we are trying to construct a b-jet, we refer to decay products originating from the b as descendants. A similar study utilising MC truth information can be found in [127].

It can therefore be deduced that, for some parton jet, there is a sample of particles that are descendants of the original parton, and so a perfect classifier would gather precisely this sample of particles into the jet corresponding to that parton. Conversely, a realistic algorithm will make errors arising from two different sources, namely

- Particles that arose from that parton not being included in the corresponding jet leading to a reduction in the jet mass - this is referred to as **signal mass lost $\mathcal{S}$**. In standard jet clustering this would occur if the chosen input parameter R is too small.
- Particles that did not come from the parton in question, being drawn in to the jet, leading to an artificial increase in jet mass - we refer to this as **background contamination $\mathcal{B}$**. For standard clustering, we would expect contamination when we select an R that is too large.

Using the above, we construct a loss function that is used to evaluate the performance of the algorithm for a given set of hyperparameters,

$$\mathcal{L} = \sqrt{w\mathcal{B}^2 + \mathcal{S}^2}, \tag{8.23}$$

where we apply a weighting $w = 0.73$ to $\mathcal{B}$.

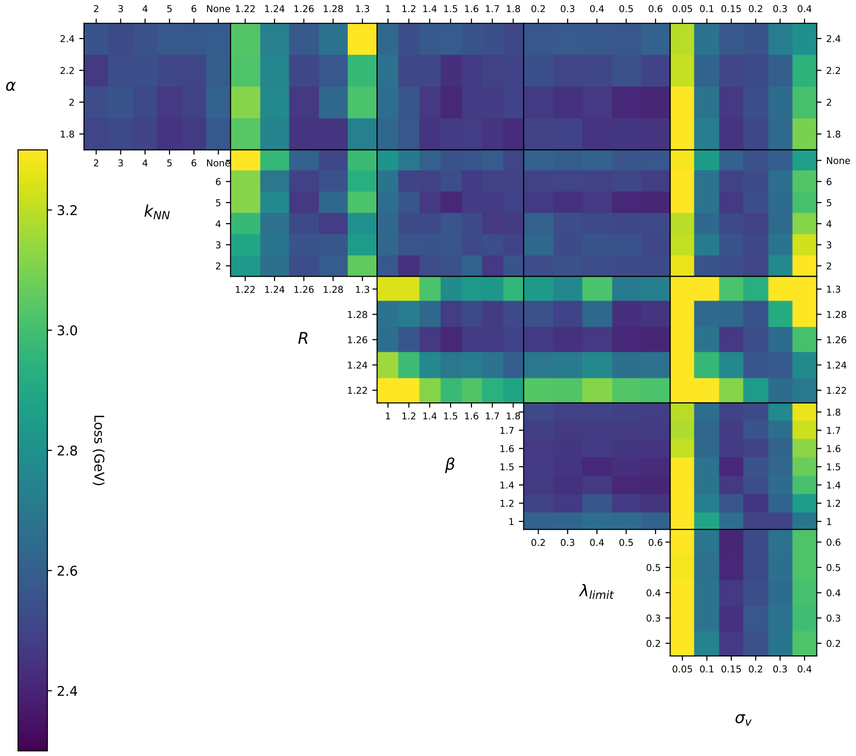


FIGURE 8.2: The loss $\mathcal{L}$ plotted over the scan of hyperparameters.

In our scan, we therefore test a variety of values for each parameter, and record the corresponding loss, as seen in Fig. 8.2. We note that while a particular set of parameters are selected for this study, there are large regions of parameter space that result in low loss values - and so do not need to be fine tuned for an acceptable clustering performance. The final values used in the remainder of this paper are

$$\alpha = 2,\ k_{\text{NN}} = 5,\ R = 1.26,\ \beta = 1.4,\ \sigma_v = 0.15,\ \lambda_{\text{cut}} = 0.4. \tag{8.24}$$

8.2.7 Infrared Safety

As we saw in a previous chapter, any good jet clustering algorithm must be IR safe. We demonstrate this for our spectral clustering algorithm by computing an IR sensitive variable, and showing it remains consistent between datasets both containing and missing IR singularities. By means of completing an exhaustive search, we again perform a scan over the hyperparameters of the model, to show that there is no combination that is IR unsafe.

To avoid the task of comparing mass spectra by eye, we compute the Jensen-Shannon score [128], which quantitatively encodes the divergence between a pairs of distributions. For two probability density distributions p and q, the Jensen-Shannon (JS) score is

$$D_{\text{JS}}(p,q) = \frac{1}{2}D_{\text{KL}}\left(p|\frac{1}{2}(p+q)\right) + \frac{1}{2}D_{\text{KL}}\left(q|\frac{1}{2}(p+q)\right) \tag{8.25}$$

where

$$D_{\text{KL}}(p|q) = \int_{-\infty}^{\infty} p(x)\log\left(\frac{p(x)}{q(x)}\right)dx \tag{8.26}$$

is the Kullback-Leibler divergence.

8.3 Results

In this section we present the results of our spectral clustering algorithm, starting with a demonstration of the IR safety. We then present a comparison of the performance of spectral clustering against the anti-k_T jet algorithm.

8.3.1 IR Safety

8.3.1.1 Jet Shape Variables

To test the IR safety of a given algorithm, we compute the following jet shape variables; jet mass, thrust, sphericity, spherocity, oblateness. For an in-depth review of jet shape variable, we refer the reader to [129].

Firstly, the jet mass is found by constructing the invariant mass from the combined four momentum of its constituents. The jet thrust is defined relative to a specified axis $\vec{n}_{\text{thrust}}$,

$$t = \max\left(\frac{\sum_a |\vec{p}_a| \cdot |n_{\text{thrust}}|}{\sum_a |\vec{p}_a|}\right), \tag{8.27}$$

where the sum is over hadrons in a jet. $\vec{p}_a$ is the three-momentum of the a^{th} jet constituent Sphericity is defined as

$$S = \frac{3}{2}\frac{\sum_a |p_T^a|^2}{\sum_a |p^a|^2}, \tag{8.28}$$

where p^a is the three-momentum, and p^a_T transverse momentum of the a^{th} particle in a given jet. Sphericity is IR safe, but not collinear safe, so spherocity is also tested, which is IRC safe, given by

$$\bar{S} = \frac{4}{\pi} \min \left(\frac{\sum_a |p^a_T|}{\sum_a |p^a|} \right)^2 . \tag{8.29}$$

The oblateness of a jet, similarly to thrust t, is defined in with respect to axes. Firstly n_{major}, which is in the plane orthogonal to n_{thrust}. There is also n_{minor}, which is orthogonal to both n_{major} and n_{thrust}. The overall oblateness is written in terms of terms containing these two axes,

$$O = F_{\text{major}} - F_{\text{minor}} = \frac{\sum_a |\vec{E}_a \cdot \vec{n}_{\text{major}}|}{\sum_a E_a} - \frac{\sum_a |\vec{E}_a \cdot \vec{n}_{\text{minor}}|}{\sum_a E_a}, \tag{8.30}$$

where the E_a are jet constituents.

8.3.1.2 Testing for IR Safety

We observe that an IR safe algorithm, when computed on a LO and NLO dataset, will give a low JS score, whereas an IR unsafe algorithm would be sensitive to soft and collinear emissions present in the NLO data, corresponding to a greater difference in the jet shape variables and hence larger D_{JS}.

We use the 3J dataset described in section 8.2.3, and compute the shape variables as detailed above. It is clear in Fig. 8.3 that spectral clustering is closely matching the performance of the k_T algorithm.

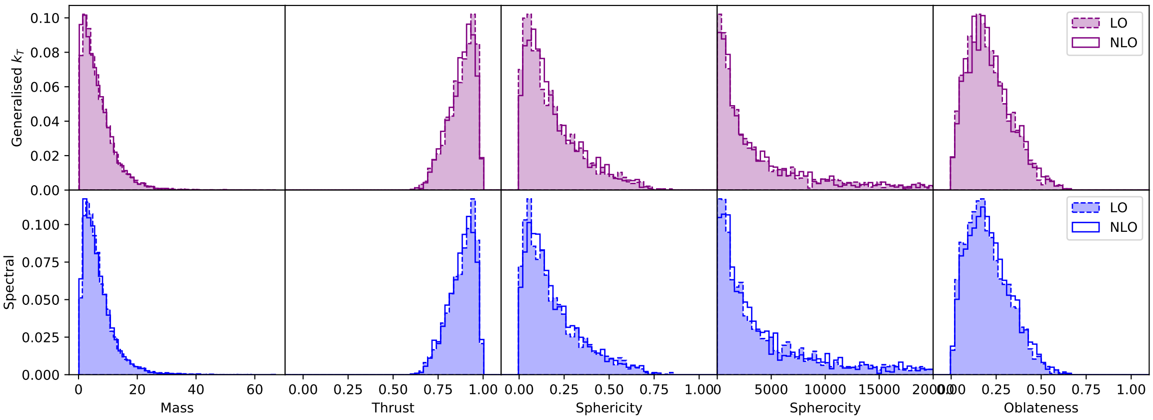


FIGURE 8.3: A comparison of the jet shape variables computed in the LO and NLO dataset, for both the generalised k_T algorithm and for our spectral clustering algorithm.

As an additional measure, we present the JS score of the jet shape variables, this time comparing the generalised k_T, spectral clustering and the iterative cone algorithm [130], which is known to be IR unsafe. The inclusion of the latter ensures that we can verify our safety measure does indeed identify IR unsafe algorithms with an enhanced JS score.

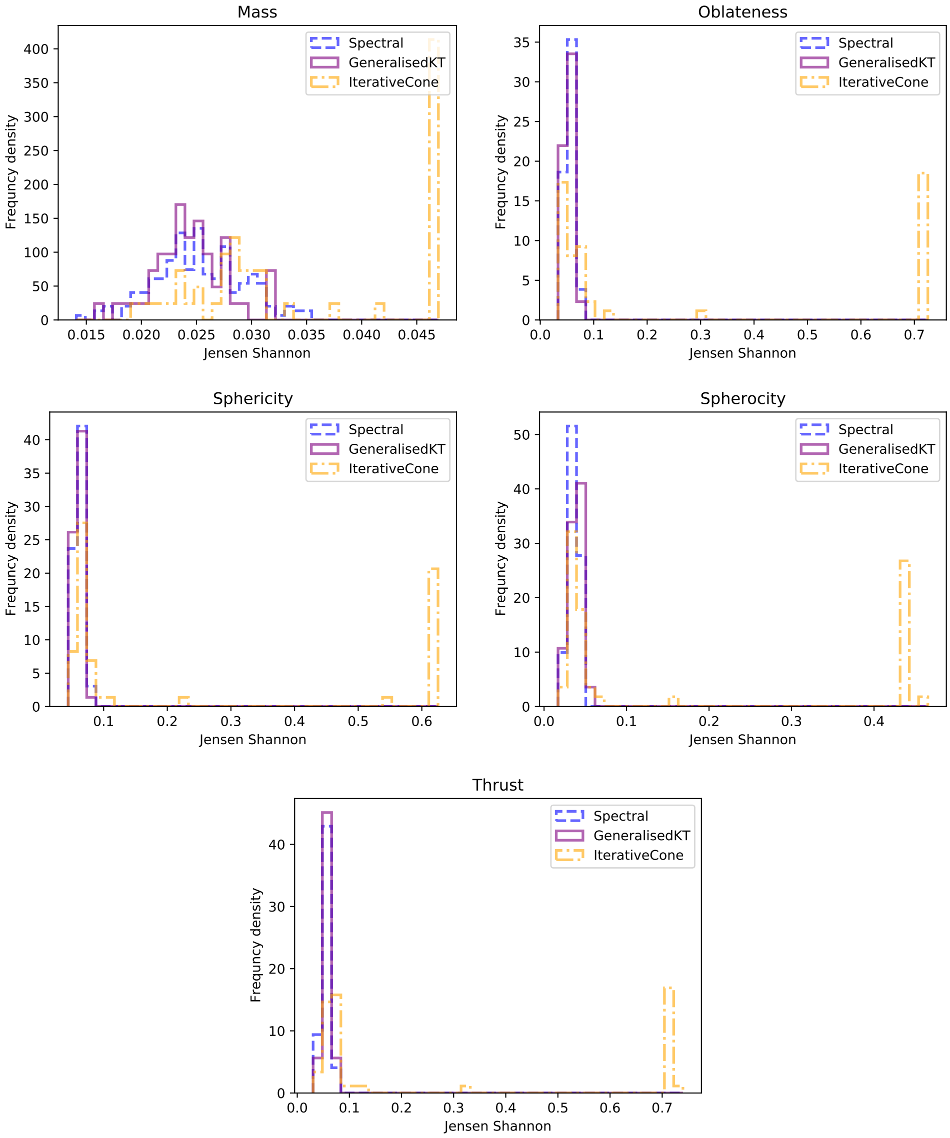


Figure 8.4: The JS scores for the five jet shape variables, computed for our SC algorithm, the generalised k_T algorithm, and the IR unsafe iterative cone algorithm. Plotted is a distribution of JS scores, where each score corresponds to a different selection of clustering parameters.

We can indeed see in Fig. 8.4, that the SC algorithm closely matches the generalised k_T, with comparatively low JS scores compared to the iterative cone algorithm.

8.3.2 Signal Jet Reconstruction

We finally consider the physics performance of our SC algorithm when compared with the anti-k_T. We consider two implementations of the anti-k_T, both with $R_{k_T} = 0.4$

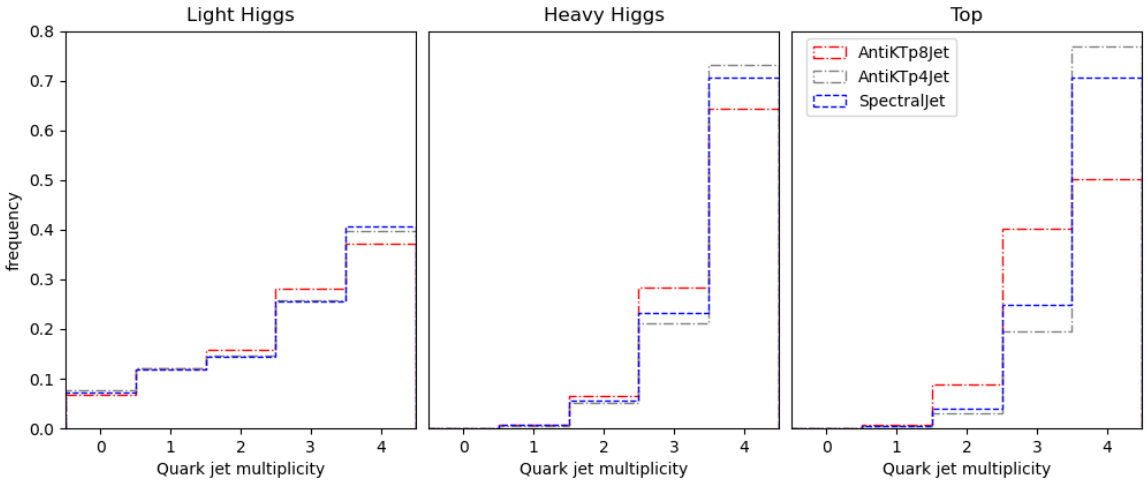


FIGURE 8.5: The jet multiplicity for each of the datasets, comparing the three algorithms AK4, AK8 and SC.

(AK4) and $R_{k_T} = 0.8$ (AK8), and measure the performance on three datasets, the LH, HH and TT described in section 8.2.3.

Firstly, we compare the jet multiplicity of the three algorithms. As we are generating processing with a known number of expected final state jets, we can identify a good jet algorithm as one that reconstructs all expected jets frequently.

We can see from Fig. 8.5 that for Higgs datasets, all three algorithms perform comparably, for the TT dataset AK8 is clearly a less optimal choice, and in fact AK4 reconstrctus the expected four jets most of the time. However as we have seen, the number of jets is not an exhaustive way of testing the jet algorithms - we require the jets reconstructed to represent the physics from which they originated, in the form of invariant mass distributions.

As a side note, it is worth mentioning the stark difference between the jet multiplicities quoted here and in [1]. A big source of jet suppression in [1] comes from the p_T cuts, which are made looser in this study. Furthermore, the implementation of b-tagging here is slightly different than in the previous study. Here, we use a fixed tagging radius of $R = 0.8$ universally, whereas previously we had matched the tagging radius to the radius used in jet clustering, this enhances the number of jets successfully b-tagged in this study. We also do not simulate any b-tagging inefficiencies here, unlike in [1] where p_T dependent b-tag probabilities are used.

From Fig. 8.6, we plot the mass distributions of the jets in each sample, reconstructed with each algorithm. In the LH sample, we observe SC leading to narrower mass peaks in reconstructing both the $m_H = 125$GeV Higgs (from all four b-jets) and for the two $m_h = 40$GeV Higgses. The stronger and weaker signal labels refer to the b-jet pairings with a larger and smaller combined mass respectively. We also note that here, the peaks from the AK8 algorithm appear sharper than those from the AK4.

For the HH data, the performance of the three algorithms are more closely matched, with SC and AK8 in particular producing narrow peaks at the expected mass for the respective Higgs the jets are reconstructing, and AK4 peaks being slightly below the expected values - perhaps due the signal mass lost $\mathcal{S}$. Referring back the Fig. 8.5, we note that the multiplicity for SC was better for the HH data than the AK8 algorithm, and so we argue here that SC is in fact the optimal choice. We also highlight a further benefit of SC, in that for both the LH and HH datasets we have used the same model, with the same hyperparameters, both achieving a strong performance. In contrast the anti-k_T algorithm performance varies for the two choices of R_{k_T} - implying an element of fine tuning in the choice of this parameter on different datasets.

We finally consider the TT dataset. The masses constructed are as follows; the hadronic W is reconstructed from the jj decay products, where the correct partons corresponding to the jets are identified using MC information, via Eq. 8.7. The hadronic t is reconstructed from the combination of the associated W described previously, along with the companion b-jet, which is tagged in the usual way. The leptonic t is reconstructed from the leptonic W and its corresponding b-jet. Again here the SC performance is comparable to the best anti-k_T algorithm, in this case AK4, and we again draw attention to the fact that this is the same variation of SC used for the other datasets.

8.4 Conclusions

We have demonstrated that an alternative to sequential recombination jet clustering algorithms, using machine learning, can achieve a comparative performance, without the need to tune the hyperparameters to the particular physics we are interested in.

By utilising spectral clustering, in which we transform data from it's original feature into a new, transformed embedding space, we are able to train a model with specific hyperparameters that can accurately reconstruct jets from various physics process, of various different kinematics and sizes. In contrast, the user input in the anti-k_T algorithm is dependent on the spread of radiation in the jets themselves, and should be adjusted accordingly - as demonstrated in varied performance of AK4 and AK8 on different physics processes. Furthermore, SC is also adaptable in that many other combinations of hyperparameters perform adequately, such that it is possible for a single ‘out-of-the-box’ configuration to work well on whatever physics is thrown at it, all while maintaining IR safety.

Finally, while there are a multitude of different ML inspired methods for jet reconstruction, SC has the added benefit of being strikingly simple when compared with deep learning methods. Only basic linear algebra is required to construct the Laplacian and its eigenvectors and corresponding eigenvalues.

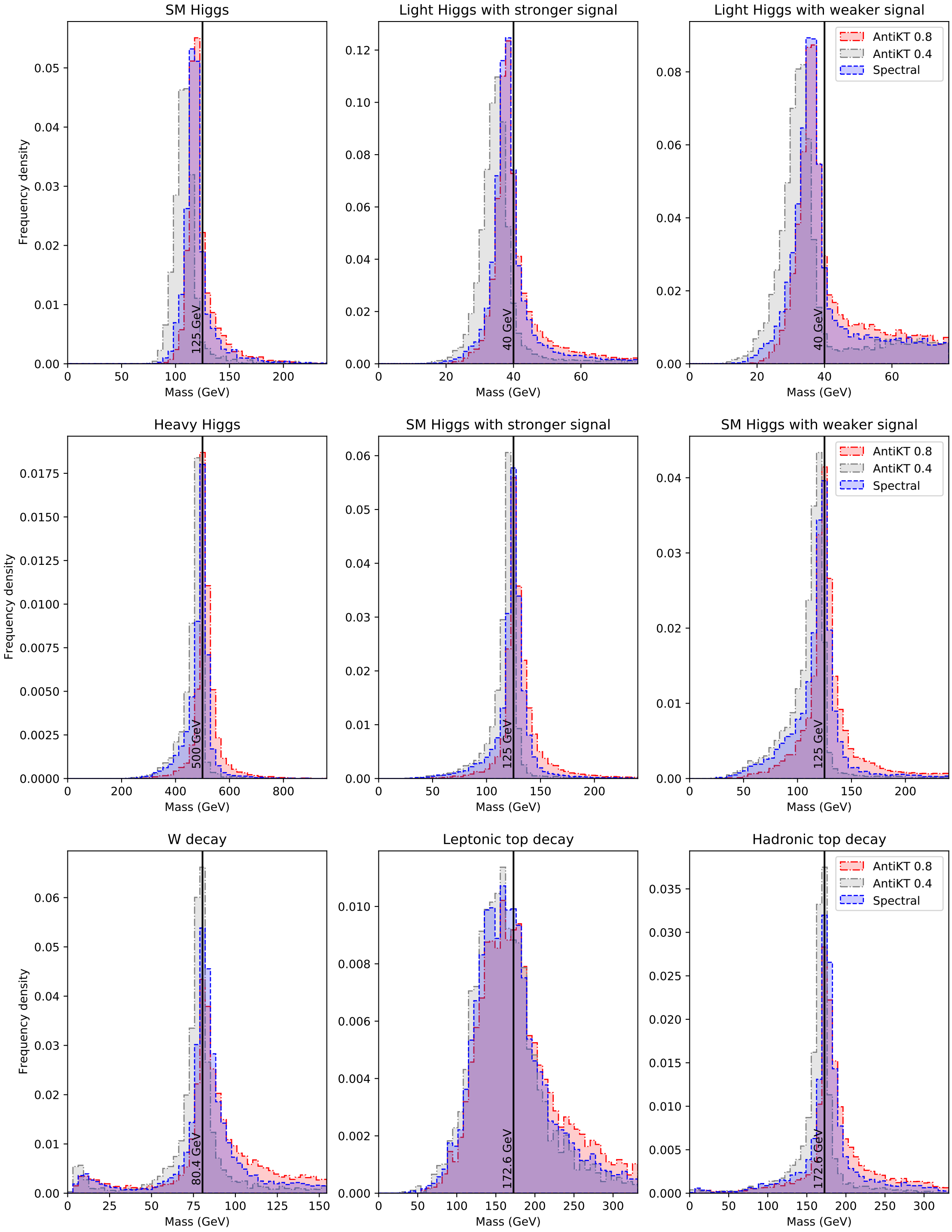


FIGURE 8.6: The jet masses for each of the datasets, HH (upper panel), LH (middle) and TT (lower), comparing the three algorithms AK4, AK8 and SC.

The simplicity and performance of SC make it a promising new candidate for jet physics, the results of this study highly motivating further study into more complex effects such as full detector simulation, pile-up and MPI, and fat jet substructure.

Part III

Final Comments

Chapter 9

Conclusions

The role of particle accelerators, both current (LHC) and potentially in the future with the high luminosity LHC (HL-LHC), is to explore new physics beyond the SM. The environment inside such experiments are noisy and highly populated with radiation from multiple sources - both from processes of potential interest and additional ‘junk’ which must be carefully removed to accurately assess the physics content of a given event. As such, it is of paramount importance that the methodologies of mapping detected final state radiation into the underlying physical process from which it came can perform well, for many different types of physics.

At the forefront of such methods are jets - sprays of hadronic radiation arising from a final state parton, which are confined to cone shapes due to the evolving nature of the QCD coupling over distance scales. Accurately reconstructing and identifying the source of such jets is a non-negotiable requirement for extracting physics from these messy final states.

An exciting potential source of new physics are so called two-Higgs-doublet-models (2HDMs), which closely match the SM, with the addition of an extended Higgs sector, in the form of additional physics Higgs states - $h, H, A, H^{\pm}$. Measurements of the Higgs boson, discovered in 2012, show that the preferred decay mode for this Higgs is into a pair of b-quarks, which in turn form b-jets in the detector. This makes 2HDMs a good candidate for assessing jet reconstruction techniques - however we note that of course jet reconstruction is universally required for any BSM models in one way or another.

The current popularity of data science and machine learning are leading advancement in a diverse range of fields, from our day-to-day lives in social media, e-commerce and banking, to AI that can outperform humans in chess, aiding medical diagnoses and self driving cars. We have shown that high energy physics can also profit from these these techniques.

Firstly, by employing the power of modern high performance computing (HPC), we are able to simulate huge particle datasets mimicking what is seen in experiments such as the LHC. In chapter 6 we test alternative jet clustering algorithms on high b-jet multiplicity datasets from $H \to hh \to bb\bar{b}\bar{b}$ decays, with a variable R parameter, which we demonstrate allows for clustering single events containing multiple jets of different angular sizes. We see that, when combined with quality cuts, there is an improvement in signal to background ratio when using variable-R in place of traditional fixed-R algorithms.

In chapter 7 we then consider more cutting edge techniques on b-jets from 2HDM-II Higgs decays. By mapping the constituents of jets in these decays into images, we apply deep learning techniques to build and train a classifier to predict which kind of physics event a set of jets came from. In this iteration, we present a proof-of-concept which highlights the ability of CNNs to learn from relatively low level information (that is, direct p_T deposits in a detector, rather than other ML methods which can require highly tuned parameters).

Finally, in chapter 8, we consider the origin of the jet formation, that is in the initial clustering of final state particles - and present a novel, ML informed, clustering algorithm based upon spectral clustering. This SC algorithm can be trained and used on a variety of different events, with jets of different angular scales, without the need for constant changes to the model parameters. It is shown that SC performs comparatively well compared with the anti-k_T clustering algorithms.

In all three of these endeavours we observe that traditional jet reconstruction techniques, while very powerful, can potentially be improved upon with more modern ML and data intensive methodologies. The field of ML is continually growing and advancing, and so therefore will the applications in high energy physics. With the plans for the HL-LHC to begin in 2027, which will drastically increase the amount of data produced, the demand for new ways to use and model data will also grow - which makes for an exciting time in high energy physics to push the current boundaries of our understanding of the fundamental building blocks of nature.

References

[1] A. Chakraborty, S. Dasmahapatra, H. Day-Hall, B. Ford, S. Jain, S. Moretti, E. Olaiya and C. Shepherd-Themistocleous, [arXiv:2008.02499 [hep-ph]].

[2] S. Dasmahapatra, H. A. Day-Hall, B. Ford, S. Moretti and C. H. Shepherd-Themistocleous, [arXiv:2104.01972 [hep-ph]].

[3] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964), 321-323

[4] S. Weinberg, Phys. Rev. Lett. **19** (1967), 1264-1266

[5] M. Bustamante, L. Cieri and J. Ellis, [arXiv:0911.4409 [hep-ph]].

[6] C. Quigg, Ann. Rev. Nucl. Part. Sci. **59** (2009), 505-555 [arXiv:0905.3187 [hep-ph]].

[7] J. Ellis, [arXiv:1312.5672 [hep-ph]].

[8] S. Dawson, AIP Conf. Proc. **1116** (2009) no.1, 11-34 [arXiv:0812.2190 [hep-ph]].

[9] J. Goldstone, A. Salam and S. Weinberg, Phys. Rev. **127** (1962), 965-970

[10] R. K.Ellis, W. J.Stirling and B. R.Webber,

[11] M. E. Peskin and D. V. Schroeder,

[12] B. P. Abbott *et al.* [LIGO Scientific and Virgo], Phys. Rev. Lett. **116** (2016) no.6, 061102 [arXiv:1602.03837 [gr-qc]].

[13] E. Corbelli and P. Salucci, Mon. Not. Roy. Astron. Soc. **311** (2000), 441-447 [arXiv:astro-ph/9909252 [astro-ph]].

[14] S. M. Faber, R. E Jackson, The Astrophysical Journal **204** (1976), 668683.

[15] A. McKellar, Publications of the Dominion Astrophysical Observatory Victoria 7, **251** (1941).

[16] A. A. Penzias, R. W Wilson, The Astrophysical Journal **142** (1965), 419421.

[17] N. Hata, G. Steigman, S. A. Bludman and P. Langacker, Phys. Rev. D **55** (1997), 540-547 [arXiv:astro-ph/9603087 [astro-ph]].

[18] Y. Fukuda *et al.* [Super-Kamiokande], Phys. Rev. Lett. **82** (1999), 2644-2648 [arXiv:hep-ex/9812014 [hep-ex]].

[19] Q. R. Ahmad *et al.* [SNO], Phys. Rev. Lett. **87** (2001), 071301 [arXiv:nucl-ex/0106015 [nucl-ex]].

[20] D. A. Camargo, A. G. Dias, T. B. de Melo and F. S. Queiroz, JHEP **04** (2019), 129 [arXiv:1811.05488 [hep-ph]].

[21] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher and J. P. Silva, Phys. Rept. **516** (2012), 1-102 [arXiv:1106.0034 [hep-ph]].

[22] V. A. Bednyakov, N. D. Giokaris and A. V. Bednyakov, Phys. Part. Nucl. **39** (2008), 13-36 [arXiv:hep-ph/0703280 [hep-ph]].

[23] D. Eriksson, J. Rathsman and O. Stal, Comput. Phys. Commun. **181** (2010), 833-834.

[24] J. F. Gunion and H. E. Haber, Phys. Rev. D **67** (2003), 075019 [arXiv:hep-ph/0207010 [hep-ph]].

[25] N. Craig and S. Thomas, JHEP **11** (2012), 083 doi:10.1007/JHEP11(2012)083 [arXiv:1207.4835 [hep-ph]].

[26] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D **2** (1970), 1285-1292.

[27] A. G. Akeroyd, A. Arhrib and E. M. Naimi, Phys. Lett. B **490** (2000), 119-124 [arXiv:hep-ph/0006035 [hep-ph]].

[28] M. E. Peskin and T. Takeuchi, Phys. Rev. Lett. **65** (1990), 964-967.

[29] I. Maksymyk, C. P. Burgess and D. London, Phys. Rev. D **50** (1994), 529-535 [arXiv:hep-ph/9306267 [hep-ph]].

[30] S. Kanemura, M. Kikuchi and K. Yagyu, Nucl. Phys. B **896** (2015), 80-137 [arXiv:1502.07716 [hep-ph]].

[31] G. Aad *et al.* [ATLAS and CMS], JHEP **08** (2016), 045 [arXiv:1606.02266 [hep-ex]].

[32] G. Aad *et al.* [ATLAS], Phys. Lett. B **800** (2020), 135103 [arXiv:1906.02025 [hep-ex]].

[33] G. Aad *et al.* [ATLAS], Eur. Phys. J. C **76** (2016) no.4, 210 [arXiv:1509.05051 [hep-ex]].

[34] V. Khachatryan *et al.* [CMS], JHEP **10** (2017), 076 [arXiv:1701.02032 [hep-ex]].

[35] A. G. Akeroyd, M. Aoki, A. Arhrib, L. Basso, I. F. Ginzburg, R. Guedes, J. Hernandez-Sanchez, K. Huitu, T. Hurth and M. Kadastik, *et al.* Eur. Phys. J. C **77** (2017) no.5, 276 [arXiv:1607.01320 [hep-ph]].

[36] G. S. Bali *et al.* [SESAM], Phys. Rev. D **71** (2005), 114513 [arXiv:hep-lat/0505012 [hep-lat]].

[37] T. Appelquist, J. Carazzone, H. Kluberg-Stern and M. Roth, Phys. Rev. Lett. **36** (1976), 768 [erratum: Phys. Rev. Lett. **36** (1976), 1161] doi:10.1103/PhysRevLett.36.768

[38] G. Altarelli and G. Parisi, Nucl. Phys. B **126** (1977), 298-318.

[39] Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze and S. I. Troian, Phys. Lett. B **165** (1985), 147-150.

[40] R. P. Feynman,, W.A. Benjamin, New York (1974).

[41] B. R. Webber, [arXiv:hep-ph/9411384 [hep-ph]].

[42] B. Andersson, G. Gustafson and B. Soderberg, Z. Phys. C **20** (1983), 317.

[43] A. M. Sirunyan *et al.* [CMS], Phys. Rev. Lett. **121** (2018) no.12, 121801 [arXiv:1808.08242 [hep-ex]].

[44] G. F. Sterman and S. Weinberg, Phys. Rev. Lett. **39** (1977), 1436.

[45] W. Bartel *et al.* [JADE], Z. Phys. C **33** (1986), 23.

[46] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock and B. R. Webber, Phys. Lett. B **269** (1991), 432-438.

[47] S. D. Ellis and D. E. Soper, Phys. Rev. D **48** (1993), 3160-3166 [arXiv:hep-ph/9305266 [hep-ph]].

[48] M. Cacciari, G. P. Salam and G. Soyez, JHEP **04** (2008), 063 [arXiv:0802.1189 [hep-ph]].

[49] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **08** (1997), 001 [arXiv:hep-ph/9707323 [hep-ph]].

[50] D. Krohn, J. Thaler and L. T. Wang, JHEP **06** (2009), 059 [arXiv:0903.0392 [hep-ph]].

[51] G. L. Bayatian *et al.* [CMS], CERN-LHCC-2006-001.

[52] T. N. Pham, [arXiv:1110.6050 [hep-ph]].

[53] I. R. Tomalin [CMS], J. Phys. Conf. Ser. **110** (2008), 092033.

[54] P. A. Zyla *et al.* [Particle Data Group], PTEP **2020** (2020) no.8, 083C01 doi:10.1093/ptep/ptaa104

[55] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, Phys. Rev. Lett. **101** (2008), 142001 doi:10.1103/PhysRevLett.101.142001 [arXiv:0806.0848 [hep-ph]].

[56] D. Bertolini, P. Harris, M. Low and N. Tran, JHEP **10** (2014), 059 [arXiv:1407.6013 [hep-ph]].

[57] D. Krohn, J. Thaler and L. T. Wang, JHEP **02** (2010), 084 [arXiv:0912.1342 [hep-ph]].

[58] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **80** (2009), 051501 [arXiv:0903.5081 [hep-ph]].

[59] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, JHEP **05** (2014), 146 [arXiv:1402.2657 [hep-ph]].

[60] J. Thaler and K. Van Tilburg, JHEP **03** (2011), 015 [arXiv:1011.2268 [hep-ph]].

[61] D. R. Cox, Journal of the Royal Statistical Society: Series B (Methodological), 20(2), 215232.

[62] A. GeÌĄron, OReilly (2019).

[63] D. P. Kingma, J. Ba, [arXiv:1412.6980 [cs.LG]].

[64] S. Haykin Prentice Hall PTR (1994).

[65] Y. LeCun, Y. Bengio and G. Hinton Nature, 521(7553), pp.436-444.

[66] S. P. Lloyd, Information Theory, IEEE Transactions on 28.2 (1982): 129-137.

[67] A. Y. Ng, M. I. Jordan and Y. Weiss, in Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, NIPS01, (Cambridge, MA, USA), (2001) p. 849856, MIT Press.

[68] A. J. Larkoski, I. Moult and B. Nachman, Phys. Rept. **841** (2020), 1-63 [arXiv:1709.04464 [hep-ph]].

[69] D. Guest, K. Cranmer and D. Whiteson, Ann. Rev. Nucl. Part. Sci. **68** (2018), 161-181 [arXiv:1806.11484 [hep-ex]].

[70] K. Albertsson, P. Altoe, D. Anderson, J. Anderson, M. Andrews, J. P. Araque Espinosa, A. Aurisano, L. Basara, A. Bevan and W. Bhimji, *et al.* J. Phys. Conf. Ser. **1085** (2018) no.2, 022008 [arXiv:1807.02876 [physics.comp-ph]].

[71] A. M. Sirunyan *et al.* [CMS], JINST **13** (2018) no.05, P05011 [arXiv:1712.07158 [physics.ins-det]].

[72] [ATLAS], ATL-PHYS-PUB-2017-003.

[73] J. Bielčíková, R. Kunnawalkam Elayavalli, G. Ponimatkin, J. H. Putschke and J. Sivic, JINST **16** (2021) no.03, P03017 [arXiv:2005.01842 [hep-ph]].

[74] E. Bols, J. Kieseler, M. Verzetti, M. Stoye and A. Stakia, JINST **15** (2020) no.12, P12012 [arXiv:2008.10519 [hep-ex]].

[75] G. Kasieczka, T. Plehn, A. Butter, K. Cranmer, D. Debnath, B. M. Dillon, M. Fairbairn, D. A. Faroughy, W. Fedorko and C. Gay, *et al.* SciPost Phys. **7** (2019), 014 [arXiv:1902.09914 [hep-ph]].

[76] A. Chakraborty, S. H. Lim, M. M. Nojiri and M. Takeuchi, JHEP **07** (2020), 111 [arXiv:2003.11787 [hep-ph]].

[77] J. A. Aguilar-Saavedra, Eur. Phys. J. C **81** (2021) no.8, 734 [arXiv:2102.01667 [hep-ph]].

[78] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee and M. Perelstein, JHEP **07** (2015), 086 [arXiv:1501.05968 [hep-ph]].

[79] S. Macaluso and D. Shih, JHEP **10** (2018), 121 [arXiv:1803.00107 [hep-ph]].

[80] G. Kasieczka, T. Plehn, M. Russell and T. Schell, JHEP **05** (2017), 006 [arXiv:1701.08784 [hep-ph]].

[81] S. Farrell, P. Calafiura, M. Mudigonda, Prabhat, D. Anderson, J. R. Vlimant, S. Zheng, J. Bendavid, M. Spiropulu and G. Cerati, *et al.* [arXiv:1810.06111 [hep-ex]].

[82] S. Amrouche, L. Basara, P. Calafiura, V. Estrade, S. Farrell, D. R. Ferreira, L. Finnie, N. Finnie, C. Germain and V. V. Gligorov, *et al.* [arXiv:1904.06778 [hep-ex]].

[83] V. V. Gligorov and M. Williams, JINST **8** (2013), P02013 [arXiv:1210.6861 [physics.ins-det]].

[84] T. Q. Nguyen, D. Weitekamp, D. Anderson, R. Castello, O. Cerri, M. Pierini, M. Spiropulu and J. R. Vlimant, Comput. Softw. Big Sci. **3** (2019) no.1, 12 [arXiv:1807.00083 [hep-ex]].

[85] P. T. Komiske, E. M. Metodiev, B. Nachman and M. D. Schwartz, JHEP **12** (2017), 051 [arXiv:1707.08600 [hep-ph]].

[86] J. Arjona Martínez, O. Cerri, M. Pierini, M. Spiropulu and J. R. Vlimant, Eur. Phys. J. Plus **134** (2019) no.7, 333 [arXiv:1810.07988 [hep-ph]].

[87] S. Carrazza and F. A. Dreyer, Phys. Rev. D **100** (2019) no.1, 014014 [arXiv:1903.09644 [hep-ph]].

[88] J. Brehmer, K. Cranmer, G. Louppe and J. Pavez, Phys. Rev. Lett. **121** (2018) no.11, 111801 [arXiv:1805.00013 [hep-ph]].

[89] J. Brehmer, K. Cranmer, G. Louppe and J. Pavez, Phys. Rev. D **98** (2018) no.5, 052004 [arXiv:1805.00020 [hep-ph]].

[90] A. Andreassen, B. Nachman and D. Shih, Phys. Rev. D **101** (2020) no.9, 095004 [arXiv:2001.05001 [hep-ph]].

[91] G. Aad *et al.* [ATLAS], Phys. Lett. B **716** (2012), 1-29 [arXiv:1207.7214 [hep-ex]].

[92] S. Khalil and S. Moretti,, "Supersymmetry Beyond Minimality: from Theory to Experiment",

[93] U. Ellwanger, C. Hugonie and A. M. Teixeira, Phys. Rept. **496** (2010), 1-77 [arXiv:0910.1785 [hep-ph]].

[94] J. F. Gunion, H. E. Haber, G. L. Kane and S. Dawson, Front. Phys. **80** (2000), 1-404 SCIPP-89/13.

[95] J. F. Gunion, H. E. Haber, G. L. Kane and S. Dawson, [arXiv:hep-ph/9302272 [hep-ph]].

[96] S. Moretti and W. J. Stirling, Phys. Lett. B **347** (1995), 291-299 [erratum: Phys. Lett. B **366** (1996), 451] [arXiv:hep-ph/9412209 [hep-ph]].

[97] A. Djouadi, J. Kalinowski and P. M. Zerwas, Z. Phys. C **70** (1996), 435-448 [arXiv:hep-ph/9511342 [hep-ph]].

[98] L. Scodellaro [ATLAS and CMS], [arXiv:1709.01290 [hep-ex]].

[99] S. Moretti, L. Lonnblad and T. Sjostrand, JHEP **08** (1998), 001 [arXiv:hep-ph/9804296 [hep-ph]].

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

[101] M. Wobisch and T. Wengler, [arXiv:hep-ph/9907280 [hep-ph]].

[102] G. P. Salam, Eur. Phys. J. C **67** (2010), 637-686 [arXiv:0906.1833 [hep-ph]].

[103] A. M. Sirunyan *et al.* [CMS], JINST **13** (2018) no.05, P05011 [arXiv:1712.07158 [physics.ins-det]].

[104] A. M. Sirunyan *et al.* [CMS], Phys. Lett. B **795** (2019), 398-423 [arXiv:1812.06359 [hep-ex]].

[105] J. de Favereau *et al.* [DELPHES 3], JHEP **02** (2014), 057 [arXiv:1307.6346 [hep-ex]].

[106] B. Tannenwald, C. Neu, A. Li, G. Buehlmann, A. Cuddeback, L. Hatfield, R. Parvatam and C. Thompson, [arXiv:2009.06754 [hep-ph]].

[107] P. Bechtle, O. Brein, S. Heinemeyer, O. Stål, T. Stefaniak, G. Weiglein and K. E. Williams, Eur. Phys. J. C **74** (2014) no.3, 2693 [arXiv:1311.0055 [hep-ph]].

[108] P. Bechtle, S. Heinemeyer, O. Stål, T. Stefaniak and G. Weiglein, Eur. Phys. J. C **74** (2014) no.2, 2711 [arXiv:1305.1933 [hep-ph]].

[109] F. Mahmoudi, Comput. Phys. Commun. **180** (2009), 1718-1719

[110] R. D. Ball *et al.* [NNPDF], JHEP **04** (2015), 040 [arXiv:1410.8849 [hep-ph]].

[111] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, JHEP **07** (2014), 079 [arXiv:1405.0301 [hep-ph]].

[112] T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178** (2008), 852-867 [arXiv:0710.3820 [hep-ph]].

[113] E. Conte, B. Fuks and G. Serret, Comput. Phys. Commun. **184** (2013), 222-256 [arXiv:1206.1599 [hep-ph]].

[114] E. Conte and B. Fuks, Int. J. Mod. Phys. A **33** (2018) no.28, 1830027 [arXiv:1808.00480 [hep-ph]].

[115] A. M. Sirunyan *et al.* [CMS], JHEP **04** (2019), 112 [arXiv:1810.11854 [hep-ex]].

[116] A. M. Sirunyan *et al.* [CMS], JHEP **03** (2021), 257 [arXiv:2011.12373 [hep-ex]].

[117] J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever and J. Rojo, Eur. Phys. J. C **76** (2016) no.7, 386 [arXiv:1512.08928 [hep-ph]].

[118] J. Amacker, W. Balunas, L. Beresford, D. Bortoletto, J. Frost, C. Issever, J. Liu, J. McKee, A. Micheli and S. Paredes Saenz, *et al.* [arXiv:2004.04240 [hep-ph]].

[119] T. Lapsien, R. Kogler and J. Haller, Eur. Phys. J. C **76** (2016) no.11, 600 [arXiv:1606.04961 [hep-ph]].

[120] [ATLAS], ATL-PHYS-PUB-2016-013.

[121] [ATLAS], ATL-PHYS-PUB-2017-010.

[122] [ATLAS], ATLAS-CONF-2020-007.

[123] [CMS], CMS-PAS-BTV-15-002.

[124] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer and A. Mück, SciPost Phys. **10** (2021) no.2, 046 [arXiv:2006.08639 [hep-ph]].

[125] A. M. Sirunyan *et al.* [CMS], Phys. Rev. Lett. **124** (2020) no.20, 202001 [arXiv:1911.03800 [hep-ex]].

[126] A. M. Sirunyan *et al.* [CMS], JINST **13** (2018) no.06, P06015 [arXiv:1804.04528 [physics.ins-det]].

[127] X. Ju and B. Nachman, Phys. Rev. D **102** (2020) no.7, 075014 [arXiv:2008.06064 [hep-ph]].

[128] J. Lin, IEEE Transactions on Information Theory **37** (1991) 145.

[129] G. Altarelli, R. Kleiss and C. Verzegnassi, doi:10.5170/CERN-1989-008-V-1

[130] G. P. Salam and G. Soyez, JHEP **05** (2007), 086 [arXiv:0704.0292 [hep-ph]].