Extracting jet signals of new Higgs physics: from data analysis to machine learning

H.A. Day-Hall

Supervisors; S. Dasmahapatra, S. Moretti, C.H. Shepherd-Themistocleous

September 28, 2020

Contents

1 Introduction

I am a student from the NGCM, an interdisciplinary doctoral training centre focused on computational problem solving across STEM fields. The investigations presented here straddle a number of topics; machine learning, high energy experimental physics and high energy phenomenology. This has been possible thanks to training from
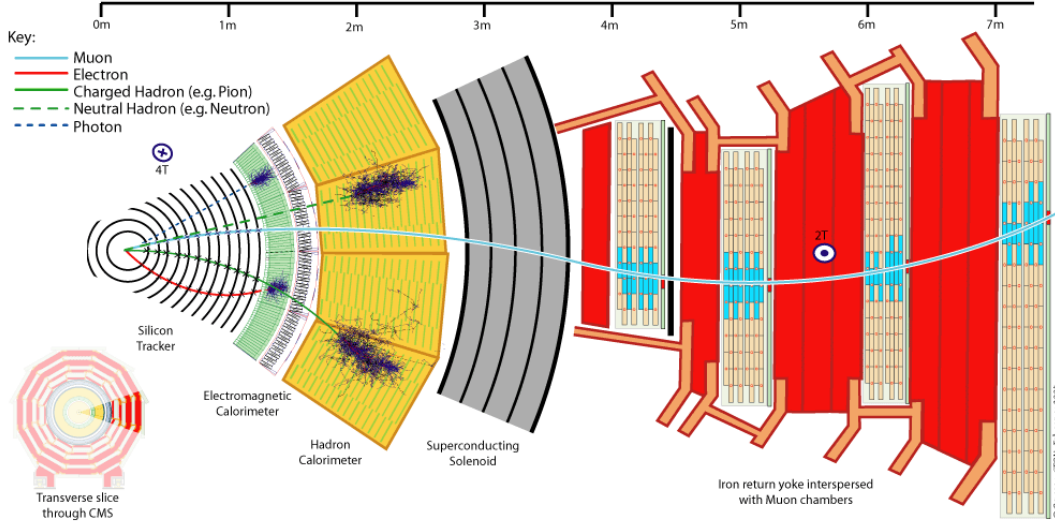
Figure 1: The CMS detector has a concentric structure with each layer sensitive to a subset of possible particles. [3]

both the physics and computer science departments during the first year of this course, and continued feedback at conferences and workshops, detailed in section 7.

Stefano Moretti (Physics, Univeristy of Southampton) provides supervision and guidance on the phenomenology aspect, Claire Shepherd-Themistocleous (CMS, Rutherford Appleton Laboratory) provides supervision with experimental expertises and Srinandan Dasmahapatra (Computer Science, University of Southampton) provides supervision with regards to machine learning.

2 Problem statement and outline

Particle colliders such as the LHC grant insight into physics by creating heavy short lived particles that do not occur without a high energy collision and only persist for a fraction of a second before decaying into lighter particles. There is then a sequence of decays, known as a shower, before something stable and long lived enough to reach the detectors is created. From the remnants of the decay that reach the detector we seek to characterise the heavy particle created in the collision.

The data pipeline can be seen as a whole in [3, 45]. Evidence of the shower is gathered from multiple concentric detectors about the collision point, the primary components being the silicon tracker for charged particles and the calorimeter for neutral particles. These can be seen in figure 1. This data is essentially a series of energy deposits associated with three dimensional coordinates, there is no time stamp available because the interaction occurs faster than the electronics can read out. At this point the Particle Flow algorithm is used to reconstruct four vector tracks from the readings [5].

One possibility under investigation in these experiments is that of an extended Higgs sector such as the two Higgs Doublet model (2HDM) [11]. The 2HDM introduces 5 new particles; two CP even (h and H, with, conventionally, $m_h < m_H$), one CP odd (A) and a pair of charged ($H^\pm$) Higgs bosons. These come about from the addition of two complex scalar doublets $\Phi_{1,2}$ from $SU(2)_L$ with the most general gauge invariant renormalisable scalar potential given by:

$$V(\Phi_1, \Phi_2) = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - [m_{12}^2 \Phi_1^\dagger \Phi_2 + \text{h.c.}]$$

$$+ \frac{\lambda_1}{2}(\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger \Phi_2)^2 + \lambda_3(\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2)$$

$$+ \lambda_4(\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) + \frac{1}{2}[\lambda_5 \ (\Phi_1^\dagger \Phi_2)^2 + \text{ h.c.}]$$

$$+ \left[(\lambda_6(\Phi_1^\dagger \Phi_1) + \lambda_7(\Phi_2^\dagger \Phi_2))\Phi_1^\dagger \Phi_2 + \text{h.c.}\right]. \tag{1}$$

Following the hermiticity of the scalar potential, m_{11}^2, m_{22}^2 and $\lambda_{1,\dots 4}$ are real parameters whereas m_{12}^2, $\lambda_{5,6,7}$ can be complex. Assuming the CP-conserving version of the 2HDM, m_{12}^2, $\lambda_{5,6,7}$ and the VEVs of the fields Φ_i are real

parameters. As a consequence of extending the discrete Z_2 symmetry to the Yukawa sector in order to avoid Flavour Changing Neutral Currents (FCNCs) at tree level, $\lambda_{6,7} = 0$, whereas the mass term m_{12}^2 breaks the symmetry in a soft way.

Amongst the many signals that these additional Higgs states could produce, of particular relevance are those involving their cascade decays, whereby a heavier Higgs state decays in a pair of lighter ones or else into a light Higgs state and a gauge boson. This is the case as the former process gives access to the shape of the Higgs potential of the enlarged Higgs sector while the latter channel is intimately related to the underlying gauge structure, which may well be larger than the SM one. The focus of these investigations is the use of decay cascades involving Higgs bosons (either amongst themselves or in association with gauge bosons) in order to understand both the Higgs potential and gauge structure which may exist beyond the SM.

The first part of this investigation will extend and recast the bounds on 2HDM model parameters. The aim being to identify model parameters that there is experimental sensitivity to after the next LHC upgrade. Existing scans were also recast to look at alternative decay channels.

The second part of this investigation looks at the best use of existing tools to process this signal, in particular the optimal way to cluster particles into jets.

In a detector, jet clustering is designed to aid reconstruction of decayed particles. The default choice for jet clustering tends to be one of three algorithms; the anti-kt algorithm [14], the Cambridge-Aachen algorithm [51] and the kt algorithm [21]. These algorithms are often turned to as they are infrared safe, excellent implementations of them are available (see `FastJet` [13]) and they are flexible enough to capture many signals with minimal parameter change.

For a cascade decay in a 2HDM model this process presents a particular challenge owing to the mass difference between the b-quark and heavy Higgs. The decays tend to be collimated and the jets often merge or overlap.

Finally there is an investigation of possible machine learning inspired jet clustering techniques. A recursive algorithm is well suited to clustering objects when the number of groups is not known at outset. Agglomerative algorithms are easier to design in a manner that is infrared safe, as they can recombine soft and collinear emissions in early steps. Spectral clustering is considered as it can be performed in a recursive agglomerative manner.

3 Two Higgs Doublet Model Parameters

This section summarises the work began during a secondment to University Cadi Ayyad, Morocco, in collaboration with Rachid Benbrik and Souad Semlali, for which my supervisor Stefano Moretti is also co-author [10]. The work extends and recasts existing scans of the 2HDM model, to locate parameter space of interest now, and after a planned luminosity and energy upgrade to the LHC. My contribution was focused on efficient data generation by integrating existing tools. This involved modifying tools to remove bottlenecks such as disk reading, creating frameworks to run in parallel without race conditions and finding efficient search strategies.

Parameter regions of interest for the process $pp \rightarrow H \rightarrow ZA \rightarrow l^+l^- b\bar{b}$ and its mirror image $pp \rightarrow A \rightarrow ZH \rightarrow l^+l^- b\bar{b}$ are established by eliminating regions ruled out by theory constraints, previous observations and experimental sensitivity requirements.

Existing studies of the relevant branching ratios are [40, 18, 17, 35]. LHC searches for the complete channels $gg, b\bar{b} \rightarrow A \rightarrow ZH$ and $gg, b\bar{b} \rightarrow H \rightarrow ZA$ have been carried out at both ATLAS [1] and CMS [33, 47], by exploiting leptonic decays of the gauge boson, $Z \rightarrow l^+l^-$ $(l = e, \mu)$, and hadronic decays of the accompanying neutral Higgs state, in particular, H or $A \rightarrow b\bar{b}$ or $\tau^+\tau^-$. Here, we consider the final state $l^+l^- b\bar{b}$ and start from the results of [1] for the $A \rightarrow ZH$ decay in order to obtain the corresponding ones for the complementary channel $H \rightarrow ZA$, and extend the analysis to higher energy and luminosity from planned detector upgrades.

The different transformations of the quark fields under the Z_2 symmetry lead to four structures of Higgs-fermions interactions: in Type-I only one doublet couples to all fermions; in Type-II one of the doublets couples to the up quarks while the other doublets couples to the down quark; in Type-X (or Lepton Specific) one of the doublets couples to all quarks and the other couples to all leptons; in Type-Y (or Flipped) one of the doublet couples to up-type quarks, and to leptons and the other couples to down-type quarks.

3.1 Scan

In this study, we identify the lightest CP-even Higgs boson of the 2HDM as the observed Higgs state at the LHC, with $m_h = 125$ GeV, and assume $\sin(\beta - \alpha) = 1$.

We scan over the following parameter range:

$$m_h = 125 \text{ GeV}, \ \sin(\beta - \alpha) = 1, 0 < m_{12}^2 < 2 \times 10^5 \text{ GeV},$$
$$130 \text{GeV} < m_X < 700 \text{ GeV}, \ m_X \geq m_Y + 100 \text{ GeV},$$
$$m_X, m_Y \text{ chosen at 10 GeV intervals.}$$
$$\tan(\beta) \in \begin{cases} \{1, \ 2, \ 3\}, & \text{if Lepton Specific} \\ \{1, \ 5, \ 10, \ 20\}, & \text{otherwise} \end{cases} \tag{2}$$

The set of values chosen for $\tan(\beta)$, and the masses, align with the choices in [1].

- For the process mediated by $A \to ZH$, we choose $m_X = m_A$, $m_Y = m_H$ and $m_{H^\pm} = m_A$. (Note that this choice is consistent with Ref. [1].)

- For the process mediated by $H \to ZA$, we choose $m_X = m_H$, $m_Y = m_A$ and $m_{H^\pm} = m_H$. (Note this choice is specular to that in Ref. [1].)

While an evident symmetry exists between the two cases, neither the constraints affecting the two processes nor their sensitivity reaches should be expected to be symmetric. On the one hand, the role played by the heavy CP-even and CP-odd Higgs states of the 2HDM in both theoretical and experimental limits is different, owing to their different quantum numbers (and hence couplings). On the other hand, their production and decay rates at the LHC are different despite leading to the same final states, including residual differences due to width effects entering their normalisation (but, as mentioned, not their kinematics), since, e.g., the A state does not decay to W^+W^- and ZZ pairs while the H state does and, conversely, the A state decays to Zh while the H state does not. However, in the alignment limit used here these decay channels are closed.

3.1.1 Theoretical constraints

Within these ranges there are several theoretical and experimental constraints for the parameter points of the 2HDM to pass, discussed below.

- Unitarity: various scattering processes require that unitarity is conserved at the tree-level at high energy. The unitarity requirements in the 2HDM have been studied in [32, 23, 4].

- Perturbativity constraints are given in [32, 12].

- Vacuum stability requires the scalar potential to be bounded from below [25].

In practice the theoretical constraints are automatically checked by the program 2HDMC [22]. The program can determine if a selected parameter combination is valid.

The choice of $m_{12}^2 = m_A^2 \tan(\beta)/(1 + \tan(\beta))^2$ enables us to reconstruct the exclusion limits at 95% confidence level (CL) given in Ref [1]. However, this choice does not actually allow to satisfy theoretical constraints in all four types of 2HDM. In contrast, our choice of m_{12}^2 above aims to simultaneously satisfy as many theoretical constraints as possible while affording one with significant parameter space amenable to experimental investigation.

2HDMC is a well engineered, fast library; however owing to the range of possible values in m_{12}^2 some curve fitting of the valid points is required to augment the MC sampling. Points that satisfy the most constraints are fitted to a polynomial and values of m_{12}^2 near the surface of the polynomial are sampled further.

3.1.2 Experimental constraints

In addition to theoretical constraints experimental constrains need to be accounted for.

- EW Precision Observables (EWPOs) [27], such as the oblique parameters S and T [42, 24], require a level of degeneracy between the charged Higgs boson state and one of the heavier neutral Higgs bosons. Here, we assume $m_{H^\pm} = m_A$ or m_H, as appropriate (see below), so that the T parameter exactly vanishes in the alignment limit.

- Exclusion limits at 95% CL from Higgs searches at colliders (LEP, Tevatron and LHC) via HiggsBounds, version 5.3.2 [7, 6, 8] are enforced. Furthermore, the ATLAS Collaboration has set an upper limit at 95% CL on the production cross section σ of the A state times its decay BR into $ZH \to l^+l^-b\bar{b}$, i.e., $\sigma(A) \times \text{BR}(A \to ZH \to l^+l^-b\bar{b})$ [1], that is not included in this tool, hence we have accounted for it separately.
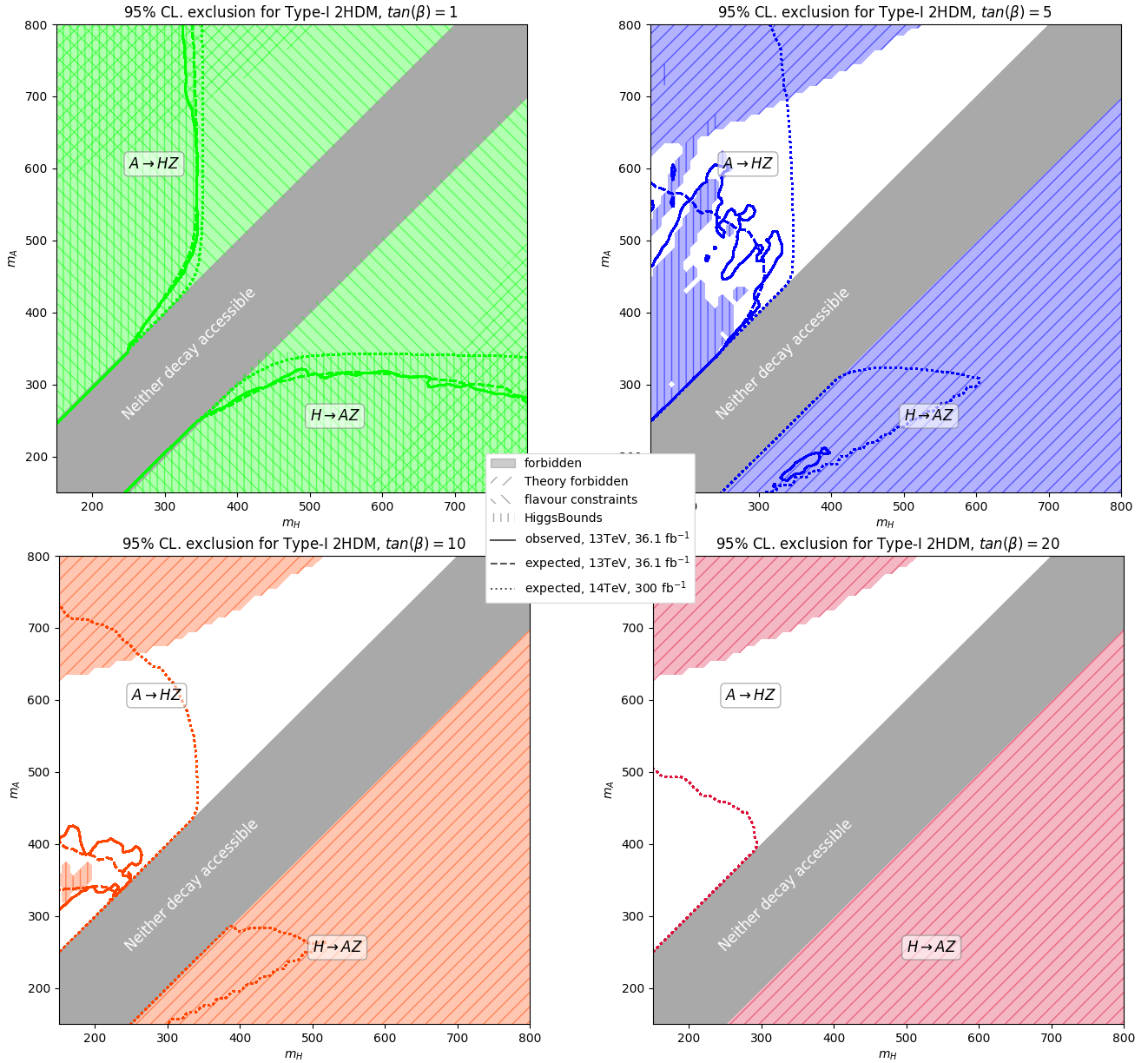
4

Figure 2: Exclusion limits at 95% CL in Type-I. The lines denoting expected and observed exclusion limits do not appear at all on some plots when the prediction never exceeds the expected or observed limit.

- Constraints from the Higgs boson signal strength measurements are automatically satisfied as we assume $\sin(\beta - \alpha) = 1$.

- Constraints of flavour physics observables, namely, $B \to X_s \gamma$, $B_{s,d} \to \mu^+ \mu^-$ and $\Delta m_{s,d}$ [27].

Comparing to observed and expected exclusion limits requires obtaining branching ratios and production cross sections. The relevant branching ratios, $A \to ZH$, $H \to ZA$, $A \to b\bar{b}$ and $H \to b\bar{b}$ are calculated with 2HDMC [22]. The production cross sections of the heavy CP-even (H) and CP-odd (A) Higgs bosons, at Next-to-Next-to-Leading Order (NNLO) in QCD, for both $gg \to H, A$ and $b\bar{b} \to A, H$, at the Centre-of-Mass (CM) energies of 13 TeV and 14 TeV, are calculated using SusHi [31, 30, 29, 28]. Of all the calculations required for the scan this one required most compute time, therefore some optimisations to the code were investigated. Firstly, much of the compute time was on unneeded disk reading and writing, this was circumnavigated by piping the input and output, such that the data that would have gone to the disk stayed in the ram. This produced a 30% speed increase. Secondly, the program was not natively vectorised. To allow the computation to be done in parallel without race conditions and with minimal possibility of introducing a new bug a python script to split in input and launch the program in
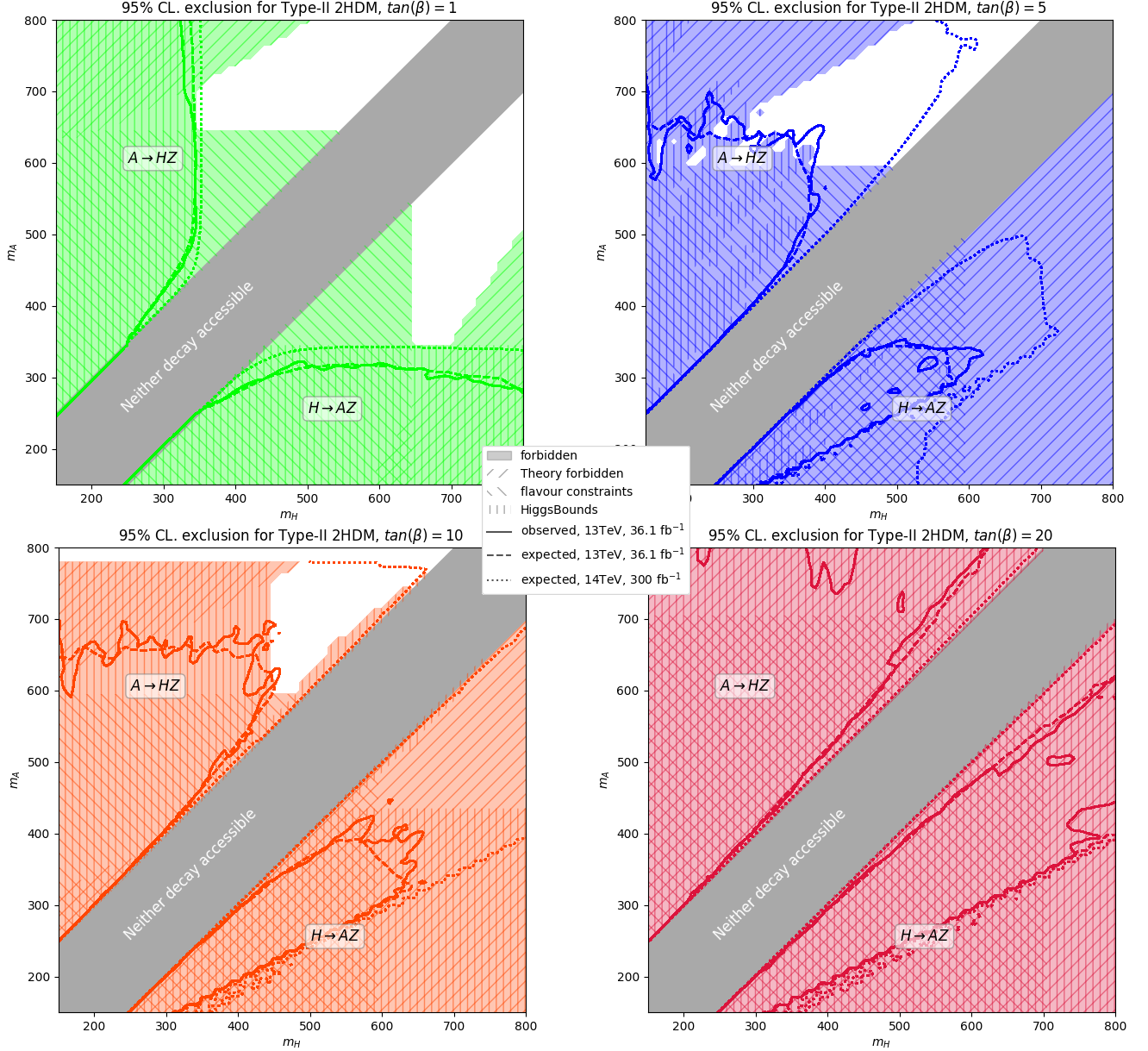
Figure 3: Like in Fig. 2 but for Type-II.

multiple, non interacting, threads was developed.

2HDMC code also includes an interface to HiggsBounds, which is used to apply the aforementioned exclusion limits at 95% CL from Higgs searches at LEP, Tevatron and LHC.

The first part of this study deals with the two production and decay processes $pp \to H(A) \to ZA(H) \to b\bar{b}l^-l^+$. The observed and expected confidence limits for all four types of Yukawa couplings in the 2HDM are produced at $\sqrt{s} = 13$ TeV, with an integrated luminosity, L, of 36.1 fb^{-1}, by combining our calculations with the data from Ref. [1]. In the second part, we rescale the expected exclusion limit to the CM energy of $\sqrt{s} = 14$ TeV, with an integrated luminosity of 300 fb^{-1}, by calculating the so called 'upgrade factor' for both signals and backgrounds, while retaining the acceptance and selection efficiencies of the analysis at the lower $\sqrt{s}$ value. The change in energy will naturally affect signals and backgrounds differently. We treat the former by using SusHi (as intimated) and the latter by using MadGraph5, version 2.6.4 [2]. (For completeness, the background is considered to be any reducible or irreducible SM process that creates a pair of b-jets plus a pair of electrons or muons, as in Ref. [1].)
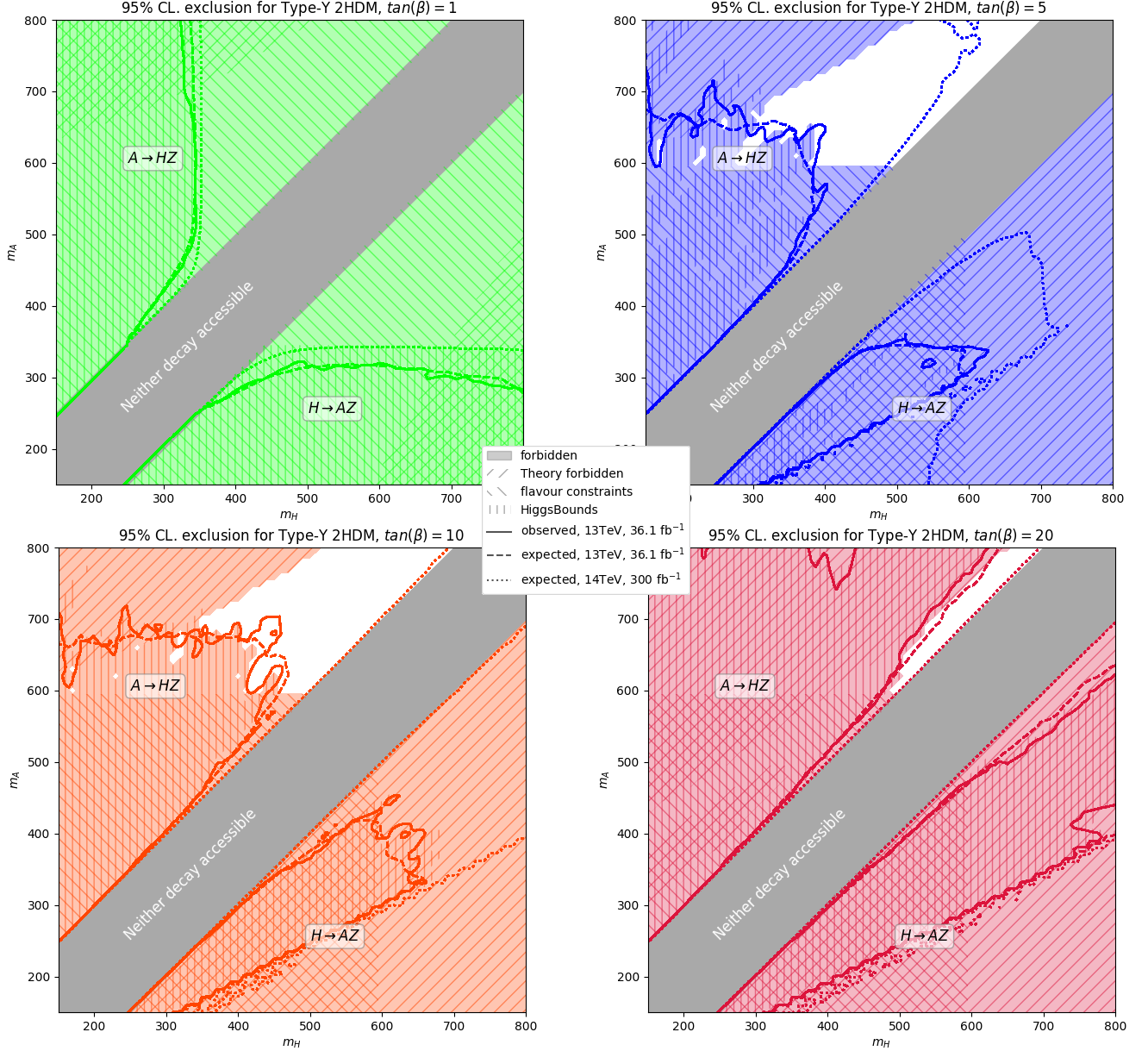
Figure 4: Like in Fig. 2 but for Type-Y (Flipped).

3.2 Numerical results

After performing a scan over the parameter space delimited by Eq. (2), we compare the prediction of the model with the observed and expected limits given in Ref. [1]. If the prediction exceeds the observed limit, then the parameter combination is excluded. When the prediction exceeds the expected limit, we anticipate that the signal would be visible above background given the energies and luminosities available, hence, the experiment is sensitive to these parameters.

Figs. 2 to 5 illustrate the outcome the scan for each Yukawa type, $\tan(\beta)$ and mass combination (m_H, m_A). Each figure provides results for one choice of Yukawa couplings and each frame in each figure provides results at one value of $\tan(\beta)$. In the top left of each plot, where $m_A > m_H + 100$ GeV, the decay $A \to ZH$ is considered while in the bottom right of each plot, where $m_H > m_A + 100$ GeV, the decay $H \to ZA$ is considered. The corridor along the diagonal between these regions is coloured grey to indicate that neither decay is accessible. If a combination of parameters is forbidden by theory, HiggsBounds or flavour constraints then the corresponding area is filled with solid colour, conversely, white areas pass all these checks and so are of interest. The hatching over the solid colour is used to indicate which of the checks causes the corresponding parameter combination to fail. There are three
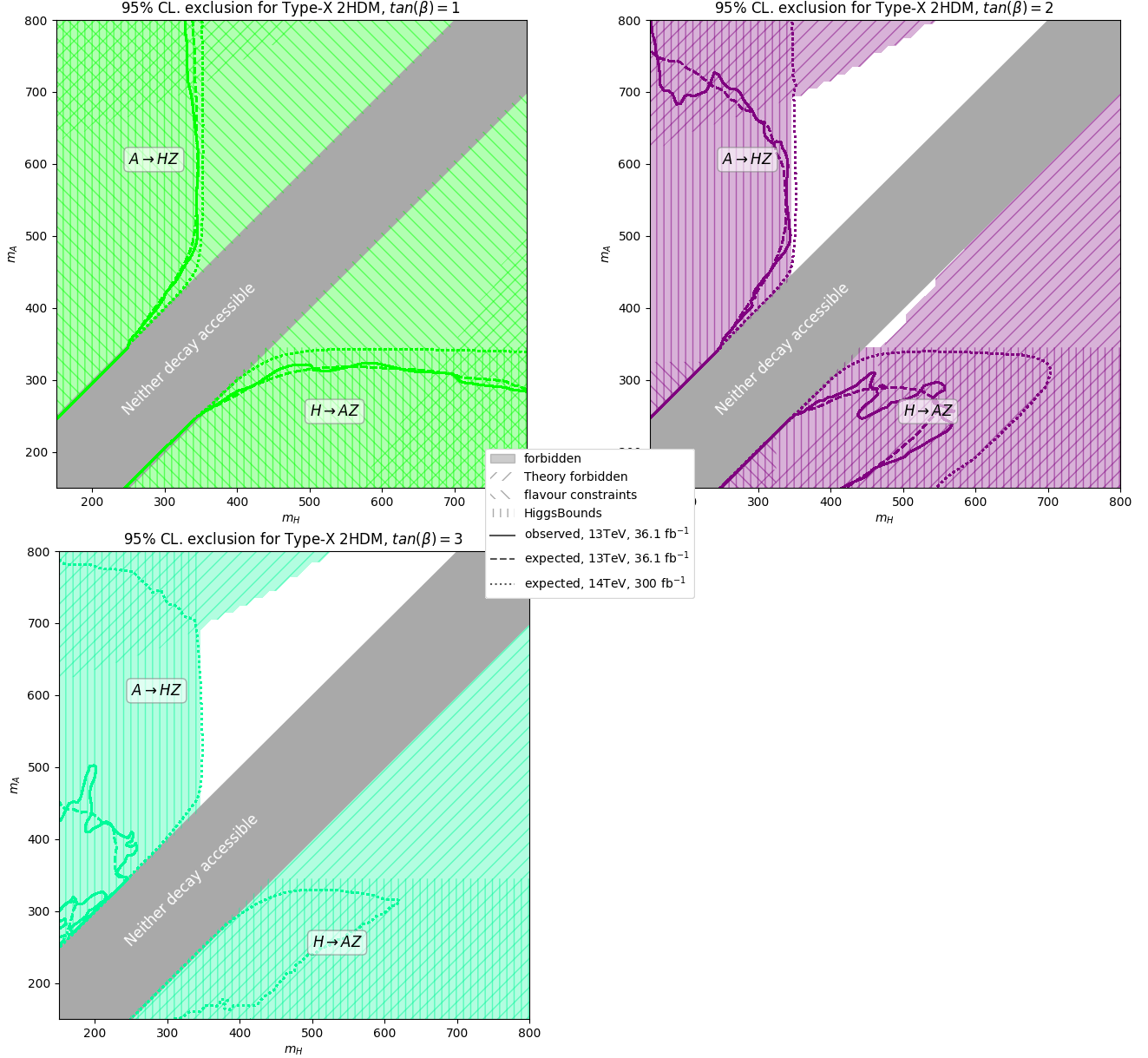
Figure 5: Like in Fig. 2 but for Type-X (Lepton specific).

boundary lines drawn over the plots: these are the observed and expected 95% CLs for the ATLAS detector in its present state, 13 TeV and 36.1 fb^{-1}, plus the expected 95% CL for an upgraded LHC and ATLAS detector at 14 TeV and 300 fb^{-1}[1]. The model predictions exceed the 95% CL inside the curve.

3.3 Conclusions

In summary, we have revisited an experimental analysis of the ATLAS Collaboration of the production and decay process $gg, b\bar{b} \to A \to ZH \to l^+l^-b\bar{b}$ performed at Run 2 with 36.1 fb^{-1} of luminosity, which had been interpreted in terms of exclusion limits over the parameter space of the four types of the 2HDM, wherein the lightest Higgs state is identified with the SM-like Higgs boson discovered during Run 1 at the LHC with mass 125 GeV. Upon validating the ATLAS interpretation in our framework, though, we have discovered that their (fixed) choice of m_{12}, a mass parameter in the 2HDM Lagrangian that softly breaks an underlying Z_2 symmetry of the 2HDM to avoid FCNCs, yields parameter space configurations which are ruled out by theoretical requirements of model consistency.

[1]We neglect here to consider the case of $\sqrt{s} = 13$ TeV and $L \approx 140$ fb^{-1}, as it only improves marginally the present situation yet it would be make the plots far too crowded.

$\tan(\beta)$	1	5	10	20
Type-I	Flavour constraints	Some masses	Many masses	Low sensitivity
Type-II	Flavour constraints	Some masses after upgrade	Some masses after upgrade	Theory constraints
Flipped	Flavour constraints	Some masses after upgrade	Some masses after upgrade	Theory constraints

$\tan(\beta)$	1	2	3
Lepton specific	Flavour constraints	Excluded by HiggsBounds	Excluded by HiggsBounds

Table 1: Table summarising the findings in Figs. 2 to 5. An overview of the possibility of each Yukawa type and value of $\tan(\beta)$ is given. Entries in red indicate that the combination has little or no mass combinations that are not forbidden while those in blue represent available parameter space accessible presently at Run 2 or after the upgrade of Run 3.

Hence, we have allowed this parameter to vary freely and subject the ensuing parameter space configurations to both the aforementioned theoretical constraints as well as those emerging from past and present experiments, thereby redrawing the actual sensitivity of such an experimental search to all four Yukawa types of the 2HDM, as a function of $\tan(\beta)$. In doing so, we have have also forecast the potential sensitivity of this channel to the 2HDM parameter space at the end of Run 3, assuming increased energy to 14 TeV and luminosity to 300 fb^{-1}. This revealed some extended coverage of the 2HDM Type-I, -II and -Y (but not -X), especially for intermediate $\tan(\beta)$ values (say, between 5 and 10), with m_A up to 800 GeV and m_H up to 700 GeV. This is somewhat beyond what is presently covered, i.e., up to 150 GeV or so in mass of either Higgs state, so as to justify further searches for this signature at the next stage of the LHC. Finally, we have recast the sensitivity of this analysis onto that of the channel $gg, b\bar{b} \to H \to ZA \to l^+l^-b\bar{b}$. However, we have found that the complementary parameter space accessible this way (i.e., $m_H \geq m_A + m_Z$) is actually entirely excluded already by existing theoretical and/or experimental constraints, so as to conclude that it is not warranted to pursue further this channel at the LHC, at least, not with a view to interpret it in the context of the standard four Yukawa types of the 2HDM[2].

4   Seeking cascade decays

The work in the section was produced in collaboration with Billy Ford, Shubhani Jain, Amit Chakraborty, Stefano Moretti, Emmanuel Olaiya, Claire Shepherd-Themistocleous and Srinandan Dasmahapatra [15]. Working with various established jet clustering techniques the objective was to determine what experimental particle and jet cuts would be needed to observe cascade decays involving light Higgs. My contribution to this was in developing alternative event processing and jet clustering tools that validated the results obtained from the standard libraries and allowed in depth inspection of points of confusion.

Remaining in the 2HDM, as in the previous section, we consider an alternative scenario involving a light Higgs. The heavier CP-even Higgs (H) is considered to be a SM Higgs at 125GeV, then should the light CP-even Higgs be < 60GeV the $H \to hh$ decay is resonant. As with the SM Higgs the dominant decay mode for the light Higgs is bottom-antibottom quark pairs [40, 18], so that the predominant final state of $H \to hh$ is four (anti)quarks[3]. Detecting this signal would require identifying these four b-quarks by clustering their decay products into jets. Various algorithms are establish for jet clustering, here, both the popular anti-kt [14] and Cambridge-Aachen algorithms [51] and the more recent variable-R method [36]. We will investigate whether such variable-R jets can improve the ability to observe particular 2HDM-II four b-jet final state topologies.

4.1   Monte Carlo and tagging

We consider two sample benchmark points, namely, $m_h = 40$ GeV and 60 GeV, while fixing $m_H = 125$ GeV, within a 2HDM Type-II (2HDM-II henceforth), which have been tested against theoretical and experimental constraints by using 2HDMC [22], HiggsBounds [8], HiggsSignals [9] as well as checking flavour constraints with SuperISO [39].

---

[2]We finally note that analyses similar to Ref. [1] performed by the CMS Collaboration exist [33, 47]. We have not used these for two reasons. On the one hand, they did not convey all the information necessary to make extrapolations to higher energies. On the other hand, they did not afford one with significantly different sensitivity to the 2HDM at present energies than what achieved by the ATLAS analysis [1] that we have adopted as benchmark.

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

Using Madgraph [2] to generate the partonic process, and Pythia [50] to shower, $\mathcal{O}(10^5)$ of $pp \to H \to hh \to b\bar{b}b\bar{b}$ events are generated, with $\sqrt{s} = 13$ TeV.

To label a jet as a b-jet Monte Carlo (MC) information is used. The b-quarks that we aim to capture in jets are identified and assigned to the closest jet in angular (ΔR) distance unless the closest jet is further away than the cut off parameter R, in which case the quark is abandoned. To mimic a more realistic tagging efficiency a 25% false negative rate is added to this, and a 10% false positive rate, in the form of c-quark jets appearing as b-jets, is applied to this selection.

4.2 Cutflow

Typical cuts used for Higgs decay chains, such as those seen in [49], tend to be restrictive with the objective of ensuing fully hadronic signature. In this example the cuts given are;

1. Remove all final state particles with a $p_T < 0.5$ GeV and $|\eta| > 2.5$.

2. Perform jet reconstruction and b-tagging in `FastJet` [13], with specified clustering algorithm and R.

3. Remove b-jets with $p_T < 30$ GeV ($*$).

4. Remove b-jets whose energy content due to electrons(muons) is greater than 90(80)%.

5. Remove b-jets whose energy content due to neutral hadrons and photons is greater than 90%.

6. Remove b-jets with zero charged hadron energy.

7. Remove b-jets with only one constituent.

For the process considered here, these cuts remove the vast majority of the signal. Taking a $p_T < 30$GeV cut on the jet culls many events, see figure 6. A more permissive set of jet cuts is seen in [48]. In [48] p_T of the leading b-jet is required to have > 20GeV, and others $p_T > 15$GeV. This reduced cutflow would require the implementation of a trigger with p_T thresholds of 15, 10, 10, 10GeV for the four (p_T ordered) b-jets. These p_T thresholds are sufficiently below our jet p_T requirement to enable 100% efficiency in the trigger.

A comparison of the b-jet multiplicity of events with typical cuts, permissive cuts and no cuts can be seen in 7. as can be seen with the comparison given in figure 7.
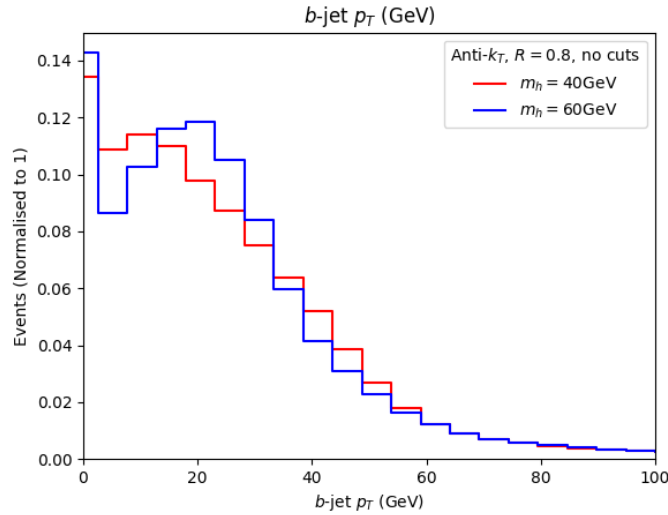
Figure 6: Transverse momentum distribution of all b-jets with no cuts imposed on the jets or on the final state hadrons. We have used here $R = 0.8$, though the pattern is similar for $R = 0.4$. Here, we have used the anti-k_T jet clustering algorithm.

From this point on the permissive cuts will be employed as they seem necessary for an attempt to retain this signal.
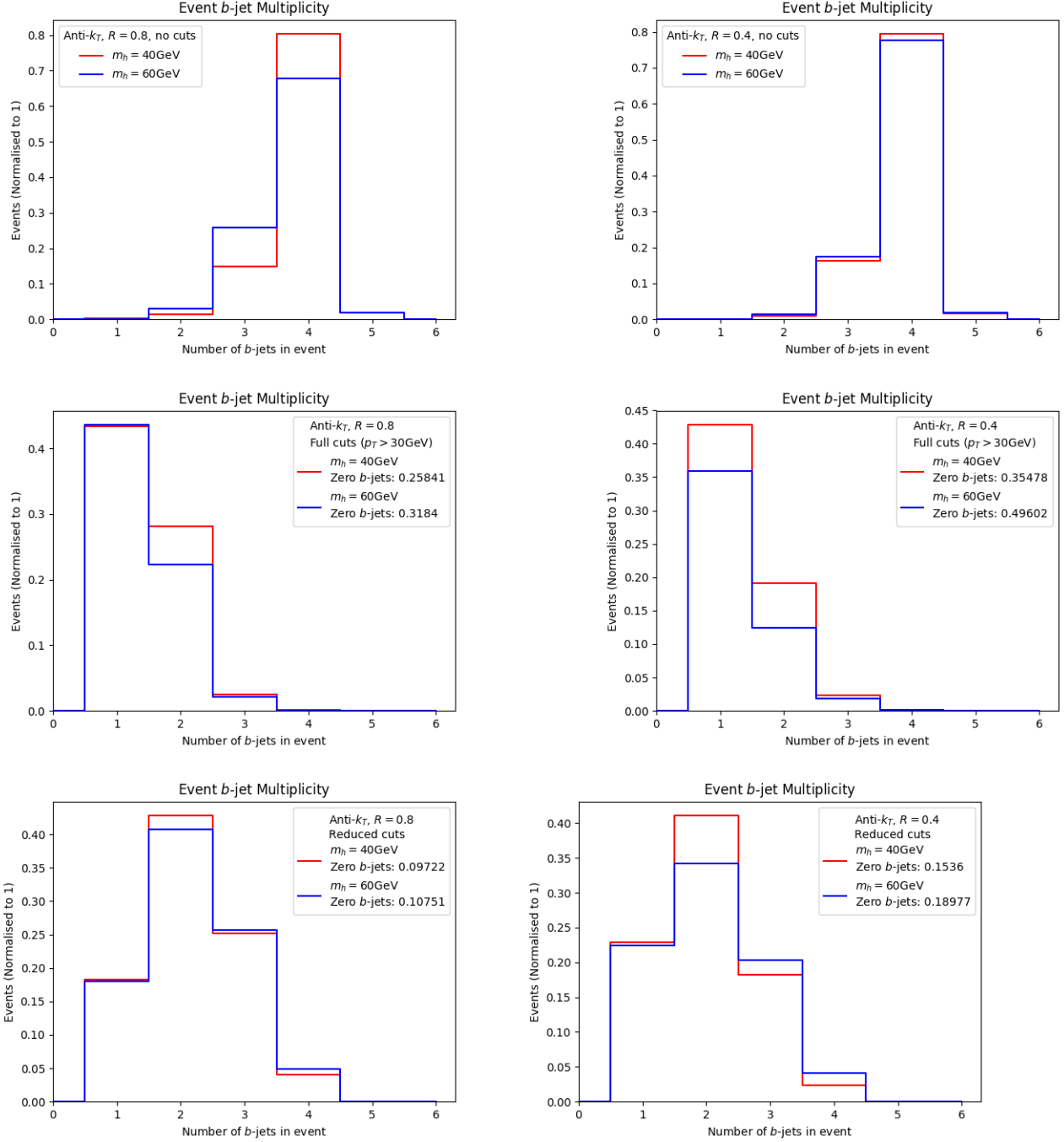
Figure 7: The distributions of b-jet multiplicities under various cutflows. The left column presents anti-kt jets where $R = 0.8$ and the right column has $R = 0.4$. The first row is the distribution with no cuts on the inputs or the jets, the second row shows the impact of typical cuts (as described in sec 4.2), the third row shows the results of the possible relaxation (suggested in sec 4.2). Note that events containing zero b-jets are not plotted but included in the legend, in the case of no cuts being applied all events contain at least one b-jet.

4.3 Variable-R clustering

The standard jet clustering algorithms, anti-kt and Cambridge-Aachen, both depend on a interparticle distance, $d_{ij} = \min(p_{Ti}^n, p_{Tj}^n)\Delta R_{ij}^2$, and a beam distance, $d_{Bi} = p_{Ti}^n \Delta R^2$. The algorithm is agglomerative, to begin all interparticle distances and beam distances are calculated, the smallest value is found. If the smallest value is an interparticle distance then those particles are combined, otherwise, if the smallest value is a beam distance then the

corresponding particle is deemed a complete jet and removed. This process repeats until all particles are removed.

The variable-R determines how far from it's nearest neighbour a combined particle must be such that it is removed and deemed a jet.

Variable-R is a slight divergence from this that allows the value of R to change with the p_T of the particle. R becomes $R_{\text{eff}}(p_T) = \frac{\rho}{p_T}$, where ρ is a tunable parameter.

Figure 8: Left: the ΔR distribution between the two b-partons originating from the same h. Right: the p_T distribution of the light Higgs boson h originating from H decay. No (parton level) cuts have been enforced here.
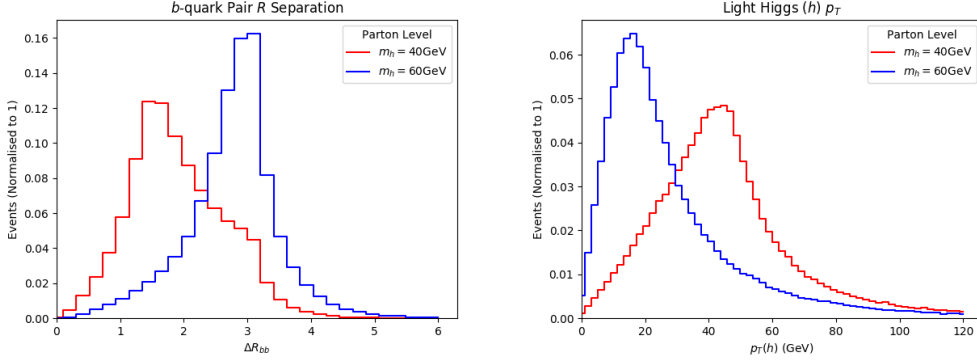
An indication of why this might be desirable can be seen by looking at the angular separation of b-quarks originating from the same h, seen in figure 8. There is a considerable range of angular separations for each decay, in general, the angular separation between the decay products a and b in the resonant process $X \to ab$ can be approximated as $\Delta R(a, b) \sim \frac{2m_X}{p_T^X}$.

Indeed the variable-R algorithm does produce notably sharper mass peaks, as can be seen in figure 9.

To quantify this improvement the significance of the signal over the background is calculated for two values of (integrated) luminosity, e.g., $\mathcal{L} = 140$ and 300 fb^{-1}. The event rate (N) for the various processes is given by:

$$N = \sigma \times \mathcal{L}. \tag{3}$$

With these event rates, the significance, Σ, is given by;

$$\Sigma = \frac{N(S)}{\sqrt{N(B_{b\bar{b}b\bar{b}}) + N(B_{Zb\bar{b}}) + N(B_{t\bar{t}})}}. \tag{4}$$

The results of this are shown in tables. 2–3. It is clear that that the variable-R approach works better than the 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.

	Var-R, $\rho = 20$ GeV	$R = 0.4$	$R = 0.8$
40 GeV	0.358	0.061	0.160
60 GeV	6.677	2.074	3.138

Table 2: Final Σ values calculated for signal and backgrounds for $\mathcal{L} = 140$ fb^{-1} upon enforcing the reduced cuts plus the mass selection criteria $|m_{bbbb} - m_H| < 20$ GeV and $|m_{bb} - m_h| < 15$ GeV for the various jet reconstruction procedures.

	Var-R, $\rho = 20$ GeV	$R = 0.4$	$R = 0.8$
40 GeV	0.524	0.089	0.234
60 GeV	9.775	3.036	4.594

Table 3: Final Σ values calculated for signal and backgrounds for $\mathcal{L} = 300$ fb^{-1} upon enforcing the reduced cuts plus the mass selection criteria $|m_{bbbb} - m_H| < 20$ GeV and $|m_{bb} - m_h| < 15$ GeV for the various jet reconstruction procedures.
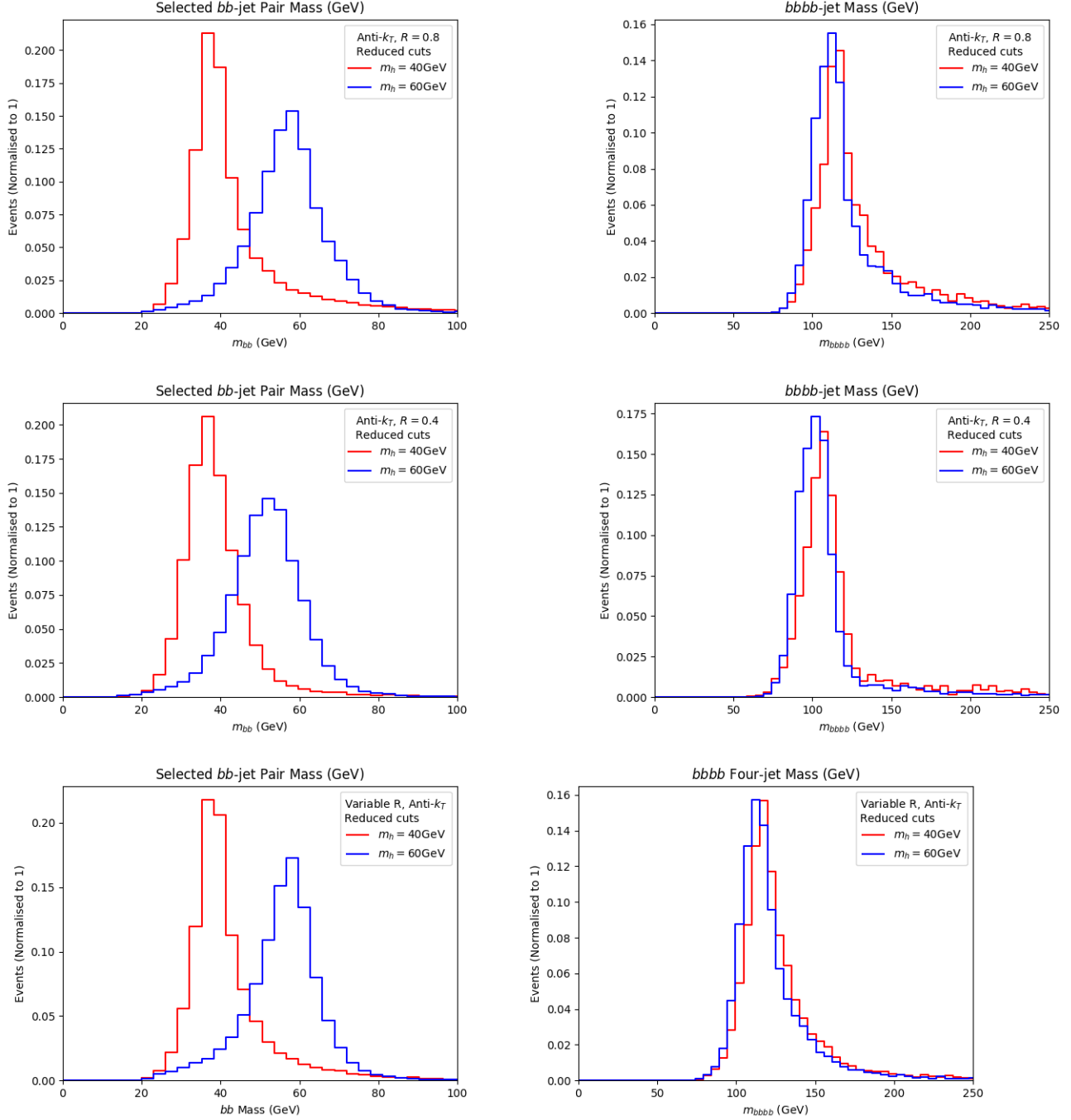
Figure 9: The distributions in invariant mass of two (left panels) and four (right panels) b-tagged jets, in presence of the reduced p_T cuts on the b-jets. The first two rows are anti-kt clusterings, with R equal to 0.8 and 0.4 respectively, The final row shows the variable-R algorithm.

4.4 Conclusions

In this paper, we have assessed the potential scope of the LHC experiments in accessing BSM Higgs signals induced by cascade decays the following prototypical production and decay channel: $gg \to H \to hh$, where H is the SM-like Higgs state and h is a lighter BSM Higgs state. The latter is extremely difficult to establish at the LHC, owing to the substantial hadronic background.

The first message we deliver is that, with current p_T limitations 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 present a reduced cut flow, based on existing $b\bar{b}\mu^+\mu^-$ analyses, and show that this indeed provides a window into

final states of $gg \rightarrow H \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ decays with $m_H = 125$ GeV and $m_h < \frac{m_H}{2}$.

Additionally, we test a variable-R reconstruction approach, and show a significant improvement in signal yield, as well as signal-to-background rates.

5   Jet clustering

This next section considers a completely novel approach to jet clustering, spectral clustering.

As mentioned previously, the default choice for jet clustering tends to be one of there algorithms; the anti-kt algorithm [14], the Cambridge-Aachen algorithm [51] and the kt algorithm [21]. They have been the default choice for some time because they have a number of desirable properties. They are infrared safe, excellent implementations of them are available (see `FastJet` [13]) and it is flexible enough to capture many signals with minimal parameter changes. These algorithms are recursive and agglomerative. A recursive algorithm is well suited to clustering objects when the number of groups is not known at outset. Agglomerative algorithms are easier to design in a manner that is infrared safe, as they can recombine soft and collinear emissions in early steps.

Finding a clustering method that compares favourably to these algorithms is challenging. Spectral clustering is a candidate that has had considerable success in other studies. In fluid dynamics spectral clustering has been used to identify the motion of vortices [26], finding that it is possible to successfully identify the vortex structures in cases with less data available. It was also seen that spectral clustering was proficient at determining the correct number of clusters to be found in the fluid. To reduce the risk of blackouts, power grids may be subdivided into 'islands'. The ideal allocation is found by minimising power flow between islands, and it was shown in [44] that spectral clustering can produce a good solution in less time than other algorithms commonly used for this problem.

To the authors knowledge this clustering algorithm has not yet been applied to jet physics, however, given its recursive, agglomerative form it could be a good fit. This is a non standard approach to spectral clustering; while the embedding step relaxes an optimisation objective the agglomerative step does not, this is described in more depth in the next section.

5.1   Theory of spectral clustering

Spectral clustering begins with embedding all data in a new space using the eigenvectors. In this new space the clustering a standard clustering will be performed. Embedding is the result of relaxing the criteria that would precisely to identify infinitely separated groups to estimate the members groups that may have some connection. This is best described in [38], a short summary is given here.

Let us imagine a graph, disconnected in n clusters. The initial aim is to identify which of the disconnected components each point belongs to. The objective used for a good clustering will be the NCut objective;

$$\text{NCut} = \sum_k \frac{W(A_k, \bar{A}_k)}{|A_k|} \tag{5}$$

Where $W(A_k, \bar{A}_k)$ measures the strength of the connections that must be broken in order to separate the cluster A_k from the graph and $|A_k|$ corresponds to the number of elements in A_k.

Membership of cluster A_k will be determined by the indicator vector h_k;

$$h_{i,k} = \begin{cases} 1/\sqrt{|A_k|}, & \text{if point } i \in A_k \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $|A_k|$ is the number of points in set A_k. A pair of points is assigned an affinity, $a_{i,j} = a_{j,i}$, of either 1 for connected i and j or 0 if points i and j are unconnected.

The graph is represented by the graph Laplacian, a square matrix with as many rows and columns as there are points. On ith diagonal it holds the value $\sum_j a_{i,j}$, which is the degree of the vertex i. Off the diagonal it has the negative affinity; $-a_{i,j}$.

$$L = \begin{pmatrix} z_0 & -a_{0,1} & -a_{0,2} & \cdots \\ -a_{0,1} & z_1 & -a_{1,2} & \\ -a_{0,2} & -a_{1,2} & z_2 & \\ \vdots & & & \ddots \end{pmatrix} \tag{7}$$

Notice that this is a real symmetric matrix, and therefore all its eigenvalues are real.

14

Considering just one cluster we find that when the Laplacian is multiplied by two indicator vectors the result is the function that NCut seeks to minimise for that cluster.

$$h'_k L h_k = \frac{1}{|A_k|} \sum_{i \in A_k, j \in A_k} \left(\delta_{i,j} \sum_l a_{l,i} - a_{i,j} \right) = \frac{W(A_k, \bar{A}_k)}{|A_k|} \tag{8}$$

Stack the indicator vectors into a matrix;

$$h'_k L h_k = (H'LH)_{kk} \tag{9}$$

and the NCut aim described earlier becomes the trace;

$$\text{NCut}(A_1, A_2, \dots A_n) \equiv \frac{1}{2} \sum_{i=1}^n \frac{W(A_i, \bar{A}_i)}{|A_i|} = \text{Tr}(H'LH) \tag{10}$$

Where $H'H = I$. Trace minimisation in this form is done by finding the eigenvectors of L with smallest eigenvalues.

Generalising this to a graph that is not disconnected only requires relaxing the requirements on the form of the indicator vectors; h_k.

Once these indicator vectors have been created the position of the points in the embedding space is determined. Each indicator vector has as many elements as there are points to be clustered, so the coordinates of a point are the corresponding elements or the indicator vectors. Using the positions in embedding space the points can be gathered agglomeratively, this part is not a relaxation of a well defined objective.

5.1.1 Impact of p_T factors

In the case of reconstructed particles, a possible input to the affinity measure is the p_T of the particles. A format commonly used for the distance is;

$$d_{i,j} = \min(p_{T,i}^{2x}, p_{T,j}^{2x}) \Delta R_{i,j} \tag{11}$$

where $\Delta R_{i,j}$ is an angular distance. This is discussed from a practical standpoint in section 5.2.2, from a theoretical standpoint it is important to investigate the impact of the p_T factor on the eigenvalue equations. There are a number of different ways of converting a distance measure into an affinity, for now only one will be considered $a_{i,j} = 1/d_{i,j}$.

Starting with the Laplacian a relation between the size of the eigenvalues and the affinities in each group can be uncovered. Let us take $z_i = \sum_{j \neq i} a_{i,j}$ where it is understood that by default any sum goes over all node (i.e. particle) indices. From this point onward it is also assumed that all $a_{i,i} = 0$, so such sums can omit the $i \neq j$. With this, equation 7 can be rewritten as

$$L = \begin{pmatrix} z_0 & -a_{0,1} & -a_{0,2} & \cdots \\ -a_{0,1} & z_1 & -a_{1,2} & \\ -a_{0,2} & -a_{1,2} & z_2 & \\ \vdots & & & \ddots \end{pmatrix} \tag{12}$$

Now imagine a group that includes the first q nodes, its relaxed indicator vector approximately contains two values; $-Q_0$ for nodes 0 to q inclusive and Q_1 for nodes $q+1$ to n inclusive, where $Q_{0/1} > 0$. As the ordering of the nodes is arbitrary (the rows and columns can have any order) conclusions drawn for this hold for any group containing q nodes. This must be approximate because we do not require that the q nodes are perfectly separated (the model is relaxed), there for the eigenvalue equation is a relaxation.

The eigenvalue equation has the form

$$\begin{pmatrix} z_0 & -a_{0,1} & -a_{0,2} & \cdots \\ -a_{0,1} & z_1 & -a_{1,2} & \\ -a_{0,2} & -a_{1,2} & z_2 & \\ \vdots & & & \ddots \end{pmatrix} \begin{pmatrix} -Q_0 \\ -Q_0 \\ \vdots \\ Q_1 \end{pmatrix} \approx \lambda \begin{pmatrix} -Q_0 \\ -Q_0 \\ \vdots \\ Q_1 \end{pmatrix} \tag{13}$$

.

The simultaneous equations that come from this can be simplified into two forms. The first q, where $i \leq q$, take the form

$$-Q_0 \lambda \approx Q_0 \sum_{j \leq q} a_{i,j} - Q_0 z_i - Q_1 \sum_{j > q} a_{i,j} \tag{14}$$

15

$$\lambda \approx \left(1 + \frac{Q_1}{Q_0}\right) \sum_{j>q} a_{i,j} \tag{15}$$

. In a similar manner the remaining simultaneous equations, where $k > q$ have the form

$$\lambda \approx \left(1 + \frac{Q_0}{Q_1}\right) \sum_{j \leq q} a_{k,j} \tag{16}$$

. In both cases the sum is over affinities that attach to the node associated with the row and cross the cluster boundary.

Clearly the size of the eigenvalue will be related to the size of the affinities that cross the boundary. More subtly, if the number of elements in each group differs strongly then the values of Q_0 and Q_1 will need to differ more to compensate. Approximating all the affinities to the same value

$$\left(1 + \frac{Q_1}{Q_0}\right) q a = \left(1 + \frac{Q_0}{Q_1}\right)(n - q)a \tag{17}$$

$$\left(1 + \frac{Q_1}{Q_0}\right) = \left(1 + \left(\frac{Q_1}{Q_0}\right)^{-1}\right) \frac{q - n}{q} \tag{18}$$

it can be seen that as $\frac{q-n}{q}$ deviates from 1 the solution grows the left side faster than it shrinks the right. This will also tend to increase the value of λ.

So essentially a split associated with a small eigenvalue will minimise the value of the crossing affinities and create groups with equal number of members. Returning to the idea of the influence of p_T, this will have impact on the former condition but not the latter. Taking $a_{i,j} \sim 1/\min(p_{T,i}^{2x}, p_{T,j}^{2x})$ with $x > 0$ affinities involving a low p_T particle will be larger, so the placement of soft emissions will be important. Alternatively, with $x < 0$ affinities involving one high p_T particle will be larger, so the placement of hard emissions will be most important.

This has found the role of p_T in an unnormalised laplacian, the same steps can be taken for the **symmetric laplacian**; $L = D^{-1/2}(D - A)D^{-1/2}$.

$$L_{\text{symm}} = \begin{pmatrix} 1 & -a_{0,1}(z_0 z_1)^{-1/2} & -a_{0,2}(z_0 z_2)^{-1/2} & \cdots \\ -a_{0,1}(z_0 z_1)^{-1/2} & 1 & -a_{1,2}(z_1 z_2)^{-1/2} & \\ -a_{0,2}(z_0 z_2)^{-1/2} & -a_{1,2}(z_1 z_2)^{-1/2} & 1 & \\ \vdots & & & \ddots \end{pmatrix} \tag{19}$$

Considering the same group of the first q rows the forms of the simultaneous equations are; for $i \leq q$

$$\lambda \approx 1 + z_i^{-1/2} \left(\frac{Q_1}{Q_0} \sum_{j>q} a_{i,j} z_j^{-1/2} - \sum_{j \leq q} a_{i,j} z_j^{-1/2} \right). \tag{20}$$

Notice that the first sum is over affinities that cross out of the group, divided by the root of the connectedness of the outside node, (z_j can be seen as the connectedness of node j), the second sum is over affinities contained within the first q nodes, also divided by the root of their connectedness.

Then for $k > q$

$$\lambda \approx 1 + z_k^{-1/2} \left(\frac{Q_0}{Q_1} \sum_{j \leq q} a_{k,j} z_j^{-1/2} - \sum_{j>q} a_{k,j} z_j^{-1/2} \right). \tag{21}$$

Again the first sum is over affinities that cross the group, and now the second sum is over affinities not in the first q nodes. Now finding the impact of p_T looks rather more complicated, so it will be addressed in stages. Firstly consider $z_j = \sum_i a_{i,j} \sim \sum_i 1/\min(p_{Ti}^{2x}, p_{Tj}^{2x})$, it's behaviour splits into 4 categories. Consider $N_j(p_T)$ to be roughly the p_T of a node that node j is connected to (Neighbour). Also, let nodes on average have the equivalent of c connections, where $0 \leq c \leq n$.

	$p_{Tj} < N_j(p_T)$	$p_{Tj} > N_j(p_T)$
$x > 0$	$z_j \sim c p_{Tj}^{-2x}$ therefore $z_j^{-1/2} \sim c^{-1/2} p_{Tj}^x$	$z_j \sim c N_j(p_T)^{-2x}$ therefore $z_j^{-1/2} \sim c^{-1/2} N_j(p_T)^x$
$x < 0$	$z_j \sim c N_j(p_T)^{-2x}$ therefore $z_j^{-1/2} \sim c^{-1/2} N_j(p_T)^x$	$z_j \sim c p_{Tj}^{-2x}$ therefore $z_j^{-1/2} \sim c^{-1/2} p_{Tj}^x$

The first sum in equation 20 is $\sum_{j>q} a_{i,j} z_j^{-1/2}$, where i is in the first q nodes and j is always outside the first q nodes. Let $w_i = \sum_{j>q} a_{i,j} z_j^{-1/2}$. If connections were randomly allocated after clustering then the probability of a connection crossing between clusters would be $q(n-q)/n(n-q)$. However as clustering is supposed to avoid breaking connections the probability number of crossing connections will be lower than this. Call this χ, so that the sum $\sum_{j>q} a_{i,j}$ has on average χc non zero terms. A similar table is constructed for w_i;

	For most j; $p_{Tj} < p_{Ti}$	For most j; $p_{Tj} > p_{Ti}$
$x > 0$	$w_i \sim \chi c^{1/2} p_{Tj}^{-x}$	$w_i \sim \chi c^{1/2} N_j(p_T)^x p_{Ti}^{-2x}$
$x < 0$	$w_i \sim \chi c^{1/2} N_j(p_T)^x p_{Ti}^{-2x}$	$w_i \sim \chi c^{1/2} p_{Tj}^{-x}$

Note that from the perspective of node i connected to node j $N_j(p_T)$ relates to the p_T two steps away in the direction of j.

The other sums in equation 20 have similar form, the distinction being is indices i and j belong to the same or opposing clusters. This changes the prefactor, χ, as it is no longer a sum over crossing vertices, but otherwise the algebra will be the same as there are no assumptions about group membership in it.

The table is recast for those two cases, and the entries are described in terms of the influence of p_T values on the magnitude of the sum. Firstly, when both indices belong to the same cluster the sum is subtracted from the value of the eigenvalue. As such larger sums are favoured.

	Node i is high p_T so mostly $p_{Tj} < p_{Ti}$	Node i is low p_T so mostly $p_{Tj} > p_{Ti}$
$x > 0$	The smaller the p_T of a node connected to i the greater its contribution to the sum. The clustering is favoured when i connects to many soft nodes. The harder p_{Ti} is not used directly.	The larger the neighbourhood p_T for a node connected to i in comparison to soft p_{Ti} the greater its contribution to the sum. The clustering is favoured when nodes connecting to i sit in a higher p_T neighbourhood. The harder p_{Tj} is not used directly.
$x < 0$	The larger the neighbourhood p_T for a node connected to i in comparison to hard p_{Ti} the greater its contribution to the sum. The clustering is favoured when nodes connecting to i sit in a higher p_T neighbourhood. The softer p_{Tj} is not used directly.	The smaller the p_T of a node connected to i the greater its contribution to the sum. The clustering is favoured when i connects to many soft nodes. The softer p_{Ti} is not used directly.

The difference between $x > 0$ and $x < 0$ can be summarised as $x > 0$ indicates that things of differing scales should be connected and the highest p_Ts should not reduce any connectedness. On the other hand $x < 0$ indicates that things of the same scale should be clustered together and the lowest p_Ts should not reduce connectedness.

The sums that cross between groups are added to the value of the eigenvalue, so increasing the value of the sum disfavours the clustering. This symmetry means that the implications for $x > 0$ or $x < 0$ will be the same as summarised above.

5.2 Method

5.2.1 Particle data

The dataset used here is the 40GeV simulated Higgs cascade decay used in section 4.1. One Standard Model (SM) Higgs at 125GeV decays to two light Higgs at 40GeV, which in turn decay to b-$\bar{b}$ quark pairs. That is $H_{125\text{GeV}} \to h_{40\text{GeV}} h_{40\text{GeV}} \to b\bar{b}b\bar{b}$.

This dataset has the desirable property of creating b-jets with a range of geometries owing to the boost provided by large mass of the SM Higgs and the high chance of overlap with 4 b-quarks in the event. Other radiation from the protons is also included.

Before any evaluation can be performed on the final state of the simulation particles that ended outside the range of the silicon tracker ($|\eta| > 2.5$) or particles with low transverse momentum ($p_T < 0.5$ GeV) are cut. This is to mimic restrictions from reconstruction accuracy.

A particle is considered a b-descendant if is found in the chain of decays from a b-quark. After cuts, 72% of events have at least 5 b-descendants and 5 non b-descendants available. Exactly how the percentage of b-descendants changes with event size is visualised in figure 10.

5.2.2 Clustering algorithm

Now the implementation of the theory described in section 5.1 will be specified. For every simulated event this process if used to select a clustering.
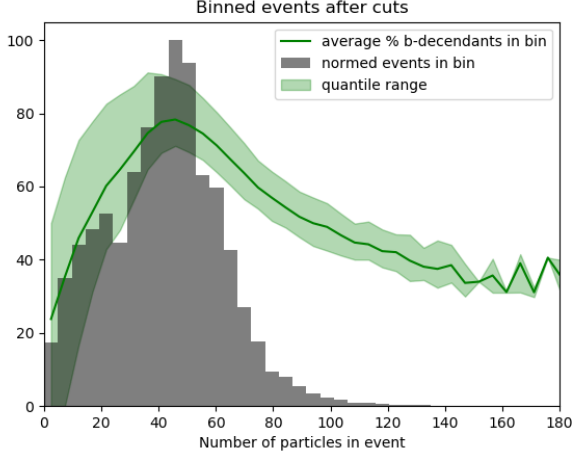
Figure 10: The end state particle in the simulated events are filtered with the standard cuts, $p_T > 0.5$, $|\eta| < 2.5$. The events are binned according to how many particles remain after the cuts. The percentage of b-descendants in the remaining event after the cuts is averaged for each bin and plotted on the same axis. After the cuts have been applied most events are left with around 50 particles. The percentage of particles that are descendant from a b-quark varies, it is at it's highest in events with 50 particles, and most variable in events will small multiplicity.

1. The particles are to be used to form the nodes of a graph, the edges of which will be weighted by some measure of proximity between the particles known as affinity. To obtain an affinity, first a distance is obtained; $d_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ where y_i is the rapidity of particle i and ϕ_i is the barrel angle is particle i.

2. The affinity must increase as particles become more similar, whereas the distance will shrink. This is done by taking an exponent so that $a_{i,j} = \exp(2d_{i,j})$.

3. These affinities allow the construction of a Laplacian. The Laplacian used is the unnormalised Laplacien, which has $\sum_j a_{i,j}$ in the ith diagonal entry (also known as the degree of node i) and $-a_{i,j}$ of the diagonal in column i row j.

4. From the Laplacian a predetermined number of eigenvectors are calculated to create the embedding space. The eigenvectors have as many elements as there are particles, and the coordinates of the ith particle in the embedding space is the ith element of each eigenvector.

5. The first clustering can be done based on a measure of distance derived from the particles p_T and its position in the embedding space. The p_T is used as in the Luclus [41], with a variable exponent, q;

$$p_T \text{ factor} = \left(\frac{p_{T\,i} p_{T\,j}}{s(p_{T\,i} + p_{T\,j})} \right)^q.$$

Where s is the invariant mass of all observed particles in the event, it is used to make the p_T factor unitless. The embedded distance is euclidean distance in the embedding space multiplied by this factor.

6. The two objects that have the smallest embedding distance are combined. In physical space the combined object is created by summing the respective four momenta, in the embedding space two methods for locating the combined object are tried.

    (a) In a `SpectralMeanJet` clustering the location of the combined object is the geometric mean of the inputs. The clustering then continues to combine things in this manner.

    (b) In a `SpectralFullJet` once two object have been combined in physical space the embedding space is recalculated from step 1.

7. When the closest object to a particle in the embedding space is further away than R then the combined object is considered to be a complete jet and removed from future clustering.

To provide a basis for comparison the results of clustering with an anti-kt algorithm (as in [14]) is also shown.

5.3   Results

This section is still subject to change as findings reveal advantages and drawbacks of parameter choices. Three choices of clustering algorithm are considered here;

- A standard anti-KT algorithm with $R = 1.36$

- A `SpectralFullJet` algorithm with 2 eigenvectors, $R = 0.63$, and $q = -0.035$

- A `SpectralMeanJet` algorithm with 4 eigenvectors, $R = 0.70$, and $q = 0$

As the data is simulated it is possible to compare the performance of clustering algorithms to Monte Carlo truth. Each event contains 4 b-quarks and for each of them it is possible to identify the particle into which they decayed, as a subset of the particles in the final state. Henceforth the detectable decay products of the b-quark will be called the descendants of the b-quark. For two reasons it is not possible for this clustering algorithm to gather all the descendants of each b-quark into one jet: firstly not all the descendants make the p_T and η cuts, so some are discarded before clustering; secondly the descendants of the b-quarks in an event are not mutually exclusive, due to interactions during hadronisation the quarks share descendants, and our clustering algorithm does produce exclusive clusters.

Knowing the parts of the final state that are descended from each b-quark creates a clear allocation of jets to quarks. For each quark, the jet that contains the greatest mass in descendent particles is tagged to represent that quark.

Mass peaks can the be constructed from the tagged jets, using all the particles in the jet, both descendant of the quarks and background. In figure 11 the masses of all jets that have been b-tagged are plotted. In figure 12 three selections are plotted; firstly only events where some trace of all 4 b-quarks is found are plotted with the mass of the descendants of the heavy Higgs in the background. Then the two light Higgs in each event are sorted by the mass of their descendants, in effect they are ranked by how well they were picked up by the detector. The mass of the light descendants is plotting in the background and over that the mass of the associated jets in each event is shown.
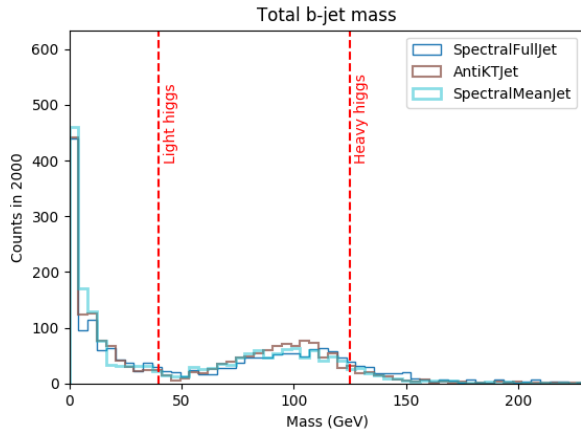
Figure 11: All tagged jets in each event are plotted. A peak just shy of 125GeV is desired as this is the mass of the decaying heavy Higgs. Another peak at 40GeV could be observed is only one light Higgs was found in the detector volume. `SpectralFullJet` gets closest, peaking close to 115 GeV and showing some contamination from background with the bins > 125 GeV. Anti-KT and `SpectralMeanJet` both give similar behaviour, falling somewhat short. The Anti-kt jet used a R of 0.63.

Figure 12: Starting with the plot on the far left, the jet mass of events where some descendants from all 4 b-quarks were found is plotted over the total mass of all descendants. The plot in the centre takes the light Higgs whose descendants have the greatest mass in each event. The mass of these descendants is plotted in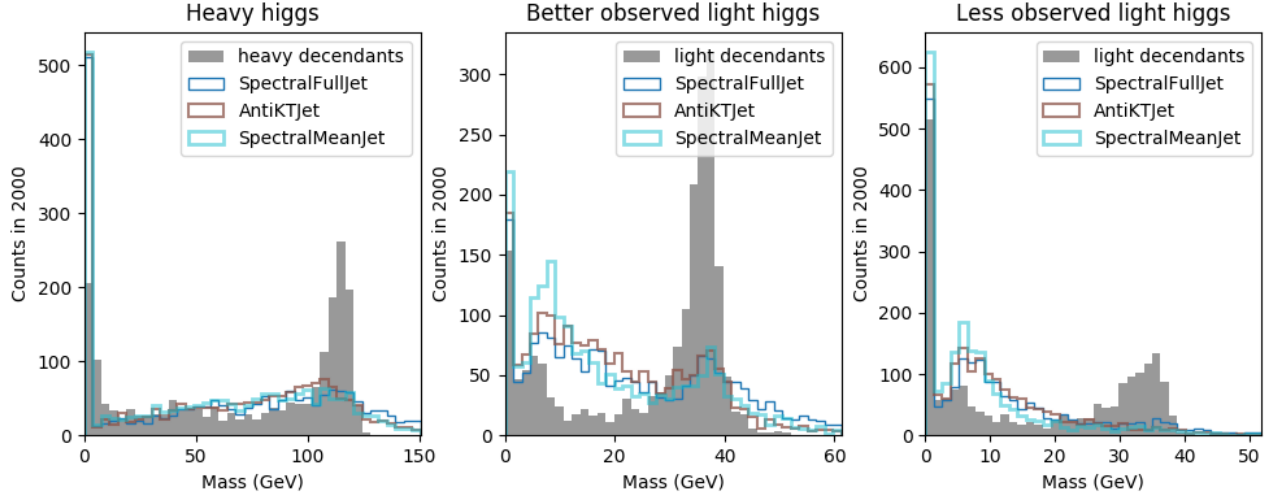 grey, and over this the masses of the jets in each event that correspond to this better observed Higgs are shown. That is, the tags in each event that have been produced by the better observed light Higgs are allocated to jets, and only the mass of these jets is tabulated - thus a good cluster will have obtained the mass of the light Higgs descendants. Finally on the left the light Higgs who's descendants have less mass is shown along with the corresponding jets. These three plots make it clear that only `SpectralFullJet` is obtaining the majority of the descendants, but it is combining them with some background as the tail of the `SpectralFullJet` distribution continues past the descendants distribution out of the frame. The behaviour of `SpectralMeanJet` does not significantly differ from Anti-KT

5.4 Conclusions

Spectral clustering has the potential to perform at least as well as traditional clustering methods such as anti-kt. Fully re-clustering at each step, as in `SpectralFullJet` is important for good performance, and thus spectral clustering may prove expensive.

Spectral clustering is certainly capable of producing sharp mass peaks, though it tends to include more background than is ideal.

5.4.1 Going forward

From here, there are three further questions that seem pressing;

1. Is spectral clustering IRC safe? Ideally this would be tested by generating samples with soft radiation and collinear splitting and comparing the clusters formed with and without these changes. Shape variables might also be of assistance here, as they have infra-red and collinear limits.

2. Does spectral clustering perform well on other data samples? Comparing the performance of methods with the same hyperparameters on different data samples would explore how well the clustering generalises. It is important to determine if the method can reliably gather any signal, or if it has been tuned to gather the signal we expect to find.

3. Does the additional information produced by spectral clustering provide any guidance in jet classification? Jet classification will be discussed more in section 6, however, it is worth noting at this point that spectral clustering effectively generates additional values associated with each jet, and perhaps these values would be of use in classifying the jet.

6 Jet classification

Jet classification attempts to identify the signal jets, commonly those containing quark descendants, and further identify the flavour of the quark. Typically this begins with a process known as expert feature selection [46, 45].

Expert feature selection, also sometimes called handcrafted features, is the process of defining variables on jets of tracks inspired by physics knowledge. These features hope to provide good discriminating power between different types of jet. The most successful example of this is perhaps the n-subjettiness, others include jet p_T and jet area.

Expert features are often criticised for requiring redesigning for each problem, and for potential information loss [16, 43]. Others have come out in favour of expert feature selection, however. In [37] expert feature selection is compared to a neural network approach that retains all information and the author reports that "the most power is gained from a Neural Network, although we have shown that a large fraction of this information can be obtained through simple observables.".

There are a number of systematic ways for generating features akin to expert features for jets; The Lund Plane [19] generates new features in a way that is easy to visualise and can create a linear sequence for a jet. This involves plotting tracks that split from the central track on the relative p_T - relative angle plane to generate density maps. Cuts can be used to impose IR and collinear safety. This feature should lend itself to visualisation of trends. Another such system has been dubbed energy flow polynomials [34]. Energy flow polynomials (EFP) constitute a complete basis for all IR collinear safe features for a jet. The author of [34] emphasises that commonly chosen expert features are found by combing only a few terms in the EFP.

Machine learning tools are capable of fitting highly non-linear signals, so it might be possible to use them to bypass expert feature selection all together.

6.1 Recursive neural tensor networks

There are many possible forms for a neural network that classifies jets;

- A feed forward Deep Neural Network (DNN). This is simple to train but requires a fixed length input. Obtaining a fixed length input is a challenge because the number of tracks in a jet is not fixed, expert features must be used, and often some form of interpolation or padding is required. This doesn't make a DNN a very natural fit for the problem.

- A Convolutional Neural Network (CNN). This can be run on an image generated by the calorimeter output. It has the advantage of interpretability because the filters it developed to scan the image can be analysed and some of it's process can be estimated. Removing symmetries from the images, however, is difficult to achieve without distorting their physical content. Rotations in particular are frequently done improperly, altering the jet mass. Further, the images are sparse, and a CNN focuses on local correlations so cannot perform optimally on spares images.

- A Recurrent Neural Network (RNN). This can read in a sequence, so if we could order the tracks it would be able to used all there data. Finding a natural ordering for the tracks is difficult, however. Previously p_T has been used, the Lund Plane is certainly a natural ordering, but it doesn't capture all the tracks.

- A Recurrent Neural Tensor Network (RTRN). RTRNs follow a tree shape, being originally developed for parsing syntax trees. This happens to coincide exactly with the shape of a jet clustering algorithm, so we could apply them to this.

To begin with the starting point considered was DeepCSV. It is a DNN, so it's foible is a fixed length input vector. As each input is one jet long, two types of variable are considered, variables that describe an aspect of the jet as a whole (p_T of the jet for instance) or variables that describe an aspect of a single track (such as p_T of that track). The first kind have a fixed length and provided all values are successfully reconstructed (which is not always true) will require no 'padding'. Padding is the manner in which missing values are filled. The second kind, variables per track, will not have a fixed length even if the reconstruction is perfect because the number of tracks in a jet varies. Unless the plan is to train a separate DNN for each size of jet (which would be statistically wasteful and computationally expensive) some form of padding will have to be chosen.

The amount of padding required for a restricted set of tracks in my dataset is shown in figure 13.

The padding decision taken in [46] was called zero padding, but as the distributions are all centred on zero before padding this is equivalent to padding with the average value of each distribution. Thus we can imagine that we are inserting "average" tracks into each event until there are sufficient tracks to fill the fixed length input. This make very little physical sense; the momentum of the additional tracks will contradict the momentum of the jet, and the angle of the tracks will have no relation to the angle of the jet at all.

This was deemed to be an undesirable approach, so new methods that didn't require fixed length input were considered.

Figure 13: Both DeepCSV and CSVv2 used zero padding to fill in values that were not available from the data. In my data sample the percentage of each variable that must be zero padded for CSVv2 is shown here.

6.2 Architectural direction

The network used by [16] was originally developed for parsing syntax trees. It the case of jet clustering it generates a state vector ($\mathbf{h}_k$) for every pseudojet. That state vector encodes the useful information about the pseudojet structure and kinematics.

The equations involved can be seen in figure 14. Qualitatively the first pseudojets being the tracks or towers must make a state vector out of the observables alone. These are known as leaf nodes in the net, and they learn a process for encoding observables into a new hidden state. When two pseudojets merge into a new pseudojet there are three kinds of input to the state vector; the left and right state vectors of the child nodes and the combined kinematics of the pseudojet. A separate procedure is learnt to combine these three things. These two processes combine all the tracks into one root state vector.

The root state vector, or final state vector, will be passed on to a DNN, along with global event information which tags the jet.

$$\mathbf{o}_k = \begin{cases} \mathbf{v}_k & \text{if } k \text{ is leaf} \\ \mathbf{o}_{k\text{left}} \oplus \mathbf{o}_{k\text{right}} & \text{otherwise} \end{cases}$$

$$\mathbf{u}_k = \sigma(W_u g(\mathbf{o}_k) + \mathbf{b}_u)$$

$$\mathbf{h}_k = \begin{cases} u_k & \text{if } k \text{ is leaf} \\ \sigma\left(W_h \begin{bmatrix} \mathbf{h}_{k\text{left}} \\ \mathbf{h}_{k\text{right}} \\ \mathbf{u}_k \end{bmatrix} + \mathbf{b}_h\right) & \text{otherwise} \end{cases}$$

Figure 14: A single node of an RNTN. The vector $\mathbf{o}_k$ is the vector of observable properties of the kth node. If the kth node is a leaf then these would be the track properties, otherwise they are a combination of the properties of the tracks below k. The vector $\mathbf{u}_k$ is the embedding of $\mathbf{o}_k$ into the latent space of the net. The vector $\mathbf{k}_k$ is the 'state' of the kth node, structural information about the tree propagates in the state of the nodes.

6.2.1 Subsequent steps

Once this net has been replicated and seen to perform acceptably there are a number of forward direction open;

1. Most RNNs and RNTNs have a target available at every node. This greatly improves the stability of the training process. A physically inspired intermediate target might be identified of increase the stability of the net as it trains.

2. RNTNs have been implemented with Long Short Term Memory (LSTM) and jet taggers with LSTM have been implemented [20]. To my knowledge, however, they have yet to be combined. This would be a very natural extension.

3. Such a net will be sensitive to the subtleties of hadronisation, there would be a number of possibility's for testing their robsutness to Monte Calro errors. Unsupervised training on known signal/background regions in data could be compared to equivalent MC trained nets. Different MC generators could be compared.

4. It would be interesting to find an interpretation of the net's latent space. Perhaps t-SNE would be suitable for this.

Points 1 and 2 should certainly be undertaken, if the structure show promise then points 3 and 4 would be worth attempting.

7 Presentations

7.1 18/02/2018, GRADnet winter school

Workshop based school on leadership skills and team structuring. Had a panel of experts from tech fields. Covered both theory and practice of good leadership.

7.2 14/06/2018 56th EMISSP (ISSP EMFCSC)

This was a summer school hosted by the EMFCSC (http://www.emfcsc.infn.it/issp) on subnuclear physics. It covered both theoretical and experimental developments.

In particular there was experimental discussion of IceCube, LIGO, ALICE and LHCb among others.

7.3 17/07/2018 STFC's Summer School in Artificial Intelligence and Machine Learning

Summer school hosted by UCL on AI and ML with a focus on physics phenomenology. A number of industry talks also, the conference being financed by industry.

Poster session was particularly useful, many posters on similar projects and discussion of challenges and drawbacks of ML. I presented a poster on my attempts to train CSVv2 and DeepCSV replicas. The questions and advice from other attendees offered valuable new perspectives.

7.4 11/07/2018 CMS UK meeting

Meeting of the UK branch of the CMS collaboration, hosted by the university of Bristol. Recent developments and analysis are discussed. Many talks on current ML techniques used at CMS.

I gave a talk on my attempt to replicate CSVv2/DeepCSV. Again the questions put by the audience offered valuable insight.

7.5 20/08/2018 48th British Universities Summer School in Theoretical Elementary Particle Physics

Summer school for particle physics hosted by the university of Oxford. The school presented basic material on many areas including QFT, lattice, strings and CFT. Some recent experimental work was also discussed.

I gave a short overview talk here about ML and neural networks.

7.6 27/10/2018 UK HEP Forum 2018: The Spice of flavour

Conference on flavour physics. Predominately experimental discussion with some theory.

7.7 19/12/2018, Annual theory meeting and young theorists meeting

This pair of back to back conferences explores recent theoretical developments in various fields and Offers young researches a chance to present their work. It was interesting to learn about similar projects going on across other areas of physics.

7.8 6/01/2019, YETI

Promotes interaction between theorists and experimentalists at the early career stage and to encourage interest in phenomenology. This year the topic was proposed collider plans. Among the workshops offered there was some practical experience with analysis tools on root files.

7.9 16/3/2019, SOTSEF outreach event

In collaboration with other students presented an engagement show at the Southampton Science and Engineering Festival (SOTSEF). The show was an accessible explanation of the experimental components of a particle collider such as the LHC. It described the major components and used bench top interactive experiments to demonstrate the principles behind them.

7.10 4/4/2019, Future of particle physics in a post Higgs landscape

Student lead discussion of the field of particle physics in the absence of major discoveries since the Higgs Boson in 2012. Lots of work on the extended Higgs sector is presented here. I gave a talk on the use of more complicated neural network architectures to classify particle jets.

7.11 1/7/2019, GRADnet Summer School

Summer school on opportunities for physics PhD student in industry. It was a useful thing to attend, although most of the options seem rather less exciting than work in academia.

7.12 8/7/2019, Decoding new physics from data: connecting theory and signatures

Pedagogical workshop on phenomenology and the impact of new observations on current theories. A very useful overview of many upcoming directions. Also a clear an accessible dissection of key statistical concepts and common foibles was presented by Prof. Glen Cowan.

I presented a poster on potential uses of RNTNs to classify the flavour of jets.

7.13 23/9/2019, 1st Mediterranean Conference in Higgs Physics

Conference on experimental and theoretical developments in Higgs Physics. I gave a talk on tools for determining parameter spaces.

7.14 29/4/2019, NEXT spring meeting

Theoretical and experimental physics talks from members of the NEXT network. I gave a talk on Spectral clustering.

8 Acknowledgements

References

[1] M. Aaboud et al. "Search for a heavy Higgs boson decaying into a Z boson and another heavy Higgs boson in the $\ell\ell bb$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett.* B783 (2018), pp. 392–414. DOI: 10.1016/j.physletb.2018.07.006. arXiv: 1804.01126 [hep-ex].

[2] J. Alwall et al. "MadGraph 5: going beyond". In: *Journal of High Energy Physics* 2011.6 (June 2011). ISSN: 1029-8479. DOI: 10.1007/jhep06(2011)128. URL: http://dx.doi.org/10.1007/JHEP06(2011)128.

[3] M. S. and. "Deep learning in jet reconstruction at CMS". In: *Journal of Physics: Conference Series* 1085 (Sept. 2018), p. 042029. DOI: 10.1088/1742-6596/1085/4/042029. URL: https://doi.org/10.1088%2F1742-6596%2F1085%2F4%2F042029.

[4] A. Arhrib. *Unitarity constraints on scalar parameters of the Standard and Two Higgs Doublets Model*. 2000. arXiv: hep-ph/0012353 [hep-ph].

[5] F. Beaudette. "The CMS Particle Flow Algorithm". In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, pp. 295–304. arXiv: 1401.8155 [hep-ex].

[6] P. Bechtle et al. "HiggsBounds 2.0.0: Confronting neutral and charged Higgs sector predictions with exclusion bounds from LEP and the Tevatron". In: *Computer Physics Communications* 182.12 (Dec. 2011), pp. 2605–2631. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2011.07.015. URL: http://dx.doi.org/10.1016/j.cpc.2011.07.015.

[7] P. Bechtle et al. *HiggsBounds: Confronting arbitrary Higgs sectors with exclusion bounds from LEP and Tevatron*. 2009. arXiv: 0905.2190 [hep-ph].

[8] P. Bechtle et al. "HiggsBounds-4: improved tests of extended Higgs sectors against exclusion bounds from LEP, the Tevatron and the LHC". In: *The European Physical Journal C* 74.3 (Mar. 2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2693-2. URL: http://dx.doi.org/10.1140/epjc/s10052-013-2693-2.

[9] P. Bechtle et al. "HiggsSignals: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC". In: *The European Physical Journal C* 74.2 (Feb. 2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2711-4. URL: http://dx.doi.org/10.1140/epjc/s10052-013-2711-4.

[10] R. Benbrik et al. *Mapping $pp \to A \to ZH \to l^+l^-b\bar{b}$ and $pp \to H \to ZA \to l^+l^-b\bar{b}$ Current and Future Searches onto 2HDM Parameter Spaces*. 2020. arXiv: 2006.05177 [hep-ph].

[11] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Physics Reports* 516.1-2 (July 2012), pp. 1–102. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.02.002. URL: http://dx.doi.org/10.1016/j.physrep.2012.02.002.

[12] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].

[13] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet User Manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[14] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[15] A. Chakraborty et al. *Revisiting Jet Clustering Algorithms for New Higgs Boson Searches in Hadronic Final States*. 2020. arXiv: 2008.02499 [hep-ph].

[16] T. Cheng. "Recursive Neural Networks in Quark/Gluon Tagging". In: *Computing and Software for Big Science* 2.1 (Nov. 2018). INSPEC:18015611, 3 (13 pp.) ISSN: 2510-2036. DOI: 10.1007/s41781-018-0007-y.

[17] A. Djouadi, J. Kalinowski, and M. Spira. "HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension". In: *Computer Physics Communications* 108.1 (Jan. 1998), pp. 56–74. ISSN: 0010-4655. DOI: 10.1016/s0010-4655(97)00123-9. URL: http://dx.doi.org/10.1016/S0010-4655(97)00123-9.

[18] A. Djouadi, J. Kalinowski, and P. Zerwas. "Two and three-body decay modes of SUSY Higgs particles". In: *Z. Phys. C* 70 (1996), pp. 435–448. DOI: 10.1007/s002880050121. arXiv: hep-ph/9511342.

[19] F. A. Dreyer, G. P. Salam, and G. Soyez. "The Lund jet plane". In: *Journal of High Energy Physics* 2018.12 (Dec. 2018), p. 64. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2018)064. URL: https://doi.org/10.1007/JHEP12(2018)064.

[20] S. Egan et al. "Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC". In: (2017). arXiv: 1711.09059 [hep-ex].

[21] S. D. Ellis and D. E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[22] D. Eriksson, J. Rathsman, and O. Stål. "2HDMC – two-Higgs-doublet model calculator". In: *Computer Physics Communications* 181.1 (Jan. 2010), pp. 189–205. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2009.09.011. URL: http://dx.doi.org/10.1016/j.cpc.2009.09.011.

[23] A. G Akeroyd, A. Arhrib, and E. Naimi. "Note on tree-levelTreeLevel unitarity in the general two Higgs doublet model". In: *Physics Letters B* 490.1-2 (Sept. 2000), pp. 119–124. ISSN: 0370-2693. DOI: 10.1016/s0370-2693(00)00962-x. URL: http://dx.doi.org/10.1016/S0370-2693(00)00962-X.

[24] W. Grimus et al. "The oblique parameters in multi-Higgs-doublet models". In: *Nuclear Physics B* 801.1-2 (Sept. 2008), pp. 81–96. ISSN: 0550-3213. DOI: `10.1016/j.nuclphysb.2008.04.019`. URL: `http://dx.doi.org/10.1016/j.nuclphysb.2008.04.019`.

[25] J. F. Gunion and H. E. Haber. "CP-conserving two-Higgs-doublet model: The approach to the decoupling limit". In: *Physical Review D* 67.7 (Apr. 2003). ISSN: 1089-4918. DOI: `10.1103/physrevd.67.075019`. URL: `http://dx.doi.org/10.1103/PhysRevD.67.075019`.

[26] A. Hadjighasem et al. "Spectral-clustering approach to Lagrangian vortex detection". In: *Phys. Rev. E* 93 (6 June 2016), p. 063107. DOI: `10.1103/PhysRevE.93.063107`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.93.063107`.

[27] J. Haller et al. "Update of the global electroweak fit and constraints on two-Higgs-doublet models". In: *The European Physical Journal C* 78.8 (Aug. 2018). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-018-6131-3`. URL: `http://dx.doi.org/10.1140/epjc/s10052-018-6131-3`.

[28] R. V. Harlander and W. B. Kilgore. "Higgs boson production in bottom quark fusion at next-to-next-to leading order". In: *Phys. Rev. D* 68 (2003), p. 013001. DOI: `10.1103/PhysRevD.68.013001`. arXiv: `hep-ph/0304035`.

[29] R. V. Harlander and W. B. Kilgore. "Next-to-next-to-leading order Higgs production at hadron colliders". In: *Phys. Rev. Lett.* 88 (2002), p. 201801. DOI: `10.1103/PhysRevLett.88.201801`. arXiv: `hep-ph/0201206`.

[30] R. V. Harlander, S. Liebler, and H. Mantler. "SusHi Bento: Beyond NNLO and the heavy- top limit". In: *Computer Physics Communications* 212 (Mar. 2017), pp. 239–257. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2016.10.015`. URL: `http://dx.doi.org/10.1016/j.cpc.2016.10.015`.

[31] R. V. Harlander, S. Liebler, and H. Mantler. "SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM". In: *Computer Physics Communications* 184.6 (June 2013), pp. 1605–1617. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2013.02.006`. URL: `http://dx.doi.org/10.1016/j.cpc.2013.02.006`.

[32] S. Kanemura, T. Kubota, and E. Takasugi. "Lee-Quigg-Thacker bounds for Higgs boson masses in a two-doublet model". In: *Physics Letters B* 313.1-2 (Aug. 1993), pp. 155–160. ISSN: 0370-2693. DOI: `10.1016/0370-2693(93)91205-2`. URL: `http://dx.doi.org/10.1016/0370-2693(93)91205-2`.

[33] V. Khachatryan et al. "Search for neutral resonances decaying into a Z boson and a pair of b jets or *tau* leptons". In: *Physics Letters B* 759 (Aug. 2016), pp. 369–394. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2016.05.087`. URL: `http://dx.doi.org/10.1016/j.physletb.2016.05.087`.

[34] P. T. Komiske, E. M. Metodiev, and J. Thaler. "Energy flow polynomials: a complete linear basis for jet substructure". In: *Journal of High Energy Physics* 2018.4 (Apr. 2018), p. 13. ISSN: 1029-8479. DOI: `10.1007/JHEP04(2018)013`. URL: `https://doi.org/10.1007/JHEP04(2018)013`.

[35] M. Krause, M. Mühlleitner, and M. Spira. "2HDECAY-A program for the calculation of electroweak one-loop corrections to Higgs decays in the Two-Higgs-Doublet Model including state-of-the-art QCD corrections". In: *Computer Physics Communications* 246 (Jan. 2020), p. 106852. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2019.08.003`. URL: `http://dx.doi.org/10.1016/j.cpc.2019.08.003`.

[36] D. Krohn, J. Thaler, and L.-T. Wang. "Jets with variable R". In: *Journal of High Energy Physics* 2009.06 (June 2009), pp. 059–059. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2009/06/059`. URL: `http://dx.doi.org/10.1088/1126-6708/2009/06/059`.

[37] J. Lin et al. "Boosting $H \rightarrow b\bar{b}$ with machine learning". In: *Journal of High Energy Physics* 2018.10 (Oct. 2018), p. 101. ISSN: 1029-8479. DOI: `10.1007/JHEP10(2018)101`. URL: `https://doi.org/10.1007/JHEP10(2018)101`.

[38] U. von Luxburg. *A Tutorial on Spectral Clustering*. 2007. arXiv: `0711.0189 [cs.DS]`.

[39] F. Mahmoudi. "SuperIso: A program for calculating the isospin asymmetry of in the MSSM". In: *Computer Physics Communications* 178.10 (May 2008), pp. 745–754. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2007.12.006`. URL: `http://dx.doi.org/10.1016/j.cpc.2007.12.006`.

[40] S. Moretti and W. Stirling. "Contributions of below threshold decays to MSSM Higgs branching ratios". In: *Phys. Lett. B* 347 (1995). [Erratum: Phys.Lett.B 366, 451 (1996)], pp. 291–299. DOI: `10.1016/0370-2693(95)00088-3`. arXiv: `hep-ph/9412209`.

[41] S. Moretti, L. Lönnblad, and T. Sjöstrand. "New and old jet clustering algorithms for electron-positron events". In: *Journal of High Energy Physics* 1998.08 (1998), p. 001.

[42] M. E. Peskin and T. Takeuchi. "Estimation of oblique electroweak corrections". In: *Phys. Rev. D* 46 (1992), pp. 381–409. DOI: 10.1103/PhysRevD.46.381.

[43] A. Radovic et al. "Machine learning at the energy and intensity frontiers of particle physics". In: *Nature* 560.7716 (Aug. 2, 2018). INSPEC:17999331, pp. 41–8. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0361-2.

[44] R. J. Sánchez-García et al. "Hierarchical Spectral Clustering of Power Grids". In: *IEEE Transactions on Power Systems* 29.5 (2014), pp. 2229–2237.

[45] S. Schramm. *ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects*. Tech. rep. ATL-PHYS-PROC-2017-236. Geneva: CERN, Nov. 2017. URL: https://cds.cern.ch/record/2291608.

[46] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[47] A. M. Sirunyan et al. "Search for new neutral Higgs bosons through the $H \to ZA \to l + l - b\bar{b}\mathrm{H} \to \mathrm{ZA} \to \ell^+\ell^-\mathrm{b\bar{b}}$ process in pp collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)055. URL: http://dx.doi.org/10.1007/JHEP03(2020)055.

[48] A. Sirunyan et al. "Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two muons and two b quarks in pp collisions at 13 TeV". In: *Physics Letters B* 795 (Aug. 2019), pp. 398–423. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2019.06.021. URL: http://dx.doi.org/10.1016/j.physletb.2019.06.021.

[49] A. M. Sirunyan et al. "Search for $t\bar{t}$H production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 06 (2018), p. 101. DOI: 10.1007/JHEP06(2018)101. arXiv: 1803.06986 [hep-ex].

[50] T. Sjostrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: http://www.sciencedirect.com/science/article/pii/S0010465515000442.

[51] M. Wobisch and T. Wengler. "Hadronization corrections to jet cross-sections in deep inelastic scattering". In: *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*. 1998, pp. 270–279. arXiv: hep-ph/9907280 [hep-ph].