University of Southampton

Faculty of Enginering and Physical Sciences
School of Physics and Astronomy

In Search Of A Detectable Two-Higgs Doublet Model At The LHC

by

Ciara Jane Byers
MPhys
ORCiD: 0009-0001-5676-0004

A thesis for the degree of
Doctor of Philosophy

August 21, 2023

University of Southampton

Abstract

Faculty of Enginering and Physical Sciences
School of Physics and Astronomy

Doctor of Philosophy

In Search Of A Detectable Two-Higgs Doublet Model At The LHC

by Ciara Jane Byers

This work presents investigations into the potential of Two-Higgs doublet models (2HDMs) and the tool that has been developed in service to these enquiries. 2HDMs propose a possible route to find the solutions needed to remedy the problems within the Standard Model, and additionally are some of the simplest possible extensions to the Higgs sector; ensuring that we have fully explored their potential seems only logical. Their simplicity has led to them being the focus of a great deal of research. The tool, `Magellan`, was developed in part to fulfil the requirements for DISCnet funding of the candidate. This tool aims to improve upon pre-existing methods of investigating 2HDMs, and possibly, in the future, further models. Research involving the use of so-called "data-intensive science" is a key part of this funding and the project was chosen with this in mind.

Contents

List of Figures

List of Tables

C. Byers E-mail: cb27g11@soton.ac.uk School of Physics & Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

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:

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

Acknowledgements

I would like to thank my supervisors, Profs. Stefano Moretti and Jonathan Flynn for all of their help and support through my PhD. In particular I would like to thank Prof. Moretti for going above and beyond to help me through what was a somewhat tumultous period personally as well as volunteering to take over as my main supervisor half-way through my candidature. Without his support I would not have gotten this far.

I would also like to thank my parents for always encouraging and feeding my interest in science. To both them, my sister and partner I will be forever grateful for their belief in me regardless of whether I had any myself at times, and for supporting me even on my most difficult days. I also thank my partner, Jordan, especially for all his patience with having to be in a three-way relationship with myself and my PhD.

I am very appreciative to all of my friends too, who have championed me all the way.

0.1 Academic Acknowledgements

I want to extend a huge thank-you to Dr David Englert, who continued to consult on `Magellan` despite finishing his own PhD several years ago and leaving academia, he was *essential* to its continued development.

I also want to thank my collaborators who worked with me throughout my time as a PhD student, in particular Dr Emmanuel Olaiya who always managed to find time to help, despite his hectic schedule.

I am grateful to Dr Jon Cockayne of the University of Southampton's Statistics group, who provided helpful discussion on MCMCs and their optimisation.

I also thank Dr Souad Semlali for providing a suitable UFO model to use in our research and for helpful discussion on 2HDMs.

I would like to thank STFC and DISCnet for funding my research, and both DISCnet and SEPnet for all the teaching provided.

I would like to acknowledge the use of the IRIDIS High Performance Computing Facility and associated support services at the University of Southampton for their regular assistance throughout my research.

Chapter 1

Introduction

The year 2012 marked the discovery of a Standard Model (SM)-like Higgs boson at the Large Hadron Collider (LHC) in [13][53]. This discovery indicated that the Higgs mechanism, a theory of mass generation for Nature's fundamental particles via spontaneous Electroweak Symmetry Breaking (EWSB), could be considered confirmed. Further, it fell in line with expectations of the SM; the mass was within the predicted range; many of the expected decays were observed following the initial discovery, with quantum numbers measured as predicted, even parity, 0 spin, 0 electric charge and colour charge. While this was unquestionably a great accomplishment and a source of great excitement at the time, a decade later and we are still left with unsolved problems in High Energy Physics (HEP). Nothing has been discovered that can clearly point physicists to a successor for the SM, to a solution that may begin to provide answers for the as yet unanswered questions in HEP. Questions such as; "What is the mechanism behind neutrino mass generation?", "What constitutes dark matter?", and, "Why does the universe have a disparate amount of matter and anti-matter?", for example.

There is the problem of energy scale to grapple with too. The Higgs boson discovered in 2012 is very light (125GeV) when compared with the Planck scale, which is of order 10^{19} GeV. Overcoming the hierarchy problem if this Higgs boson is fundamental in Nature, i.e. not a composite state, is difficult without invoking physics Beyond the SM (BSM) where there is an enlarged Higgs sector.

This work presents investigations into the potential of Two-Higgs Doublet Model (2HDM) discovery at the LHC and the physics toolbox that has been developed in service to this. 2HDMs offer a possible route to finding the solutions physicists need, and additionally, are some of the simplest possible extensions to the Higgs sector, ensuring that we have fully explored their potential seems only logical. Their simplicity has led to them being the focus of a great deal of research.

Chapter 2

Theoretical Background

2.1 The Standard Model

Our current best understanding of the forces of Nature (excluding gravity, which will not be covered in this thesis) comes in the form of the SM. The SM is a gauge theory composed of the Quantum Field Theories (QFTs) for each of the Strong, Weak and Electro-Magnetic (EM) forces[1] and of the matter fields they interact with. In a QFT concepts from quantum mechanics, classical mechanics and special relativity coalesce into a theory in which both the matter and forces are represented with quantum fields. Particles are excitations of these fields instead of being point like objects or 'billiard balls' as they are often considered in classical physics. The matter fields are fermionic, i.e. they have spin $\frac{1}{2}$ [2]; they can be decomposed into a pair of Weyl 2-spinors, one being left-handed and the other right-handed. In addition we have fields of spin 0 and spin 1, these are bosonic fields and are embedded in Lorentz scalars and vectors respectively.

The key to the effectiveness of the SM lies in the symmetries that are built in and naturally give rise to both the interactions between particles and to the conservation laws we see in Nature. From a mathematical point of view the SM is a $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$, non-Abelian gauge theory [91][124][150][114]; which means that the Lagrangian density is invariant under local transformations of the fields. Each of these forces has one or more associated mediator bosons, the amount of these is dependent on the group representation for that force.

The $SU(3)_c$ gauge group, where the subscript "c" indicates "Colour", is the gauge group responsible for the Strong force and interactions, the theory describing these and the associated fields is known as Quantum Chromodynamics (QCD). *Colour* is the name given to the charge this force couples to, there are 3 fundamental Colours that can be taken by the quarks, often referred to as 'red', 'green' and 'blue'; meanwhile the anti-quarks appropriately take on 'anti-Colours' corresponding to each of these. The

[1] Though we are able to combine the Weak and EM into one.

[2] Note that this is referring to *fundamental* matter, clearly composite particles such as mesons, baryons, etc. can still be bosonic.

bosons associated with $SU(3)_c$ are the *gluons*, which come in 8 Colour representations (as $SU(3)$ has 8 unique parameters since it must be both hermitian and traceless, these are equivalent to the Gell-Man matrices [81]), each of these are a mixture of a Colour-antiColour pairing. There is Colour confinement in QCD, meaning that only Colourless particles or composite particles may be observed, this is a result of the peculiar behaviour where the Strong force has asymptotic freedom, i.e. it becomes asymptotically weaker as the energy scale increases. The gluons only interact with themselves and the quarks as nothing else carries Colour charge, they are also massless like the photon.[3]

We have a single boson for $U(1)_Y$, 'B', and three for $SU(2)_L$, 'W_1, W_2, W_3'. These are not *physical* fields, Spontaneous Symmetry Breaking (SSB) has not yet occurred: the bosons are also massless as a result. These fields recombine after Symmetry Breaking (SB) via the Higgs mechanism to give the 3 massive Weak bosons, $W^{\pm}$, Z^0, and the massless photon, γ, that are actually observed in Nature. Additionally, the gauge group $SU(2)_L$ couples only to left-handed fermions, as indicated by the subscript "L". This is because in Nature, only left-handed fermions carry *Weak-Isospin*, right-handed ones do not carry it at all. It is not known why this should be the case, indeed this constitutes one of the unsolved mysteries left out of the SM.

In the case of the $U(1)_Y$ gauge field, the charge the field couples to is *Hypercharge* which relates the Weak-Isospin previously mentioned and the familiar EM charge as;

$$Y = 2(Q - T_3), \tag{2.1}$$

where[4] Y is the Weak Hypercharge, Q is the electric charge and T_3 is the third component of the isospin, known as the 'Weak-Isospin'. This relation is known as the Gell-Mann–Nishijima relation.

In the SM we have three 'generations' of matter particles, each containing an up-type quark, down-type quark, one electrically charged lepton and one neutrino corresponding to it. As well as this, each generation contains an antiparticle for each of those particles[5]. The quarks all carry Strong Colour charge, fractional electric and Weak charges[6]; as such they can interact via all the SM forces. The electrically charged leptons can naturally interact via the EM force, while the uncharged cannot. Leptons all carry Weak charge and can interact via the Weak force.

The fields describing the EM and Weak forces do not appear until after the Higgs mechanism spontaneously breaks the symmetry of the SM causing:

[3] This largely concludes the extent to which we will discuss QCD here as it is not the most relevant topic to this thesis.

[4] note that the coefficient, '2', on the RHS of the equation is arbitrary, its' purpose is simply to fix the normalisation of the $U(1)$ generator relative to the ones of $SU(2)$

[5] Though neutrinos are in fact their own antiparticles

[6] The antiquarks carry the same charges but with opposite sign.

$$SU(2)_L \otimes U(1)_Y \to U(1)_{EM}, \tag{2.2}$$

Left-handed fermions have, $T = \pm\frac{1}{2}$, while the right-handed have, $T = 0$, i.e. the left-handed are doublets under Weak-isospin and the right-handed are singlets. This means that the Weak force cannot transform right-handed particles as it can with the left-handed ones. Naturally we will be discussing the workings of the Higgs mechanism later in this chapter.

The Higgs mechanism is needed in the SM in order to introduce masses to the fermions. Fermions are described by the Dirac equation:

$$(i\hbar\gamma^\mu - mc)\psi(x) = 0 \tag{2.3}$$

$$(i\not{\partial} - mc)\psi(x) = 0 \tag{2.4}$$

where $\hbar$ is the Planck constant and γ^μ are the gamma matrices:

$$\gamma^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}; \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}; \gamma^2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}; \gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{2.5}$$

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^4 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{2.6}$$

which has a Clifford algebra of the anticommuntation relation:

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu} I_4 \tag{2.7}$$

and where ψ in 2.3 is the matter field given as a Dirac spinor:

$$\psi(x) = u(\bar{p})e^{-ip\cdot x} = \begin{bmatrix} \phi \\ \frac{\bar{\sigma}\cdot p}{E_{\bar{p}+m}}\phi \end{bmatrix} \tag{2.8}$$

where σ are the Pauli matrices and ϕ is an arbitrary two-spinor (specifically a vector of $\mathbb{C}^2$).

We can also represent this in the Weyl notation by changing basis s.t. :

$$\gamma^0 \rightarrow \begin{Bmatrix} 0 & I_2 \\ I_2 & 0 \end{Bmatrix} \quad ; \quad \gamma^5 \rightarrow \begin{Bmatrix} -I_2 & 0 \\ 0 & I_2 \end{Bmatrix} \tag{2.9}$$

which enables a straightforward form of the chiral projections of ψ:

$$\psi_L = \frac{1}{2}(1-\gamma^5) \quad ; \quad \psi_R = \frac{1}{2}(1+\gamma^5) \tag{2.10}$$

Dirac mass terms then appear in a Lagrangian as, $-m\bar{\psi}_L\psi_R$, but the Lagrangian **must** be covariant and we already know that the left and right handed parts are not treated identically by transformations under $SU(2)_L$. Clearly, if we have such a term initially in the Lagrangian we will lose our covariance, something else is needed. The Higgs mechanism causes *dynamic* mass generation, neatly side-stepping this issue by having the cross term appear after spontaneous symmetry breaking.

Physicists recognised relatively early in the 20th century that both the EM and Weak forces displayed similarities in their behaviour; their interactions were governed by vectorial Yukawa type couplings to spin-one fields. The mediators were expected to be massive due to the short range of the Weak force, therefore this needed to be worked into the theory. Initially attempts to produce massive gauge bosons required manual SB and a lack of renormalizability hindered progress [129; 86]. It was not until 't Hooft proved that spontaneously broken gauge symmetries are renormalizable even when they contain massive gauge bosons in 1971 [102] that they were able to move forward.

2.1.1 The Higgs Mechanism

Despite being referred to as the "Higgs" mechanism this was discovered independently and at roughly the same time by Brout and Englert [72], Guralnik, Hagen and Kibble [93] and of course Higgs himself [101].

The Higgs field is a $SU(2)_L$ complex, scalar doublet:

$$\phi = \begin{pmatrix} \phi^c \\ \phi^n \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.11}$$

where ϕ^c are the charged components and ϕ^n are the neutral ones. Our motivation for defining the charged and neutral parts will become clear as we progress. Here, ϕ_i with, $i = 1, 2, 3, 4$, are real scalar fields. The Lagrangian for the Higgs field is:

$$\mathcal{L} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) - \mathcal{L}_Y, \tag{2.12}$$

where $\mathcal{L}_Y$ is the Yukawa sector Lagrangian - the section interacting with the quarks. The first term in Equation 2.12 is describing the couplings of the scalar (Higgs) to the

gauge field and the last term $\mathcal{L}_Y$ is describing the interactions of the scalar and fermions. D_μ is the covariant derivative for, $SU(2)_L \otimes U(1)_Y$, as the Higgs does not interact via the Colour charge. This covariant derivative is defined as follows:

$$D_\mu = \partial_\mu - ig\frac{\sigma^a}{2}W^a{}_\mu - i\frac{g'}{2}YB_\mu, \tag{2.13}$$

here σ^a with, $\alpha = 1, 2, 3$, are the three Pauli matrices while g and g' are the couplings for $SU(2)_L$ and $U(1)_Y$ respectively. In the case of the Higgs the hypercharge, Y is $+\frac{1}{2}$ [7]. The $W^a{}_\mu$ and B_μ are the gauge fields for, $SU(2)_L \otimes U(1)_Y$.

Finally the Higgs potential, $V(\phi)$ is:

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \tag{2.14}$$

where λ and μ are constants with $\mu^2 < 0$ and $\lambda > 0$.

If we look to the global minimum through examination of this potential we can observe SSB in action:

$$\frac{\partial V}{\partial \phi_\alpha{}^*} = \phi_\alpha[-\mu^2 + \lambda(|\phi_\alpha|^2 + |\phi_\beta|^2)] = 0, \tag{2.15}$$

where, $\alpha = n, c$, and, $\beta \neq \alpha$. There are two solutions:

$$\phi_c = \phi_n = 0 \quad \text{or} \quad |\phi_\alpha|^2 + |\phi_\beta|^2 = \frac{\mu^2}{\lambda}, \tag{2.16}$$

only one of these solutions is viable, however, recall that ϕ_c consists of the charged components of the field; if $\phi_c \neq 0$ then charge conservation is broken. This violates the laws of physics we see in Nature so we are left only with the solution:

$$\phi_c = 0 \quad ; \quad |\phi_n|^2 = \frac{\mu^2}{\lambda}, \tag{2.17}$$

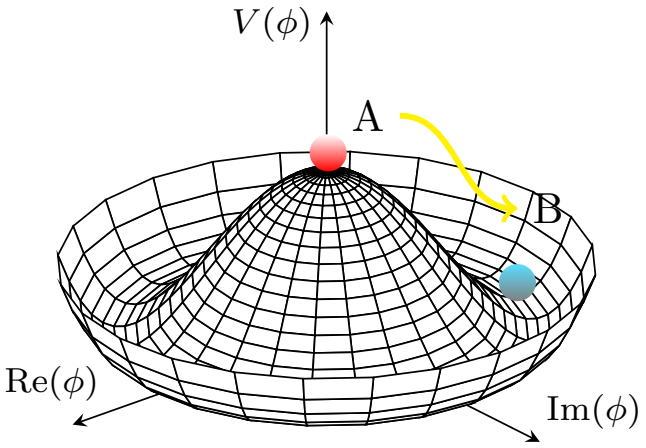
For convenience we express ϕ_n in terms of 'v', the vacuum expectation value (VEV), and 'h, η', the real and imaginary parts of the neutral ϕ component

$$\phi_n = \frac{1}{\sqrt{2}}(v + h + i\eta) \quad ; \quad v^2 = \frac{2\mu^2}{\lambda}, \tag{2.18}$$

We visualise the potential in Figure 2.1, we can see that there are numerous degenerate solutions, these solutions are degenerate as they are connected by an $SU(2)$ gauge transformation.

We started with four degrees of freedom in our complex, scalar Higgs doublet, when we attempt to make a choice of vacuum state however, we break this symmetry. After the

[7] Though as previously mentioned by Equation 2.1 there is an alternate commonly used parameterisation where it is 1 instead.



2.1 Diagram showing the shape of the Higgs potential. At point 'A' we have a red ball representing the Higgs boson before SB, in the local minima. In this state the potential appears as a well rather than the 'Mexican hat' shape. After SB the Higgs drops to point 'B', or any of the equivalent points in the ring around the origin of the same $V(\phi)$ value.

symmetry is broken we have three longitudinal modes from those that are no longer physical. As an example let us choose:

$$\phi_1{}^2 = \phi_2{}^2 = \phi_3{}^2 = 0 \quad ; \quad \phi_4{}^2 = \frac{2\mu^2}{\lambda} = v^2, \tag{2.19}$$

this choice is convenient as we know that $U(1)_Q$ is an unbroken symmetry in this process, as the Higgs has no electric charge the VEV should be chosen to lie within the neutral part of ϕ, i.e. ϕ_n from Equation 2.11. We can also see from looking at Figure 2.1 that there are solutions where the imaginary part of our doublet, η, from Equation 2.18, is equal to zero which brings us to Φ:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \tag{2.20}$$

where, h, is the *physical* Higgs boson. We still have three unphysical degrees of freedom to deal with, these can be "gauged away" as they are in fact a trio of Goldstone bosons as described [86][8].

An interesting point on the symmetries here, $SU(2)_L \otimes U(1)_Y$, has an approximate global symmetry in the limit where, $g' \to 0 \quad ; \quad h_f \to 0$. In this limit the Higgs sector has a global $SU(2)_R$ symmetry, therefore it is invariant under, $SU(2)_L \otimes SU(2)_R$, where $SU(2)_L$ is the *global* upgrade on the SM chiral gauge symmetry. This symmetry is not lost in the scenario where we have non vanishing Yukawa couplings and more specifically, if, $\mathbf{y}_u = \mathbf{y}_d$. When EWSB occurs both of, $SU(2)_L, SU(2)_R$, are broken

[8] For brevity we do not go into full details here, suffice to say that when a gauge symmetry is broken a Goldstone mode is created. This is normally a massless particle but in this case they become part of the Weak bosons instead and give rise to their masses.

however, a diagonal subgroup from them, $SU(2)_{L+R}$, forms a *custodial* symmetry. In the limit where, $g' \to 0$, the masses of the W, Z bosons are identical. Additionally, they form a triplet of this custodial symmetry. This means that the following relation/s are protected by the custodial symmetry:

$$\frac{{m_W}^2}{{m_Z}^2} = \frac{g^2}{{g'}^2 + g^2} = \cos^2\theta_W, \quad \texttt{or} \quad \rho \equiv \frac{{m_W}^2}{{m_Z}^2 \cos^2\theta_W} = 1, \tag{2.21}$$

We return now to focus on the broken symmetries themselves, they are often referred to as being 'eaten' by the Weak bosons. More mathematically we can express the Weak bosons in terms of a recombination of the original fields like so:

$$A_\mu = \cos(\theta_\omega) B_\mu + \sin(\theta_\omega) {W_\mu}^3, \tag{2.22}$$

$${W^\pm}_\mu = \frac{1}{\sqrt{2}}(W_1 \pm W_2), \tag{2.23}$$

$$Z_\mu = \sin(\theta_\omega) B_\mu - \cos(\theta_\omega) {W_\mu}^3 \tag{2.24}$$

here the, A_μ, field is the photon, and the, Ws, and, Z, are the usual Weak bosons. Along with this, the angle, θ_ω, is the Weak (Weinberg) mixing angle which can be defined as:

$$\cos\theta_\omega = \frac{M_W}{M_Z} \tag{2.25}$$

where, M_W, and, M_Z, are the masses of the W and Z bosons. If we now take what we have learnt and feed it back into the potential in Equation 2.14, then expand this about the VEV:

$$V(\Phi^\dagger \Phi) = \frac{{\mu}^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 \tag{2.26}$$

the first term here has the form of a mass term with, $m_h \equiv \sqrt{2\lambda} v$, while the second is a self-interaction parametrised by λ.

Next we consider the gauge-Higgs interaction from Equation 2.12, inserting our new Φ in place of ϕ:

$$|D_\mu \Phi|^2 = \frac{1}{2}(\partial_\mu h)^2 + \frac{1}{4} g^2 (v+h)^2 {W_\mu}^+ W^{-,\mu} + \frac{1}{8}(v+h)^2 (g^2 + {g'}^2) Z_\mu Z^\mu \tag{2.27}$$

We can now see in Equation 2.27 that the 2nd and 3rd terms are both mass terms for the $W^\pm$ and Z bosons respectively. The masses we obtain from this are[9]:

[9] We will not show the full derivation for the sake of brevity

$$m_W{}^2 = \frac{g^2 v^2}{4} \quad ; \quad m_Z{}^2 = \frac{m_W{}^2}{\cos^2\theta_w} \tag{2.28}$$

We also notice that A_μ does not appear in Equation 2.27, this is because the Higgs does not couple to the photon, and hence it is massless.

The Higgs field's doublet Nature allows for interaction with the fermions, thus they also gain mass upon SSB. The 3rd term in Equation 2.27 corresponds to this interaction. Specifically, we have the couplings, $\mathcal{Y}_\psi{}^{ij}$, where, $\psi = \ell, u, d$, i.e. the leptons, up-type quarks and down-type quarks. Examining the Lagrangian for the Yukawa sector to see this:

$$\mathcal{L}_Y = \mathcal{Y}_\ell{}^{ij} \bar{L}_{L,i} \Phi \ell_{R,j} + \mathcal{Y}_u{}^{ij} \bar{Q}_{L,i} \widetilde{\Phi} \ell_{R,j} + \mathcal{Y}_d{}^{ij} \bar{Q}_{L,i} \widetilde{\Phi} d_{R,j} + h.c. \tag{2.29}$$

which contains both left and right-handed multiplets for gauge invariance, and is a left-handed $SU(2)$ doublet with a Hypercharge value of, $Y = -1$, defined as:

$$\widetilde{\Phi} \equiv i\sigma^2 \Phi^* \tag{2.30}$$

Now we can read the masses directly off of Equation 2.29, for a general fermion, f, with field, ψ, as we know the form these terms will take:

$$\mathcal{L} \supset -m_f \bar{\psi}\psi = -m_f(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R) \tag{2.31}$$

where in the far RHS we have decomposed ψ into the two chiral components. Combining this with Equation 2.29 we get:

$$\mathcal{L} = -\sum_f \left[-\frac{\mathcal{Y}_f}{\sqrt{2}} v(\bar{\psi}_{f,L}\psi_{f,R} + \bar{\psi}_{f,R}\psi_{f,L}) - \frac{\mathcal{Y}_f}{\sqrt{2}} h(\bar{\psi}_{f,L}\psi_{f,R} + \bar{\psi}_{f,R}\psi_{f,L}) \right] \tag{2.32}$$

where the first term is the mass of a fermion arising from the coupling to the VEV, and the second term is the coupling with the Higgs boson itself. We have a Yukawa coupling for each individual quark and each charged lepton to the Higgs. The Cabibbo-Kobayashi-Maskawa (CKM) matrix comes into play here as the mass-eigenstate (physical) basis is not diagonalised. The CKM allows us to move into the Weak-eigenstate basis, where we can then see coupling between quarks in different Weak-Isospin doublets via $W^\pm$ interactions [90].

2.1.2 The Mass Of The Higgs Boson

The mass of the Higgs boson itself can seen in Equation 2.26, in the self interaction term:

$$V(\Phi)_{\text{HiggsSelf}} = \frac{\lambda}{4}(v + h)^4 \tag{2.33}$$

which leads us to a mass of:

$$m_H = \sqrt{2\lambda}v \tag{2.34}$$

where λ is the self-coupling parameter for the Higgs. This coupling is actually a free parameter in the SM, we must rely on measurements to establish it's value. The measured value is, $m_H = 125.10 \pm 0.14$ GeV, [90] and the VEV is, $v = (\sqrt{2}G_F)^{-\frac{1}{2}} \approx 246$ GeV which suggests that, $\lambda \simeq 0.13$.

The potential must have a real value thus it is necessary that the parameters, m_{11}^2, m_{22}^2, and, λ_{1-4}, are real. λ_{5-7} and m_{12}^2, however, can in general be complex.

2.2 Unsolved Problems In The SM

We mention here just a handful of the problems currently present with the SM:

- **Baryon asymmetry** - In Nature there is an abundance of matter but a notable lack of anti-matter by comparison despite the SM predicting equal amounts of each. An explanation for this scenario requires sizeable CP asymmetry. The complex phases within the Cabibbo-Kobayashi-Maskawa (CKM) matrix are far too small to account for the ratio of matter to antimatter in the universe [64]. This is one of the 3 'Sakharov conditions' that are required for observed matter-antimatter ratio. The other two are; baryon number violation and a departure (of the universe at large) from thermal equilibrium, as if we did not have this then any process would be as likely to happen in either direction and thus there would be no way to create the baryon asymmetry from our starting point at the big bang [80].

- ***CP*-violation** - One problem that we have with the SM is the presence of CP-violation, with C being charge inversion and P being Parity. Both the Strong and EM forces acting on particles preserve C and P symmetry both separately and as the product, CP; the Weak force, however, violates C and P separately as well as CP in the case of neutral kaon decays [54] and more recently neutral charm and B mesons [113]. While there have been candidate theories [108] involving phases in the CKM mixing matrix, there is no conclusive evidence that these are truly the source of CP-violation.

- **Dark matter/energy** - It is well known in the physics community that there appears to be a form of invisible matter providing galaxies with additional velocity that cannot be explained merely by the mass of visible matter alone.

This is seen through observations on galaxy rotations and gravitational lensing [90].

- **Neutrino masses** - SM neutrinos should be massless, however, various experiments [79] [87], have demonstrated that this cannot be the case. These experiments saw evidence that neutrinos change flavour as they propagate, such oscillations are only possible if neutrinos have mass as they require there to be multiple mass eigenstates [127].

- **The hierarchy problem** - The measured mass of the Higgs boson is surprisingly low at about 125 GeV, this is very far from the gravitational energy scale, $\mathcal{O}(10^{19})$ GeV. This presents a problem because first order radiative corrections[10] to the Higgs mass are quadratically proportional to scale Λ, the potential New Physics (NP) scale energy scale at which the SM becomes invalid. The corrections to the Higgs mass are as follows:

$$\delta {m_h}^2 \equiv {m_h}^2 - ({m_h}^0)^2 \propto {\Lambda_{NP}}^2 \left[+\frac{1}{4}(9g^2 + 3g'^2) - {\mathcal{L}_t}^2 \right] \tag{2.35}$$

where ${m_h}^{0^2}$ is the *bare Higgs mass*, and Λ_{NP} is the potential New Physics (NP) scale scale. ${\mathcal{L}_t}^2$ is the Yukawa coupling to the top quark. This means that the mass of the Higgs is very sensitive to parameters like the Yukawa couplings, if there is no new physics up to, $\frac{{m_h}^2}{\Lambda^2}$ 10^{-34}, the bare mass required to produce the measured Higgs mass would be very, very, large. If the mass corrections are of a similar size to the mass itself then this is fixed through cancellations, **but** we have what is known as a "*finely tuned*" cancellation of the terms. This is generally considered to be unnatural [76]. To solve this 'hierarchy problem' we need new physics at the TeV scale.

2.3 The 2HDM

2.3.1 Physics Of The 2HDM

In the 2HDM, we extend the SM scalar sector by including two complex doublets (i.e. one additional doublet alongside the pre-existing Higgs doublet), Φ_1 and, Φ_2, which eventually gives rise to two CP-even scalars, h and, H, one pseudo-scalar, A, and two charged Higgs bosons, $H^{\pm}$. Note that this leaves us with the same Weak bosons as seen in the SM, as we go from, $2 \times 4 = 8$, d.o.f. with our two doublets, then this breaks to give us 5 Higgs bosons, leaving 3 d.o.f. once again to be "eaten" by the Weak bosons and form their longitudinal modes. Extensive reviews on these models can be found in Refs. [92][36][69].

[10] Which arise due to renormalisation

As we did for the SM Higgs, we can consider a general potential for the doublets in question. The most general, renormalizable, scalar potential of the two doublets is as follows[11]:

$$\begin{aligned}\mathcal{V}_{gen} = & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - \left[m_{12}^2 \Phi_1^\dagger \Phi_2 + \text{h.c} \right] + \\ & + \frac{1}{2}\lambda_1 \left(\Phi_1^\dagger \Phi_1 \right)^2 + \frac{1}{2}\lambda_2 \left(\Phi_2^\dagger \Phi_2 \right)^2 + \lambda_3 \left(\Phi_1^\dagger \Phi_1 \right) \left(\Phi_2^\dagger \Phi_2 \right) + \lambda_4 \left(\Phi_1^\dagger \Phi_2 \right) \left(\Phi_2^\dagger \Phi_1 \right) + \\ & + \left\{ \frac{1}{2}\lambda_5 \left(\Phi_1^\dagger \Phi_2 \right)^2 + \left[\lambda_6 \left(\Phi_1^\dagger \Phi_1 \right) + \lambda_7 \left(\Phi_2^\dagger \Phi_2 \right) \right] \Phi_1^\dagger \Phi_2 \right\}\end{aligned} \tag{2.36}$$

where $v_{1,2}$ are the VEVs of each doublet, $\tan\beta = v_2/v_1$, is given by also known as CP-odd mixing angle. m_{11}^2, m_{22}^2, m_{12}^2 are the mass squared parameters and, λ_i $(i = 1, ..., 7)$, are dimensionless quantities describing the coupling of the order-4 interactions.

Six of the parameters we are left with are real, (m_{11}^2, m_{22}^2, λ_i with $i = 1, ..., 4$), and four are a priori complex, (m_{12}^2 and λ_i with, $i = 5, ..., 7$). Therefore, generally, the model has 14 free parameters. Under appropriate constraints, however, it is possible to reduce this number.

λ_{5-7} and m_{12}^2 can in general be complex, but in the case of the 2HDM the non-zero imaginary parts of the parameters can be removed via a re-phasing transformation.

After EWSB, each doublet acquires a Vacuum Expectation Value (VEV) that can be parametrised as follows:

$$\langle \Phi_1 \rangle = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ \cos\beta \end{pmatrix} \quad ; \quad \langle \Phi_2 \rangle = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ \sin\beta \end{pmatrix} \tag{2.37}$$

which gives, $v \equiv \sqrt{{v_1}^2 + {v_2}^2}$, as the total vacuum expectation value experienced by the gauge bosons.

If we can introduce the requirement for initial CP-conservation we remove several degrees of freedom. The potential is explicitly CP-conserving if and only if there exists a basis of the scalar fields in which, m_{12}^2, λ_5, λ_6, and, λ_7 are all real. Even with this explicit CP-conservation it is still possible for the vacuum to break CP spontaneously. Spontaneous CP-violation of the vacuum can occur if and only if the scalar potential is explicitly CP-conserving **and** there is no basis in which the scalars are real [94]. By requiring CP-conservation, one loses four further d.o.f. taking the number of free parameters down to 10.

We also have an enlarged Yukawa section in the Lagrangian, as both doublets have equal numbers of couplings:

[11] We avoid unnecessary steps such as defining the Lagrangian of the 2HDM, as clearly it will be the same as the SM but with a second Higgs term, two VEVs and a potential dependent on two Higgs fields.

$$\mathcal{L}_Y = \sum_{i=1}^{2} (\mathcal{Y}_\ell{}^{ij} \bar{L}_{L,i} \Phi \ell_{R,j} + \mathcal{Y}_u{}^{ij} \bar{Q}_{L,i} \widetilde{\Phi} \ell_{R,j} + \mathcal{Y}_d{}^{ij} \bar{Q}_{L,i} \widetilde{\Phi} d_{R,j}) + h.c., \tag{2.38}$$

In general, the Yukawa matrices corresponding to the two doublets are **not** simultaneously diagonalisable; presenting us with a problem. The off-diagonal elements would lead to tree-level Higgs mediated Flavour Changing Neutral Currents (FCNCs) which have been tightly bounded by experiment.

Turning to the Glashow-Weinberg-Paschos (GWP) theorem [85] [122] we find that such FCNCs are absent if, and only if, at most one Higgs multiplet is responsible for providing mass to fermions of each electric charge[12]. All that is needed to enforce is the application of a discrete $\mathbb{Z}_2$-symmetry onto the doublets, i.e. s.t. $\Phi_1 \to +\Phi_1$, and, $\Phi_2 \to -\Phi_2$. This method allows the lack of FCNCs to arise in a natural way. This $\mathbb{Z}_2$ symmetry can be softly broken through a mass parameter (denoted by m_{12}) mixing the two Higgs doublets without falling outside of current experimental limits. The soft $\mathbb{Z}_2$ breaking condition relies on the existence of a basis wherein, $\lambda_6 = \lambda_7 = 0$. Requiring the soft breaking then removes two additional d.o.f., taking the total number of free parameters down to 8.

Next, m_{11}^2, and, m_{22}^2, can both be expressed as a function of the other parameters; this is due to the fact that the scalar potential is required to be in a local minimum at the VEVs of the fields. So, globally, with restrictions to CP-conservation and soft $\mathbb{Z}_2$-SB applied, there remain 7 free parameters in the 2HDM. Finally, the mass of the light Higgs boson has been measured to high precision, factoring this in we have arrived at just 6 free parameters.

Under the $\mathbb{Z}_2$-symmetry we have introduced, the fermions are required to have a definite charge; there are numerous ways that these charges can be assigned and these correspond to the four types of 2HDM. Table 2.1 details the model types and the couplings of the neutral Higgs scalars to fermions (relative to the SM value of $\frac{m_f}{v}$).

Model	h			H			A		
	u	d	l	u	d	l	u	d	l
Type-I	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\cot\beta$	$-\cot\beta$	$-\cot\beta$
Type-II	$\frac{\cos\alpha}{\sin\beta}$	$-\frac{\sin\alpha}{\cos\beta}$	$-\frac{\sin\alpha}{\cos\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\sin\beta}$	$\cot\beta$	$\tan\beta$	$\tan\beta$
Type-III	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\sin\beta}$	$-\frac{\sin\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$-\cot\beta$	$\cot\beta$	$-\tan\beta$
Type-IV	$\frac{\cos\alpha}{\sin\beta}$	$-\frac{\sin\alpha}{\cos\beta}$	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$-\cot\beta$	$-\tan\beta$	$\cot\beta$

2.1 Couplings of the neutral Higgs bosons to fermions, normalised to the corresponding SM value, (m_f/v), in the 2HDM Type-I, II, III and IV.

[12] This has the useful side effect of providing an additional source of CP-violation while leaving flavour conservation intact [125]

2.3.2 Benefits Of The 2HDM

While 2HDMs alone often cannot fully explain the discrepancies in the SM, a second Higgs doublet naturally arises embedded in many models of BSM physics that *can* remedy these 'flaws' in the SM. For example, there is a class of axion models [49; 105], which can explain the lack of observed CP-violation in the Strong sector. The 2HDM can be found embedded in realisations of more complete theories, such as Supersymmetry (SUSY), e.g., the Minimal Supersymmetric Standard Model (MSSM) requires the existence of at least two doublets. The representative model chosen in our first research paper, the 2HDM Type-II, coincides with the MSSM [92] [69] more generally; and the CP problem is clearly solved within a particular MSSM with explicit CP-violation [126].

There are also certain realisations of composite Higgs models with pseudo-Nambu-Goldstone bosons [119; 32; 4; 61; 62], which can give rise to an effective low-energy theory with two Higgs doublets. The additional source of CP-violation present in this type of enlarged (pseudo)scalar sector could also provide an answer to the matter-antimatter asymmetry. There are also some realisations of the 2HDM with the appealing features of being able to provide an explanation for neutrino mass generation [8], to provide a candidate for dark matter [106] or to accommodate the muon g-2 anomaly [48; 37; 148]

Discussion of these models, however, is beyond the scope of this thesis. The focus is on discovery of a 2HDM signal, motivated by such models, but not studying them extensively.

Chapter 3

Magellan - A Markov Chain Monte Carlo Driven Parameter Scanner

3.1 Standard Methods Of Parameter Space Investigation

High Energy Physics (HEP) often involves problems with high dimensional parameter spaces, this presents a problem for physicists due to the "curse of dimensionality"[60] - a term used to refer to the exponential increase of sampling needed to build up a sample density compared to what can be attained for a far smaller sample when exploring a lower dimensioned space. This makes fully mapping out a parameter space difficult. We do not know if the parameter space is fractured throughout or what regions our valid points may reside in; this can lead to wasting computing resources exploring regions that turn out to be devoid of any useful points. The region that *does* contain useful points is known as the 'typical set'[1].

For models with high dimensional parameter space, the standard way of exploring it and extracting bounds is done by projecting the full parameter space onto bi-dimensional planes, defined by any two model parameters. The statistical procedure here is to maximise the (log-)likelihood on all of the other remaining parameters. Physicists may also attempt to reduce the dimensionality of the problem by setting one or more parameters to a constant value; this is not ineffective, however, it does result in the loss of useful information contained in the regions with differing values for those parameters and certainly cannot be considered to provide a full picture.

`Magellan` was borne out of the desire to improve upon this standard method, we sought to tie together the benefits of `2HDMC` [74], `HiggsBounds` [24], `HiggsSignals` [26] (such as the constraints from both theory and observation) and `T3PS` [116] to create a

[1] This concept comes from the geometry of high-dimensional probability distributions. The ways in which density and n-volume behave with increasing dimensionality leads to a very small region containing useful information [33]; it is a waste of computation time to look elsewhere.

smooth pipeline which may be used to explore the parameter spaces of different types of 2HDM. We hoped that it would be possible for this to be generalised to more models with extended Higgs sectors and the further to other BSM theories. But first, the 2HDM was chosen as a proof of concept model, we chose this in particular due to its combination of being both physically interesting and, crucially for our purposes, being characterised by a multi-dimensional parameter space that is far from trivial.

An important task for phenomenologists is to extract limits on BSM theories from the data collected from relevant experiments; to facilitate the rigorous statistical analyses for this goal a range of global fits have been developed in the last few decades. One of the earliest of such global fits for an array of models (including the 2HDM), was the package GFitter [17], with a global EW fit to constrain New Physics (NP). There are many more toolkits published in the literature which focus mainly on SUSY and the variants thereof. SFitter [112], SuperBayeS [16; 140; 141], Fittino [21; 22; 23], Lilith [109; 31] and MasterCode [38; 39; 40; 41; 42] all provide global analyses for SUSY. Gambit [10; 11] covers a considerably wider range of BSM theories, including the MSSM in Ref. [11] and the 2HDM in Ref. [128], specifically.

The standard procedure for these global fitting packages generally involves the use of all relevant experimental data as well as theoretical arguments that can constrain the model. Within Magellan these constraints fall into three main sources/categories: measurements of the 125 GeV Higgs boson, seen at the LHC, properties (i.e., production and decay signal strengths), direct and indirect searches for the extra Higgs bosons needed in the model and, finally, theory restrictions from perturbativity, unitarity, triviality and vacuum stability. The statistical analysis is then performed, using a (log-)likelihood function to express how plausible different parameter values for the given sample(s) of data are.

Something to contend with is the disconnected nature of BSM parameter spaces. They may not be comprised of a single region of allowed points with all points outside that region being disallowed. Instead there may be pockets of allowed points spread out throughout large areas of disallowed points. We do not want to accidentally miss these points when performing our mapping, thus it is important that our methodology ensures such points are captured. The parameter space of 2HDMs is large and multi-dimensional, making it difficult to pinpoint ideal regions for searches. There are some excellent pieces of software already in existence which can be used to check a χ^2 measurement which is calculated using measured Higgs rates and masses [26], validity against exclusion bounds from experiment [24] and theoretical properties of a model [74]. This toolbox aims to combine such packages along with the Markov Chain Monte Carlo (MCMC) generator `T3PS` [116] in order to streamline searches in the parameter space.

The 2HDM parameter space can be considered as six-dimensional[2], following the pre-established standard methods of fitting we would extract bounds by projecting the

[2]Following the discussion in section 2.3

full parameter space onto bi-dimensional planes, which would be defined by any two model parameters in pairs. In `Magellan` we attempt to improve upon this method. It should be trivial to extend `Magellan` to other extended Higgs sector-theories and simple to extend the statistical techniques and tools here to any BSM theory. Later in this work we will demonstrate the effectiveness and efficiency using the 2HDM, for now, however, we will cover the tool itself.

First `Magellan` performs an efficient scan over the parameter space using an MCMC approach with `T3PS` [116]. `T3PS` is a `Python` 2 based MCMC generator. It initiates and manages the chains as they run and generates each point in the chain. After this first step, instead of projecting the parameter space onto bi-dimensional planes by maximising the (log-)likelihood over the rest of the parameters, we keep all of the information on all the model parameters, simultaneously. This means that the parameter space of the model can be projected onto a series of bi-dimensional planes but retain all parameter correlations so that it is possible to map some select region of one of this bi-dimensional sub-spaces onto any other bi-dimensional plane. Hence the user has full control of the entirety of the parameter space. This means that we can investigate the associations between different properties of the lower dimensional sub-spaces which constitute the full parameter space.

The use of an MCMC ensures that we capture disconnected regions in the parameter space. It achieves this using multiple chains simultaneously, each starting at randomised positions and, as per the Metropolis-Hasting algorithm, occasionally moving to a less likely point in order to ensure sufficient exploration of the parameter space. `Magellan` also allows for a "heating algorithm" to be added if the user feels it is needed, this will temper the chains via `T3PS`'s built in mechanism. Tempering is a group of methods used to increase the speed at which an MCMC converges, a popular type of Tempering is simulated Tempering (or Annealing MCMC)[83][115]. In Simulated Annealing (SA) we temper our MCMC by replacing the prior distribution, $\pi(\theta)$, with a *set* of tempered distributions, $\pi_k(\theta)$, where, $k \in [0, \infty]$, and is known as the *inverse-temperature parameter*. There are then different ways this can be used, *Powering up* is an example.[95] In this method we have, $\pi_k(\theta) \propto \pi(\theta)^k$. Here, small values of, k, act to flatten/widen peaks as well as to raise troughs that may exist in, π.

When it comes to choosing a definition for $\pi_k(\theta)$ a common choice is to base it on the Boltzmann-Gibbs transform of, $\pi(\theta)$,

$$\pi_k(\theta) \propto \exp^{\frac{\pi(\theta)}{T_k}} \tag{3.1}$$

which bears strong similarity to the calculation performed in thermal dynamics to find the probability that energy magnitude increases: $P(\Delta E) = \exp^{-\frac{\Delta E}{k \cdot t}}$
in which k is the Boltzmann constant and t is temperature. Returning to Equation 3.1,

provided that the expression on the RHS is integrable[3] the sequence of temperatures is set to decrease such that, $\pi_k(\theta) \to 1$. There is no set method one should use for the temperature reduction though they generally fall into one of three categories:

- Geometric reduction: $t = t \cdot \alpha$, where α is the *temperature reduction function*
- Linear reduction: $t = t - \alpha$,
- Slow-Decrease : $t = \frac{t}{1+\beta t}$, where β is an arbitrary constant.

Implementing such a method causes the algorithm to draw samples from a smaller and smaller neighbourhood which should, in theory, draw closer to the maximum/maxima of, $\pi(\theta)$. An in-depth discussion of the scenarios each of these is best suited to a thorough review of Tempering methods in MCMCs, is beyond the scope of this thesis.

There is a website for `Magellan` which has interactive dashboards and the data from one of our investigations using `Magellan`. It can be accessed via a public link `Magellan` website. Through this website, the user can explore the entirety of the parameter space and survey the phenomenological features of a BSM model with ease. Eventually the intention is for it to be possible for new data analyses to be imported onto the website, thus allowing a live description of the BSM theory at hand.

The second application concerns the analysis of the extra CP-even Higgs boson decaying into two light Higgs bosons at the LHC within the 2HDM Type-II. This analysis is already public on the Magellan website. There, also the analyses of both the heavy CP-even and CP-odd Higgs boson decaying into tau pairs are published. These latter studies represent extensions of the main analysis carried out in chapter 4.

3.2 Experimental Constraints

- `2HDMC` calculates the six oblique parameters, S, T, U, V, W and X which can then be compared to current experimentally measured values. For this purpose it is incorporated into `Magellan` and by default the oblique parameter results are taken as part of the output of the scanner section. In BSMs the oblique parameters provide corrections to the gauge boson propagators, in this case due to the additional Higgs bosons, and thus affect the precision observable values for a given model. The precision observables themselves include high precision measurements from experiments; at the Z-pole[4] the properties of the $W^{\pm}$ bosons as well as the masses of the SM-like Higgs and top quark[66]. These measurements have a very high level of precision and provide sensitivity to the level of radiative corrections on the observables, as well as strong constraints on any physics model which modifies the EW sector.

[3] It seems unlikely that this would not be the case for any potential user of `Magellan`, however if it **not** then this explicitly cannot be used.

[4] Including the decay widths of the Z, the effective weak mixing angle, left-right and forward-backward asymmetries, the mass of the Z and couplings to fermions.

- In addition, `2HDMC` calculates a_μ, which is the contribution to the anomalous magnetic moment of the muon coming from the extended Higgs sector.

- A χ^2 value is determined by taking the one given by `HiggsSignals`, one calculated using the values from `HEPfit` [65] or from the values in `Gfitter` [96][5]. These are incorporated into the scanner, specifically in the "ParameterProcessor".

- `HiggsSignals` takes the predictions for 2HDM scalar sectors and compares them to the signal rate and mass measurements of the SM-like Higgs at the LHC. Through this comparison and a *peak-centred* χ^2 *method* a likelihood estimate is calculated. This attempts to minimise the total χ^2, given by:

$$\chi_{tot}{}^2 = \chi_\mu{}^2 + \sum_{i=1}^{N_H} \chi_{m_i}{}^2, \tag{3.2}$$

  Where $\chi_\mu{}^2$ are the signal strength modifiers $\chi_{m_i}{}^2$ are the Higgs masses and N_H is the number of neutral Higgs bosons in the user's model. The calculation of $\chi_\mu{}^2$ involves various uncertainties such as those on the squares of the signal strength squared to 1σ, the statistical and systematic experimental uncertainties, that of the luminosity etc. We highly recommend the third chapter of [26] for a detailed look at this.

  `Gfitter` and `HEPfit` have produced values of best fit for S and T at U=0 [97][65] which are used to compute χ_{ST}. The user may wish to switch to the values found in the same papers for $U \neq 0$ which can be done in `ParameterPointProcessor/src/EWPO.cpp`, these are combined with `HiggsSignals` output χ^2 value[6], assuming that the values are independent and thus a correlation matrix is not required, like so

$$\chi^2_{tot} = \chi^2_{HS} + \chi^2_{ST}, \tag{3.3}$$

  where χ^2_{HS} is the χ^2 calculated by `HiggsSignals` using a peak-centred method and χ^2_{ST} is the one for an S and T compatibility measure given by

$$\chi^2_{ST} = \frac{(S - S^{exp}_{bestfit})^2}{\sigma^2_S(1-\rho^2_{ST})} + \frac{(T - T^{exp}_{bestfit})^2}{\sigma^2_T(1-\rho^2_{ST})} - 2\rho_{ST}\frac{(S - S^{exp}_{bestfit})(T - T^{exp}_{bestfit})^2}{\sigma_T\sigma_S(1-\rho^2_{ST})}, \tag{3.4}$$

[5] This is decided by the user's preference.

[6] By definition χ^2 distributions that are independent from one another can be summed to give a new distribution that is *also* a χ^2 distribution

where the best fit values, $S^{\text{exp}}_{\text{best fit}}$, and, $T^{\text{exp}}_{\text{best fit}}$, their uncertainties $\sigma_{S/T}$ and the correlation parameter, ρ^2_{ST}, are taken from the fit result of the `Gfitter` group [96][7].

- The initial version of `Magellan` used `HiggsSignals` 2 [25] beta version, which includes experimental data tables collected until September 2018. `HiggsSignals` compares the 2HDM predictions for the scalar sector with the SM-like Higgs boson signal rate and mass measurements at the LHC, giving rise to a likelihood estimate. The `HiggsSignals` authors released a recent validation (see Ref. [27]) against the ATLAS and CMS combined analysis of the LHC Run1 data at 7 and 8 TeV [12; 1].

- `2HDMC` interfaces with `HiggsBounds` to provide checks on points against the current experimental limits. This includes results from LEP, the Tevatron, and LHC. For exact details on the included results please check the `HiggsBounds` website `HiggsBounds` gitlab or latest paper [24]. `HiggsBounds` works by calculating the ratio of, $Q_{model(X)}$, to, $Q_{expec(X)}$, where the former is the model prediction for each Higgs signal topology and the latter is the *expected* 95% C.L. exclusion bound on that quantity. `HiggsBounds` evaluates which search topology, X_0, has the highest value of this ratio, and thus the highest statistical sensitivity for exclusion of that model prediction. It then calculates the ratio

$$\frac{Q_{model}(X_0)}{Q_{obs}(X_0)} \tag{3.5}$$

where, $Q_{obs}(X_0)$, is the cross section for the chosen topology. The parameter point is excluded if this ratio is larger than 1.

As the toolbox is set up for the example of a 2HDM there are several constraints that may apply only to these models that the user may wish to either remove completely or alter the values of in order to fit the model they are interested in[8]:

- The first such limit concerns the mass of $m_{H^\pm}$ is in the file; "`job_submission/MCMC/utils/dat_to_csv.py`", it is set to, $m_{H^\pm} \geq 580$ GeV,[118][147]. Values failing this limit will be dropped from the data so the user must be sure to comment it out or remove it fully if they are not using a model with this bound.

- There is a bound on the lowest value $\tan\beta$ may take of 5[147], this can be found in; "`job_submission/MCMC/utils/dat_to_csv.py`".

[7] This was chosen by the authors as opposed to the `HEPfit` option as the work they are taken from came out slightly more recently, however both are built in to `Magellan`

[8] Ideally we would like to simplify the way a user interacts with this aspect to avoid them needing to hunt down constraints; however, time constraints prevented a more generalised overall interface for `Magellan` being created.

- In addition to the constraints on, κ_{qq}, discussed previously, there is a set of bounds on, κ_{uu}, only, taken from [135], these are

$$-0.9 < \kappa_{uu} < -0.7 \quad or \quad 0.7 < \kappa_{uu} < 1.1 \tag{3.6}$$

  The values in Equation 3.6 are applied after the MCMC has finished its' runs, directly to the output data within the file `utils/dat_to_csv.py`. See subsubsection A.2.5.1 for details on this.

3.3 Statistics

3.3.1 Bayesian Statistics

Deriving its name from the well known statistical theorem, 'Bayes' theorem, Bayesian statistics builds on said theorem to create a method of data analysis and parameter estimation that aims to incorporate both observed and unobserved parameters into a joint probability distribution [63]. It is fundamental to the scientific method, reflecting the process of continuously evaluating our current understanding (our prior) then updating this based on the most recent data and the accompanying likelihood in order to improve our understanding (to get our posterior). The key difference between Bayesian and Frequentist statistics is that frequentists do not attach a probability to either their hypotheses nor to any unknown properties, but instead use fixed values generally, with the unspoken assumption that their hypothesis is true.
Equation 3.7 shows Bayes' Theorem, the components are defined as follows:

$$\underbrace{P(A|B)}_{Posterior} = \frac{\overbrace{\mathcal{L}(B|A)}^{\mathcal{L}ikelihood}\overbrace{P(A)}^{Prior}}{\underbrace{P(B)}_{Evidence}} \tag{3.7}$$

where, A, B, are events, $P(A)$ is the probability of event A and, $B|A$, indicates "B given A" either within the calculation of probability, $P()$, or likelihood, $\mathcal{L}()$ [123].

- **Posterior** (probability): This is the probability of the parameters, 'A', given that we have observed evidence, 'B'. As can be seen in Bayes' theorem, this incorporates the information known about the probability of observing that evidence, our prior understanding and the Likelihood function for our parameters & evidence.

- **Likelihood**: Typically in colloquial speech probability and likelihood are used interchangeably, however in Bayesian statistics (or indeed statistics generally), the likelihood is a property of the parameter values and probability of the sample. The likelihood function is the subtlest part of the equation, it is the

probability of observing the values seen in the evidence *for given parameters*. I.e., the probability that certain parameter values could produce the evidence. When the likelihood function is evaluated on some sample of data it tells us which parameter values are the most likely for that data.

- **Prior** (probability): This is the probability distribution that comes from our understanding *before* we take new evidence into account. For example, before the discovery of the SM-like Higgs boson at the LHC the prior probability distribution for the mass of the particle covered a much broader range of masses than the posterior after a signal was detected in the region of 125 GeV.
- **Evidence**: Sometimes referred to as the 'Marginal Likelihood', this is the probability of generating the observed data with the prior distribution.

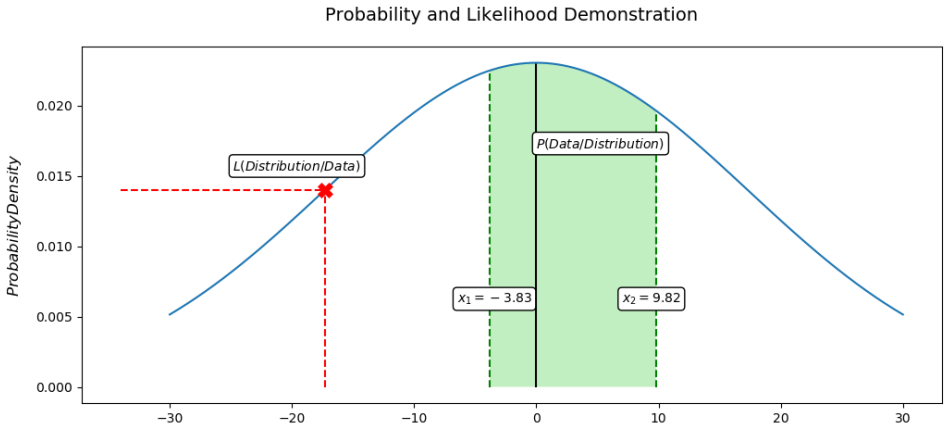
Bayes' factor is defined as the ratio of the likelihood of two different models being considered for some data D:

$$\mathcal{B}_{12} = \frac{\mathcal{L}(\mathcal{D}|M_1)}{\mathcal{L}(\mathcal{D}|M_2)}, \tag{3.8}$$

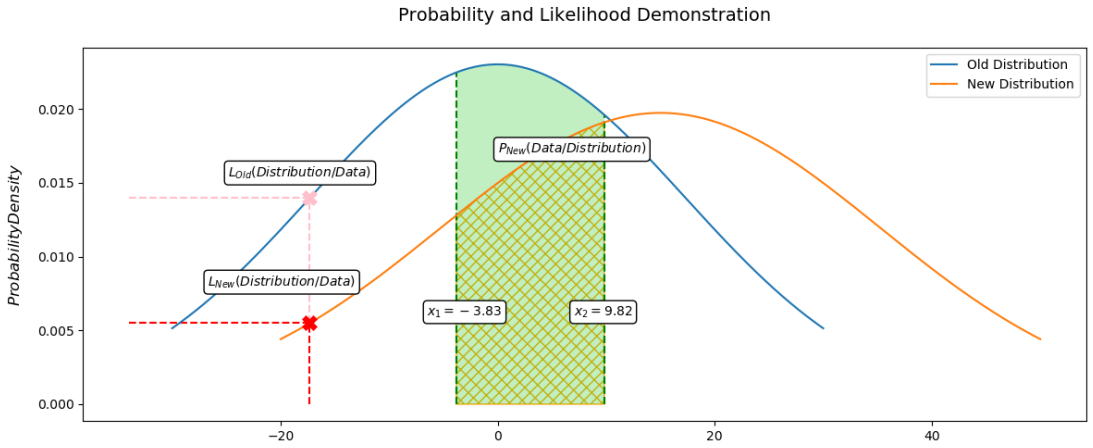
where, $M_{1,2}$, are the models under consideration.

Before moving further we feel it is necessary to emphasise the difference between likelihood and probability as it is an important distinction to make.

Simply looking at Figures 3.1 and 3.2 it is easy to gain an overall idea of the difference between probability and likelihood; probability is given by the area under a distribution between points of interest and likelihood is the probability density at a specific, unchanging point for a chosen distribution. We see the probability across the same region for two different distributions in Figure 3.2 differs, probability is ***of*** *the data* for a ***given*** *distribution*. We also see the likelihood for two different distributions



3.1 We see a partial plot of the normal distribution centred on 0. The section under the curve filled in, in green is the *probability* that a point lies between, x_1&x_2. The red cross indicates where the likelihood of the corresponding x value.



3.2 We see two partial plots of normal distribution centred on 0 and ≈ 16. The section under the blue curve filled in, in green is the *probability* that a point in the blue distribution lies between, x_1&x_2. The orange hatched region is the probability that a point in the orange distribution lies between, x_1&x_2. The pink cross indicates where the likelihood of the corresponding x value for the blue distribution, this is the same point as the red cross in Figure 3.1. The red cross now indicates the likelihood of the same x value but for the orange distribution.

at the same x-value; the likelihood changes for the two distributions and we can say that it is the likelihood ***of*** *a distribution* ***given*** *our data value.*

When using a likelihood function we must consider our results/data as parameters of the likelihood function itself unlike when considering a probability function. On the other hand the probability is a function of the data. We also note that a single likelihood value cannot provide very much information on its' own (though it still has use), it is only when we are able to consider the *ratios* of likelihoods that we begin to understand our data. This is clear in Figure 3.2, we have two likelihood values for the same data point which individually do not provide much information but through comparison can tell us which of our two distributions is a better fit to our data point.

Likelihood, generally denoted $\mathcal{L}$, is defined for a given process (specific to that process) and has a value equal to that of the proposed parametric Probability Density Function (PDF), f, for that process. The likelihood and PDF are evaluated with model parameters, $\boldsymbol{\theta}$, and a data sample $\mathbf{x}$:

$$\mathcal{L}(x, \theta) = f(x, \theta), \tag{3.9}$$

Our interpretation of our likelihood is determined by which argument we consider to be free [63]:

- When, $\boldsymbol{\theta}$, is held fixed and, $\boldsymbol{x}$, is free, we can view the likelihood as a PDF with a specific set of parameters, $\boldsymbol{\theta}$. Just as we would have with a regular PDF, the integral over, $\boldsymbol{x}$, of the likelihood is one.

- When we consider, $\boldsymbol{x}$, to be fixed and, $\boldsymbol{\theta}$, free; the likelihood is *not* equivalent to a PDF of, $\boldsymbol{\theta}$, any longer. Instead this likelihood is a single PDF value for a specific sample, $\boldsymbol{x}$, with parameter, $\boldsymbol{\theta}$, one of a family of PDFs indexed by, $\boldsymbol{\theta}$. The integral of the likelihood over, $\boldsymbol{\theta}$, for a specific, $\boldsymbol{x}$, will generally not give a value of one (though this is not impossible in some cases). In this scenario we may view the likelihood as a measure of the relative preference for each value of, $\boldsymbol{\theta}$.

3.4 Markov Chain Monte Carlo simulations

We will now discuss the basics of how a Markov Chain Monte Carlo (MCMC) simulation works and what makes them useful in our attempt to search and map the 2HDM parameter space.

An MCMC is a method used in statistics and computing for a variety of purposes; there are many applications within mathematics, computational biology, speech processing, medicine and naturally computational physics amongst others. The focal point of this method for our purposes is the relationship to likelihood; the algorithm essentially "seeks out" areas with the highest likelihood, saving a great deal of time compared to performing a simple grid-like scan. We will briefly cover several key concepts/terms which will prove relevant to our understanding of MCMCs:

- The *Markov property* is the name given to the 'memoryless-ness' of a stochastic process; memoryless in statistics means that information on previous states of a system is not retained in the present state and does not affect any future state. Another way of describing this is to say that we can make predictions about future states with the knowledge of the present state only and this will produce results that are as good as those made with the full history of a process; given the present state the future and past states are independent of one another. A Markov chain is a sequence of steps between points in a larger parameter space such that each step takes one to a point/state that satisfies the Markov property. Strictly speaking, a real, (rather than idealised mathematical) Markov chain in an MCMC does **not** have this perfect independence from point to point. Any two terms within the chain, $X_t \& X_{t+n}$, are not completely independent; however, the larger the number of steps between them is, the closer they get to being independent. Ways to mitigate this problem will be discussed later in this section.

- 'State space' is a term originating in computer science for the mathematical representation of a given problem which fully defines all possible states the problem can take at any given time. The size of an explored region of a state space is given by, B^D, where, B, is the branching factor, i.e. how many possible

steps (or directions) are available at any given point and, D, is how many steps have been taken[9].

MCMCs are commonly used to find optimal points within a large parameter space, or simply map out the available space after constraints are applied. Within `Magellan`, the parameter space is defined by implementing all known theoretical and experimental constraints; this is largely covered by `2HDMC`, `HiggsBounds` and `HiggsSignals`. In effect an MCMC hones in on regions with highest likelihood as a fundamental part of the algorithm [10], this reduces the time to find valid regions compared to searching via an inefficient grid-like scan across the parameter space.

3.4.1 The Likelihood Function

The likelihood function, $\mathcal{L}(d|\theta)$, uses the value of χ^2_{tot}, which itself was defined previously in Equation 3.3, and it is as follows:

$$\mathcal{L} = \exp\left(-\frac{\chi^2_{\text{tot}}}{2}\right) \tag{3.10}$$

The MCMC algorithm first attempts to find the minimum of the likelihood, before before it converges to thermal equilibrium. This means there is a "warm-up" or "burn-in" period of sorts, to account for this the first 200 steps[11] are discarded within every chain. This is done before the application of theoretical constraints (perturbativity, unitarity, triviality and vacuum stability). In general the choice of prior distribution can impose unintended effects on the sampling, (unless it has been chosen with extreme care), therefore we use a uniform distribution as our initial prior. This is an *uninformative* prior, meaning it does not carry information on our parameter space within it. An uninformative prior is essentially determined by the "principle of indifference" - we do not have information to suggest that there is a preference for any parameter value and thus assume they are all equally probable.

`Magellan` employs an MCMC method based on `T3PS` [116] for the parallel processing of parameter scans. This falls into the category of MCMCs that use the standard Metropolis-Hastings [117; 100] algorithm, which is outlined succinctly below.

- **Step 1)** First a starting point is drawn from the prior distribution $\pi(\theta)$. This point begins the chain. The likelihood corresponding to this point is given by,

[9] Specifically, B, is the branching factor, D, is the search depth and, B^D, is the size of the state space thus far. In order to avoid confusion between the full, pre-existing state space that is equivalent to our constrained parameter space and the *explored* state space which is the region of the state space that has been explored/mapped up to a given time we will use parameter space to refer to the full space and explored/mapped state space for the explored sub-space therein.

[10] Though this does not mean that any given MCMC chain will reach this theoretical highest region as spaces may be disconnected and even within the region containing the highest likelihood the chain still may need to reach a considerable length before locating it.

[11] We will examine the reasons behind this choice later in this section

$\mathcal{L}(\mathcal{D}, \theta)$. This is for observed data $\mathcal{D}$ and parameter point θ The posterior probability associated with this is proportional to, $\mathcal{L}(\mathcal{D}|\theta)\pi(\theta)$.

- **Step 2)** Next a candidate point θ', for the next 'step' the chain should take is selected. The candidate is part of the *proposal distribution* $q(\theta', \theta)$. In the case of `Magellan`, $q(\theta', \theta)$ is a Gaussian distribution, centred around the previous point θ, which has a standard deviation of a, this is typically known as the *step-size*. Generally this step-size will be small in comparison to the size of the parameter space overall, so the candidate point will be relatively close to the previous point. The likelihood corresponding to this new point is: $\mathcal{L}(\mathcal{D}, \theta')$.

- **Step 3)** Then the ratio of the posterior probabilities for each of the two points is calculated:

$$r = \frac{\mathcal{L}(\mathcal{D}|\theta')\pi(\theta')q(\theta', \theta)}{\mathcal{L}(\mathcal{D}|\theta)\pi(\theta)q(\theta', \theta)} \tag{3.11}$$

  $q(\theta', \theta)$, is symmetric in the Metropolis-Hastings algorithm, and in our case, as previously mentioned, is also Gaussian, thus it cancels out in the ratio and we are left with:

$$r = \frac{\mathcal{L}(\mathcal{D}|\theta')\pi(\theta')}{\mathcal{L}(\mathcal{D}|\theta)\pi(\theta)} \tag{3.12}$$

- **Step 4)** If, $r \geq 1$, then the proposed point is accepted; otherwise the acceptance of such a point has a probability of r. If a point is rejected at this point then the algorithm will return to **Step 2)** and repeat the process from there.

- **Step 5)** Once a candidate is accepted it will then be added to the chain and `Magellan` will repeat the process from **Step 2)**. This will continue to cycle until the chain reaches a predetermined length or until the user cancels it or runs out of allotted computing time, (on a cluster).

This algorithm shapes the evolution of a Markov chain within in a parameter space. Another interesting feature that we can make use of when we are analysing the points obtained via MCMC, provided they are run with identical[12] ranges, is that the density of points is proportional to the likelihood. This in turn means that they can be interpreted within a Bayesian Statistical framework, we can state that a region with a large density of points is more likely than one with a low density. If we could run infinitely long chains we would expect to completely map out the parameter space, thus if we run multiple chains up to a large length we expect to map the majority of it. What this means is that if we consider some plane of the space, x_1, x_2, and apply a

[12]The reason that we require identical runs here is because if any aspect of the likelihood or parameter space is altered then the density of points found in different regions will also be altered. For example, if we go from searching within a range of, $0 < x_1 < 10$, to, $9 < x_1 < 10$, clearly the likelihood of 9.5 has changed as there are far fewer values, x_1, can take

fresh constraint to this plane we can then switch to a few of the surviving points in another plane, y_1, y_2, and see how this has effected the points from this view; we can then infer constraints in these parameters as a result of the constraints on entirely different ones. Indeed, we make use of this feature in [43], which will be discussed in chapter 4.

Each individual chain is independent from any others and as such it is possible to run multiple chains in parallel [110]. This is advantageous for the user as it reduces the wall-clock time the scan takes overall. The dimensionality and size of the space the MCMC scan can be run over is somewhat arbitrary, though needless to say, higher dimensionality will require a greater time to fully scan. With the first iteration of `Magellan` we ran over a 6-dimensional parameter space for a 2HDM, (m_H, $m_{H^\pm}$, m_A, $\sin(\beta-\alpha)$, $\tan\beta$, Z_7), on Southampton university's `IRIDIS4`.
`IRIDIS4` is a cluster which uses dual 2.6 GHz 'Sandybridge' processors (Intel Xeon E5-2670) for standard jobs. The process of running 8 chains in parallel for 60 hours each was repeated numerous times with an average chain length of around 10000 being reached in that time. It is important to recall that not every point in a chain will be a valid point in the allowed parameter space, however. But this is something that has been accounted for, `Magellan` is set up to remove such points at the end of scans on command.

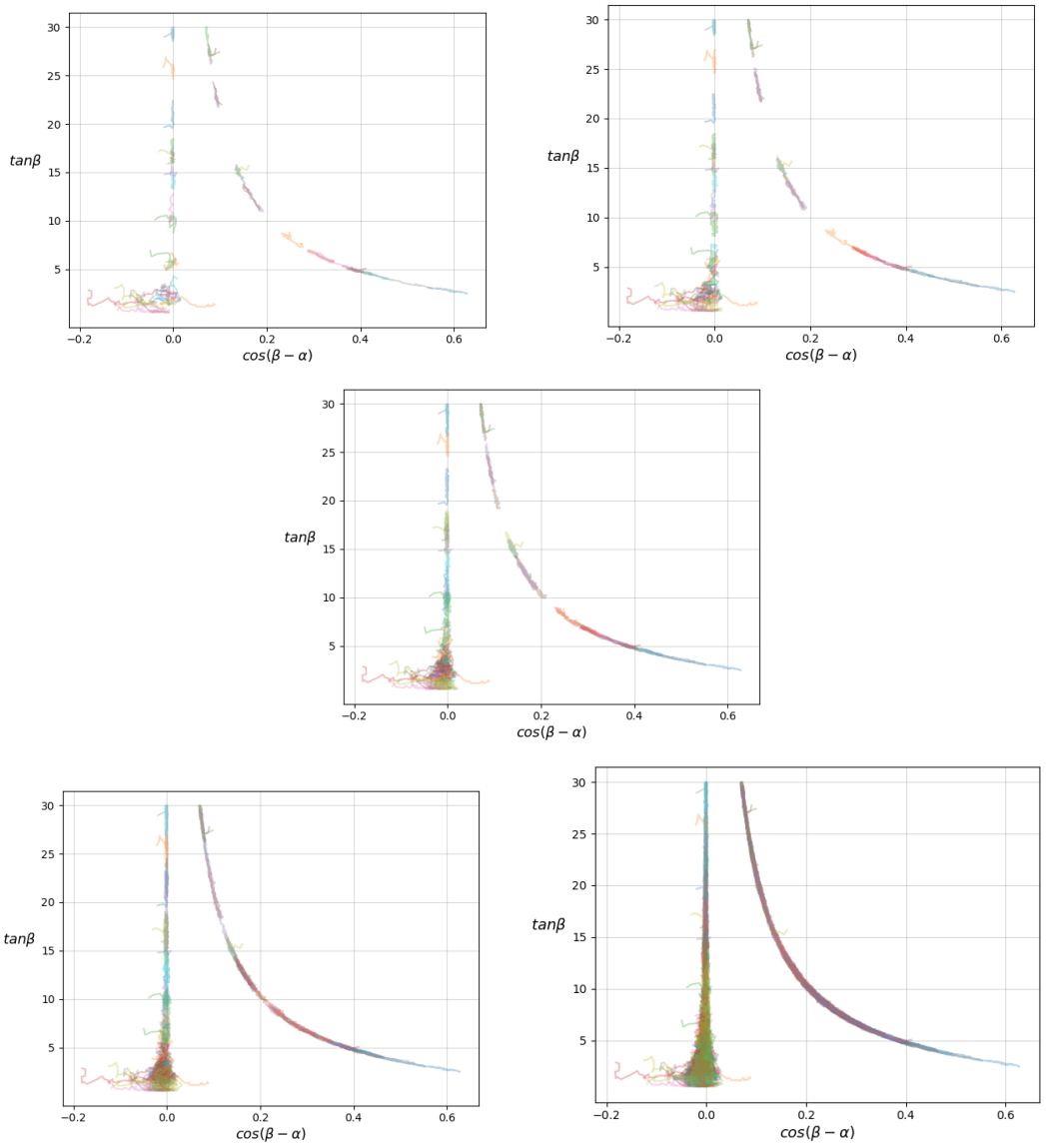
With the parameters defined as in Table 3.1, the most recent version of `Magellan` produced a data sample consisting of 1,092,985 points, before applying constraints from theory (stability, unitarity and perturbativity). This was performed on `IRIDIS5`, Southampton university's next generation of computing cluster following `IRIDIS4`; this uses dual 2.0 GHz 'Skylake' processors for standard jobs such as those mentioned here. This cluster gave us an average chain length of 17078 when we repeatedly ran 8 chains in parallel once again for 60 hours each. We consider this to be quite a successful improvement as the available parameter space was reduced a great deal (for the 2HDM type-II specifically, but also in general) in-between these tests; as such having a chain length increase of approx 70% shows that the computing system in use is very important.

Having specified exactly which type of MCMC we implement we can discuss the concept of convergence within it. Under ideal circumstances we expect all MCMCs to undergo three distinct phases (assuming there is no time/length limit imposed) [34]:

- Convergence to the 'typical set': Our chains start at random, independent points and begin taking steps stochastically to explore the parameter space. Due to the acceptance conditions on each potential step, over time they will move towards the typical set - or in the case of high dimensions split apart by constraints, the typical sets.

- The chains encounter and enter the typical set/s: the chains will make their first 'pass' through the typical set during which any bias left from their initial sampling is rapidly removed.
- The third phase consists of continued exploration of the typical set, the refinement of the Markov chain continues but it is at a much reduced rate compared to the previous stage.

As an example we run a scan on the Type-II with the parameter ranges defined in Table 3.1



3.3 Plots showing the progressive stages of the, $(\tan\beta,\ \cos(\beta-\alpha))$, plane as the MCMC chains populate the parameter space. Shown at chain lengths of 25, 50, 100, 200 and 1000. Each colour is a different chain and there are a total of 64 chains shown here.

Parameter	min	max	step-size
Z_7	-15.0	15.0	0.3
m_H [GeV]	150	2000.0	15.0
$m_{H^\pm}$ [GeV]	150	2000.0	15.0
m_A [GeV]	100	2000.0	20.0
$\cos(\beta-\alpha)$	-1.0	1.0	0.01
$\tan\beta$	0.5	30.0	0.3

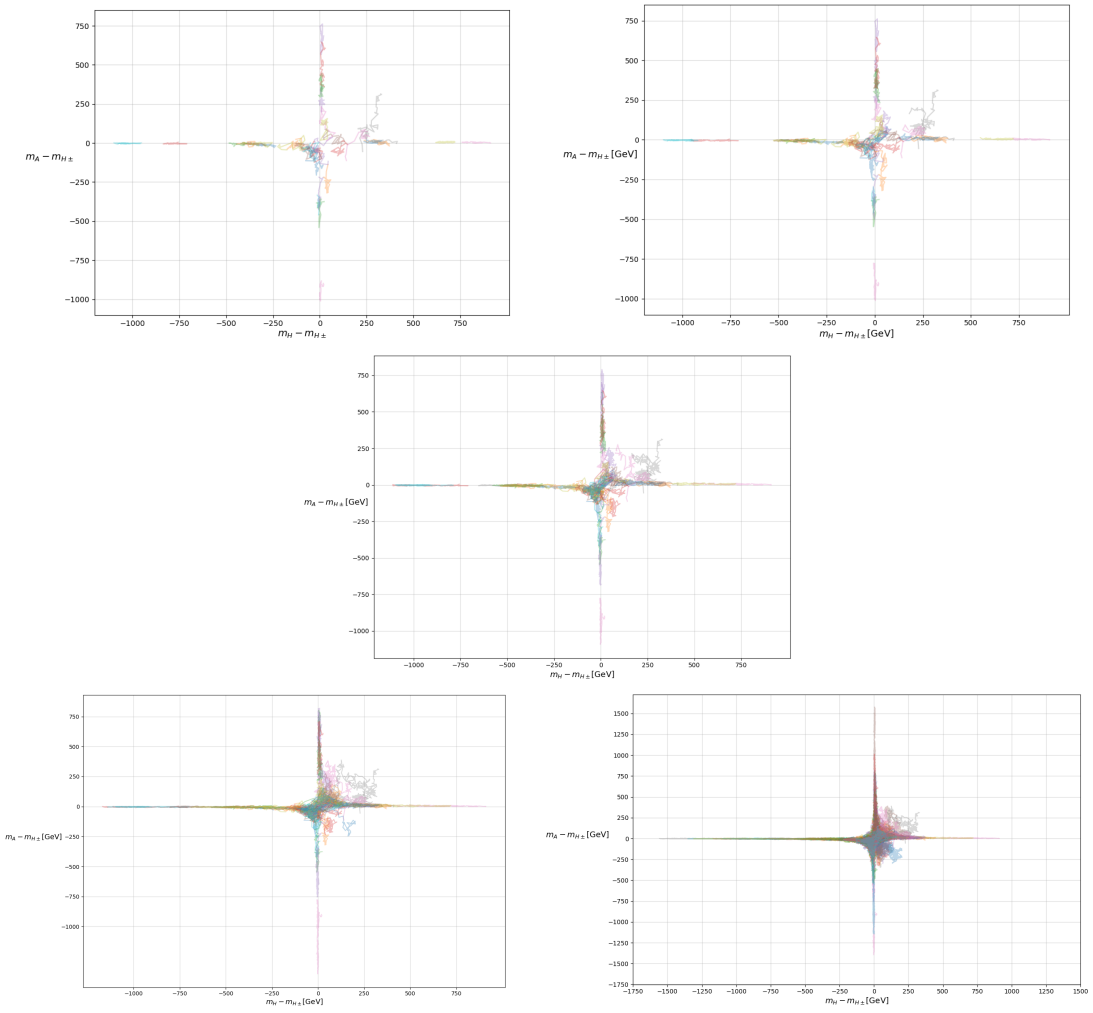
3.1 Range and step-size of the 6-dimensional 2HDM parameters used in the MCMC scan. (note that these are for demonstration purposes only and not necessarily representative of actual bounds for a 2HDM)

In Figs. Figures 3.3 and 3.4 there are plots of the progression of 64 independent chains through two different snapshots of the parameter space. In Fig. Figure 3.3 we have the, ($\tan\beta$,$\cos(\beta-\alpha)$), plane while Fig. Figure 3.4 depicts the, ($m_A - m_{H^\pm}$,$m_H - m_{H^\pm}$), plane. In the case of Figure 3.3 as the parameter space in the plane builds up we begin to see shapes very familiar to those who have studied the 2HDM type-II in recent years, showing that our MCMC is successfully mapping out the parameter space. Note that both the, ($\tan\beta$,$\cos(\beta-\alpha)$), plane and the, ($m_A - m_{H^\pm}$,$m_H - m_{H^\pm}$), plane contain the *same* 64 chains; this means we can select some subset of points in one and project it over to the other plane quite easily. `Magellan` by default has 76 parameters saved for every point in a chain, this includes the input variables (naturally how many of these there are and the specific variables used is dependent on the basis of choice) as well as all the others that can be calculated by `2HDMC`, `HiggsBounds` and `HiggsSignals`. The likelihood is also calculated, naturally, and stored as part of each point.

3.4.1.1 Retained Information

The scan performed by `Magellan` calculates a large amount of information on the phenomenology of the model in use; a large variety of this is retained as part of the parameter points, these variables include:

- Branching fractions:
  - light Higgs boson, "h":
    - BR($h \to bb$)
    - BR($h \to \tau\tau$)
    - BR($h \to \gamma\gamma$)
  - pseudo-scalar Higgs boson, "A":
    - BR($A \to tt$)
    - BR($A \to bb$)
    - BR($A \to gg$)



3.4 Plots showing the progressive stages of the, $(m_A - m_{H^\pm}, m_H - m_{H^\pm})$, plane as the MCMC chains populate the parameter space. Shown at chain lengths of 25, 50, 100, 200 and 1000. Each colour is a different chain and there are a total of 64 chains shown here.

    * BR($A \to \tau\tau$)
    * BR($A \to Zh$)
    * BR($A \to ZH$)
    * BR($A \to \gamma\gamma$)

- heavy Higgs boson, "H":
    * BR($H \to tt$)
    * BR($H \to bb$)
    * BR($H \to gg$)
    * BR($H \to \mu\mu$)
    * BR($H \to \tau\tau$)
    * BR($H \to Z\gamma$)
    * BR($H \to Zh$)

        * BR($H \to W + W-$)
        * BR($H \to ZZ$)
        * BR($H \to ZA$)
        * BR($H \to AA$)
        * BR($H \to \gamma\gamma$)
    - charged Higgs bosons, "$H^{\pm}$":
        * BR($H^{\pm} \to tb$)
        * BR($H^{\pm} \to W^{\pm}h$)
        * BR($H^{\pm} \to W^{\pm}H$)
        * BR($H^{\pm} \to W^{\pm}A$)

- Coupling (modifiers):
    - κ_{uu}
    - κ_{dd}

- Cross section:
  Dependent on user defined process. Calculated with `MadGraph5@NLO`.

- EWPOs: S, T , U , V , W , X, $\Delta\rho$.

- `HiggsBounds` information:
    - Sensitive channel
    - $\sigma_{pred}/\sigma_{lim}$

- Model input parameters: e.g. Z_7 , mH , $mH^{\pm}$, mA , $\cos(\beta-\alpha)$, $\tan\beta$.

- Physical widths: Γ_h , Γ_H , $\Gamma_{H^{\pm}}$, Γ_A.

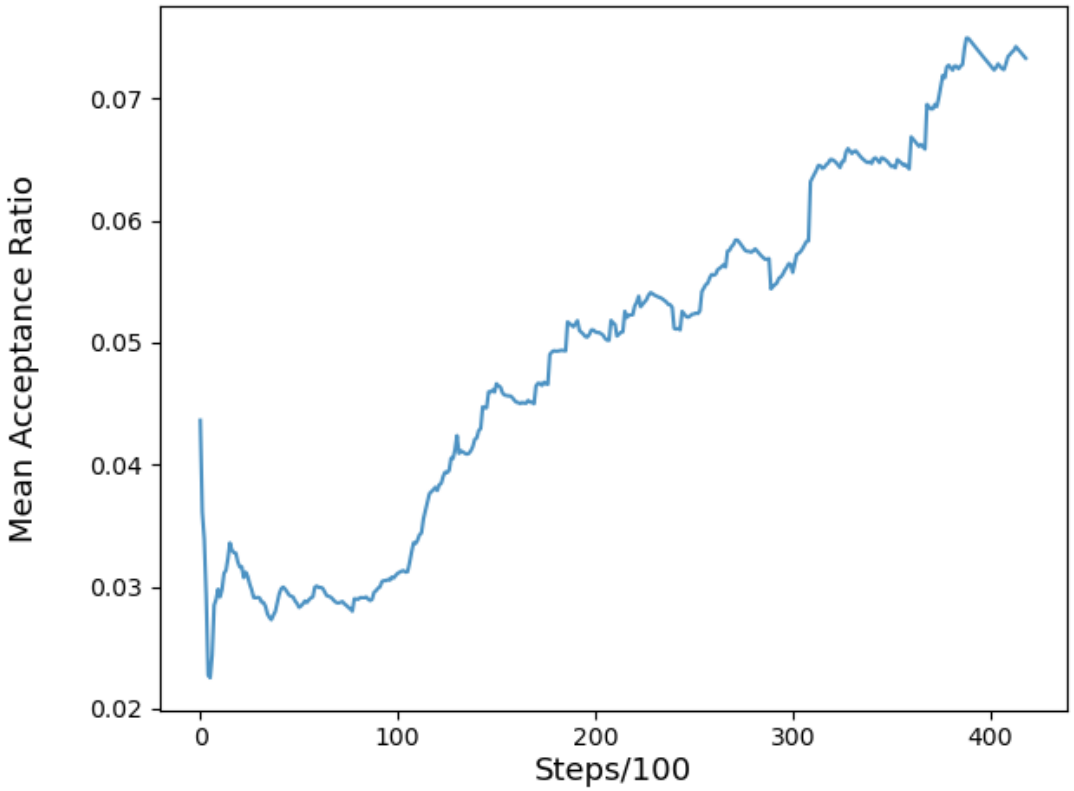
- Potential shape parameters: m_{2_12} , λ_1 , λ_2 , λ_3 , λ_4 , λ_5 , λ_6 , λ_7.

- Quartic couplings: Z_4 , Z_5 .

- Theoretical constraints: stability, unitarity, perturbativity.

- χ^2 values: $\chi^2_{\text{HiggsSignals}}$, χ^2_{ST} , χ^2_{Tot} .

The choice in how many points to disregard for the burn-in period is quite complicated in this instance as we have information within the point, from 2HDMC, that is not used in the likelihood. This is partially an active choice to improve the 'mixing' of the chains (i.e. how well they move between different regions of the parameter space), especially as the inclusion of the constraints from information would result in a parameter space that is further disconnected; and partially as it is very easy to select only points with specific `2HDMC` values after the chains have run. It is the theoretical



3.5 Plot showing that the mean acceptance ratio increases with the number of steps taken, showing how the MCMC moves into more likely regions.

constraints of perturbativity, unitarity and positivity of the potential in particular that we refer to here. `2HDMC` outputs a binary response for each point to indicate whether or not it has passed these constraints, we keep this as part of the data within a point and simply select only those points with a passing value for the constraints while discarding the rest. The main drawback of this is that it may slow the MCMC's progression towards the 'maximum likelihood', or at least towards a local maximum likelihood.

Whether or not this is of importance for a particular user depends on their goals with `Magellan`. If one simply wishes to map out the available parameter space, *regardless* of how likely a given point within may be, then there is no reason to alter the default behaviour. However, if a user wishes to hone in on regions of highest likelihood then the recommendation is to incorporate one or more of the passing/failing constraints into the MCMC algorithm. The simplest way to do this is to make a copy of the relevant ParameterPointProcessor (see Appendix A) and edit it to add a statement that causes immediate exit from that script (with `exit( -1 );`) in an if statement for when the value of the desired condition is equal to 0.[13][14] We also advocate for testing each of these constraints *individually* to decide which to implement, it is quite possible

[13] this file is written in C++ for reference

[14] Naturally this must be done *after* the values for this are calculated. The ParameterPointProcessors are well commented and searching "Theory" should bring you to where the calculation happens.

that they have different impacts from one another and that adding one in may have a larger impact than expected. We also suggest the user try each one and plot traces of the likelihoods for the entirety of their chains[15] in order to check whether or not convergence is happening faster/ which constraint has the most effect.

We have now discussed why the constraints were not used as part of the MCMC algorithm, but not why we choose to discard 200 points at the start of each chain. Figure 3.6 shows two plots of the, $\chi^2{}_{Tot}$, values for our demonstrative chains. Initially the values fluctuate quite wildly before settling into their 'stationary' states[16]. This burn-in period seems to last no more than the first 100 steps, as we can see looking at both plots. Zooming in on the first 200 steps in the RHS plot in order to see how some of the individual chains are changing during this period and conclude that all of them have converged by the end of these steps making it a reasonable number of points to discard.

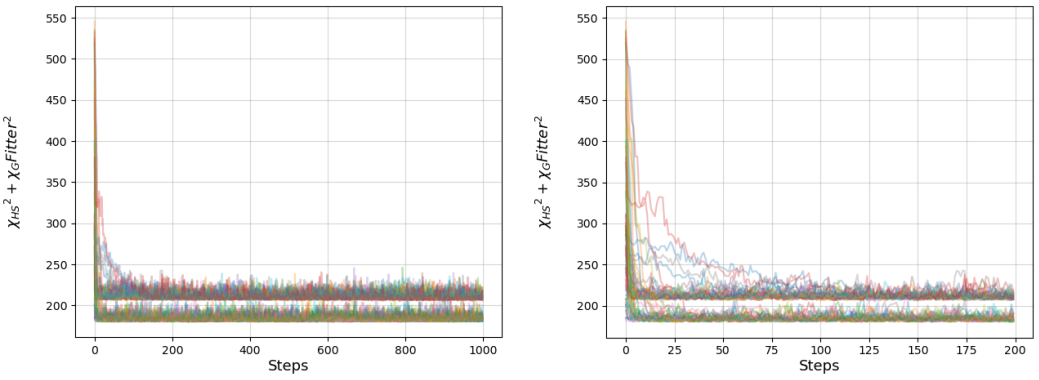
We note that the user may not see *two* distinct bands, and this is not a problem. There happened to be two highly constrained solutions in this case which meant the chains became very locked into one or the other, but if we compare this to an earlier trace plot, made when the constraints were looser:

The key feature that shoes convergence is the flattening of the chains, whether this occurs in one mass, two, three, etc., it does not matter[17]

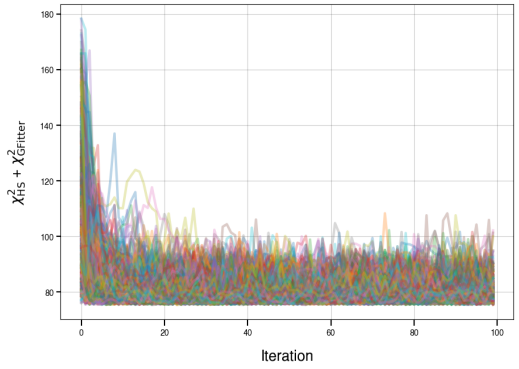
[15] These should be like the plots in Figure 3.6, though with a single line for each constraint being tested.

[16] So-called because on average they are not changing; though there is, of course, small fluctuation still occurring.

[17] Though if the number of bands is equal to the number of chains and they exhibit flattening without touch one another at all this could indicate that the chains are all getting stuck in a local likelihood maxima and that the parameter space is highly fractured. In this case it may be helpful to attempt to temper the MCMC in question.



3.6 Plots showing the, $\chi^2{}_{Tot}$ value, where, $\chi^2{}_{Tot} = \chi_{HS}{}^2 + \chi_{Gfitter}{}^2$, value of the example chains for each step taken. The LHS shows a larger number of steps, where we can see that the, $\chi^2{}_{Tot}$, values converge into two distinct bands. These two bands correspond to the "alignment" and "wrong-sign" solutions for the 2HDM we have used. The RHS shows only the first 200 steps, here we can see that the, $\chi^2{}_{Tot}$, values are beginning to approximate thermal equilibrium by 100 steps. Each colour is a different chain and there are a total of 64 chains shown here.



3.7 A trace plot created from data using an older version of `Magellan`, with looser constraints. Unlike in Figure 3.6 we do not see two distinct bands, instead there is just one mass of chains together. We can also see that it is following the same behaviour in terms of convergence, however. The fluctuations of the chains start much larger but quickly reduce as they converge, just like the plots in Figure 3.6.

As mentioned previously, a key feature of the MCMC scanning method is that the results can be interpreted in the Bayesian statistical framework, meaning that the density of the points in the parameter space is proportional to the posterior probability of the model describing the data. It must be stressed that this is **not** the case if one combines the results of two scans on the same model that have differing bounds on one or more parameters. For example, if we run a scan with a parameter x set to 0 to 10 and then a second scan with everything left the same but x instead set to -10 to 10 then we are not actually dealing with the same 'true' probability for each point as one of these only contains a subset of the possible points contained in the other, artificially boosting the probability in the reduced space. Whether or not this is an issue depends on the reason for performing the scan, as will be shown later we have used `Magellan` for different purposes. `Magellan` was used to map out a parameter space and make statements about the likelihood of different regions using the Bayesian framework interpretation of the point density; but we also used it simply to find valid points remaining to a heavily constrained parameter space (in fact they are both for the 2HDM, but multiple years passed in between and the older parameter space has been constrained much more severely).

3.4.1.2 Limitations

- As we cannot run our MCMCs forever we will never reach a state where our samples are completely uncorrelated. This does reduce the *effective sample size*. It also increases uncertainty on the posterior quantities we seek (i.e. measurements such as variance and mean)

- It is possible for it to appear as though a Markov chain has converged to its equilibrium distribution, which is what we would like to happen, when in fact it has not. This is called *pseudo-convergence*, it happens different areas of the state

space are not well connected; meaning that the Markov chain dynamics lead to the chain taking many iterations to move from one part of the space to another. If this is to such an extent that the chain takes longer to move between two regions than the entire length of simulation then a trace plot can appear to show convergence.

- Users can make the mistake of using a large number of chains, in order to try and place a chain in every region of the state space, possibly to avoid the previous problem of pseudo-convergence. However, there is an unspoken assumption that any region where pseudo-convergence would occur has indeed had a chain placed in it. Naturally this is not something that we can be confident we have achieved, and thus we would have erroneous results that might still tell us we have converged when in fact we have not. Short runs, no matter how many there are, are not good at detecting pseudo-convergence. A second problem that can occur is that with short runs is that we may miss other bugs, even entire regions that our chains have simply been unable to reach.

- In an ideal world we would run a large number of chains *and* run them for an extended period of time. As we do not live in this idealised world we instead recommend running at least one chain for 1+ days and examining the resulting trace plot. This can highlight some of the potential issues we have discussed and allow a user to circumvent them.

We do stress that `Magellan` is slightly different from a typical MCMC as we have a method for checking whether or not each point is valid in `2HDMC`. The issues we have highlighted will present the users of more typical MCMCs with points that are **not** truly in the state space if they are not careful in their choices; this is not true for us as any invalid point is marked as such. It will still affect the time taken to move between regions however, so we do recommend multiple chains.

For further information on how MCMCs work we recommend the textbook [82] by Prof. C J Geyer. For information on installation and running of `Magellan` please refer to Appendix A

Chapter 4

LHC Data Interpretation Within The 2HDM Type-II Via Magellan

After having discussed the features of `Magellan` we now move to the first investigation performing with it. In this investigation we aimed to determine whether it could improve upon the standard procedure for setting bounds on the 2HDM parameter space. Thus far no experimental evidence has surfaced *for* the existence of the 2HDM, thus we aim to set bounds upon it specifically with our new toolbox. As mentioned in section 3.1 we chose the 2HDM for our test model since it is both physically interesting and, crucially for our purposes, characterised by a multi-dimensional parameter space that is far from trivial. Such a space is ideal for displaying the effectiveness of `Magellan`. Recall that the new aspect of the method within `Magellan` is that the parameter space of a BSM theory can be projected onto any bi-dimensional plane with the underlying features of those points still being retained, allowing one to investigate the associations between the properties of multiple lower dimensional sub-spaces of the full parameter space.

As the 2HDM parameter space is six-dimensional, (after we apply soft Z2 symmetry breaking and require, $\lambda_6 = \lambda_7 = 0$), it is not possible to view the entirety of it at once and observe bounds in that way.

The process we chose for illustrative purposes is that of the associated production, $pp \to Zh \to 4f$, within the Type-II 2HDM [2]

The additional four Higgs states of a generic 2HDM [92; 36] provide a range of potential new observables through which experimentalists could test LHC data for the presence of the 2HDM; or at least seek initial evidence of the existence of through the scalar sector. Thus it is worthwhile performing a detailed investigation into the ambit of the LHC in discovering these new Higgs bosons should they exist.

Following on from the discussion of the 2HDM in section 2.3, there are several alternative bases in which the 2HDM can be described: the *general parametrisation*, the *Higgs basis*, where one of the doublets has zero VEV, and the *physical basis*, where the physical masses of the scalars are used. As of the discovery of the 125 GeV Higgs

boson, herein the, h, state, it is common to parametrise this theory using the *hybrid basis* [94], where the parameters allow us direct control on both the *CP*-even and *CP*-odd Higgs boson masses as well as the hVV couplings, ($V = W^{\pm}, Z$), the, $Aq\bar{q}$, vertices and the Higgs quartic couplings. The parameters in this basis are:

$$\underbrace{\cos(\beta-\alpha)}_{\substack{\text{determines the} \\ g_{hVV}, \text{ \& } g_{HVV}, \text{ couplings}}} \quad , \quad \underbrace{m_h, \quad m_H}_{CP\text{-even Higgs boson masses}} \quad , \quad \underbrace{\tan\beta}_{\text{ratio of the vevs}} \quad , \quad \underbrace{Z_4, \quad Z_5, \quad Z_7,}_{\substack{\text{Higgs self-coupling} \\ \text{parameters}}} \tag{4.1}$$

with, $m_H \geq m_h$, $0 \leq \beta \leq \pi/2$, and, $0 \leq |\sin(\beta-\alpha)| \leq 1$. We can replace two of these quartic couplings with the remaining (pseudo)scalar by examining those masses in terms of the quartic scalar couplings in the Higgs basis:

$$m_A^2 = m_H^2 \sin^2(\beta-\alpha) + m_h^2 \cos^2(\beta-\alpha) - Z_5 v_1^2, \tag{4.2}$$

$$m_{H^{\pm}}^2 = m_A^2 - \frac{1}{2}(Z_4 - Z_5)v^2 \tag{4.3}$$

Hence we can swap the self-couplings, Z_4, and, Z_5, with the scalar masses given above meaning that the 7 free parameters can be recast into 4 physical masses and 3 parameters related to the couplings of the 2HDM scalars to gauge bosons, fermions and scalars themselves, respectively:

$$\cos(\beta-\alpha),\ m_h,\ m_H,\ m_A,\ m_{H^{\pm}},\ \tan\beta,\ Z_7\ . \tag{4.4}$$

Looking at this list we can quickly reduce it by one further parameter, m_h [1], as it has been measured to a high degree of accuracy at the LHC. This brings us to 6 free parameters. We are also aware that, Z_7, enters only the triple and quartic scalar interactions.

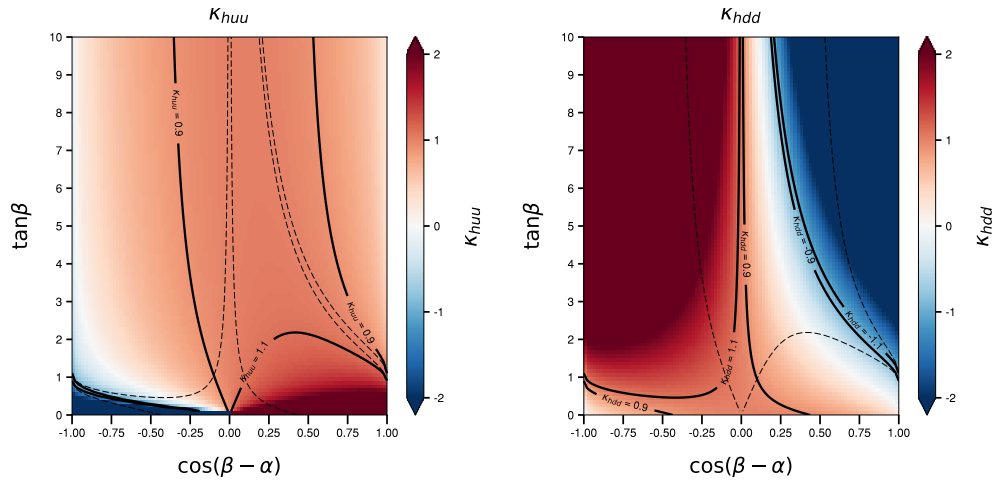
The couplings of the neutral Higgs bosons to fermions, normalised to the corresponding SM value (m_f/v, henceforth denoted by κ_{hqq}, for the case of the SM-like Higgs state coupling to a quark q, where, $q = d, u$), can be found in Table 2.1.

We will be concentrating on the 2HDM Type-II. We have two different limiting scenarios which give rise to one distinct region, ($\cos(\beta-\alpha), \tan\beta$), parameter plane each [77]. A way of to understand these scenarios is by examining the behaviour of, κ_{hqq}, as a function of the angles, α, and, β. Taking the limits, $\beta-\alpha \to \frac{\pi}{2}$, (given in Equation 4.5) and, $\beta+\alpha \to \frac{\pi}{2}$, (given in Equation 4.6), the (recalling Table 2.1) the couplings become:

$$\begin{aligned} \kappa_{hdd} = -\frac{\sin\alpha}{\cos\beta} &= \sin(\beta-\alpha) - \cos(\beta-\alpha)\tan\beta \xrightarrow[\beta-\alpha=\frac{\pi}{2}]{} 1 \text{ (middle-region)}, \\ &= -\sin(\beta+\alpha) + \cos(\beta+\alpha)\tan\beta \xrightarrow[\beta+\alpha=\frac{\pi}{2}]{} -1 \text{ (right-arm)}, \end{aligned} \tag{4.5}$$

[1] It is also possible to identify the Higgs discovered at the LHC with H, as both it and h are CP-even. This is known as "inverted Hierarchy", but it is not what we study here.

$$\begin{aligned}\kappa_{huu} = \frac{\cos\alpha}{\sin\beta} &= \sin(\beta-\alpha) + \cos(\beta-\alpha)\cot\beta \xrightarrow[\beta-\alpha=\frac{\pi}{2}]{} 1 \text{ (middle-region)},\\ &= \sin(\beta+\alpha) + \cos(\beta+\alpha)\cot\beta \xrightarrow[\beta+\alpha=\frac{\pi}{2}]{} 1 \text{ (right-arm)},\end{aligned} \tag{4.6}$$



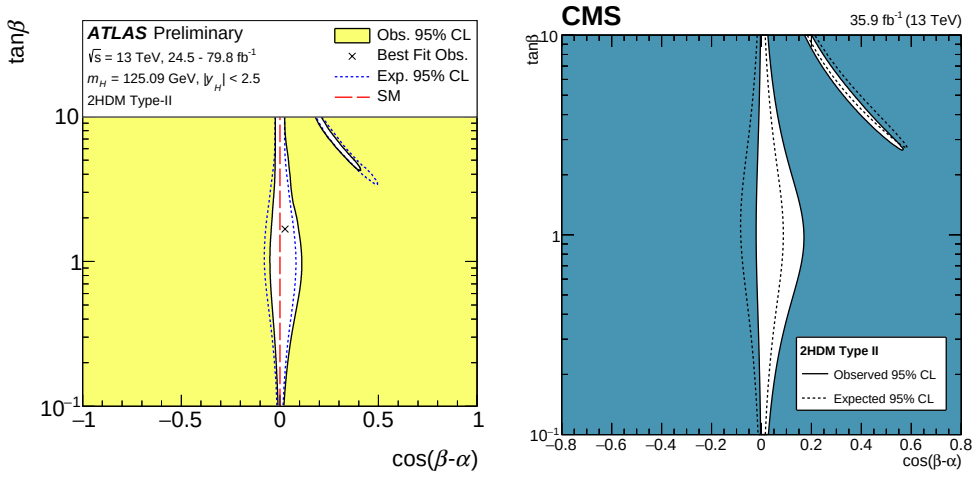
4.1 Couplings for the light *CP*-even Higgs to the up type (left) and down type (right) quarks, normalised to their corresponding SM value, shown in the, $(\cos(\beta-\alpha), \tan\beta)$, plane.

We can see the dependence of the couplings, κ_{hdd}, and, κ_{huu}, on, $\cos(\beta-\alpha)$, and, $\tan\beta$, in Figure 4.1. The, $\beta-\alpha \to \frac{\pi}{2}$, case corresponds to the SM limit of the theory, often known as the 'middle-region' due to its position and shape. We can see this within the RHS plot of Figure 4.1, it is the contour region where, $0.9 \leq \kappa_{hdd} \leq 1.1$. This is the same as saying the region assuming a, κ_{hdd}, range within 10% of the SM couplings. This is also called the 'alignment' region/solution, as the couplings for the up and down type quarks have the same sign as the SM.

In the LHS plot there is a narrow blue 'arm' corresponding to the, $\beta+\alpha \to \frac{\pi}{2}$, here we get an opposite sign, (relative to the SM), for the coupling between the down type quarks and the SM-like Higgs boson, h. This is known as the wrong-sign Yukawa coupling region/scenario. This region where the coupling is negative, and similarly to before has an assumed a range within 10% of the SM coupling value when we take the *magnitude*: $-1.1 \leq \kappa_{hdd} \leq -0.9$.

These two regions, alignment and wrong-sign, both fall within the O(10%) discrepancy for the couplings between SM-like Higgs boson to the up type quarks, κ_{huu}. The LHS plot of Figure 4.1 shows this with the contours for, κ_{hdd}, plotted in dashed lines alongside the filled lines of the, κ_{huu}, contours.

At the time of researching for this project, the most up-to-date 125 GeV Higgs boson combined signal strength analyses were ATLAS [15] and CMS [134]. The plots in Figure 4.2 are taken from these and show that the hypotheses of, $\kappa_{hdd} = 1$, and, $\kappa_{hdd} = -1$, are both still allowed. Considering the scenario from a theory perspective,



4.2 Allowed regions of $(\cos(\beta-\alpha), \tan\beta)$ parameters in 2HDM Type-II, obtained from the compatibility with the observed couplings of the 125 GeV boson, when identified as the light Higgs boson, h of the model. The plot show the most up-to-date available results from ATLAS [15] and CMS [134], seen on the left and right plot respectively. There are two distinctive regions visible in these plots, these are known as the middle and right arm regions and. The right arm region corresponds to the area where we find the wrong-sign solution. Note that it is constrained to a very narrow area in general and to a very small range of $\cos(\beta-\alpha)$ values; this is a result of flavour physics measurements heavily constraining 2HDM parameters.

a study [20] based on Renormalisation Group Equations (RGEs) showed that, in order for a model to be valid up to 'higher energies' (beyond 1TeV), the allowed parameter space shrinks to the positive sign of, $\kappa_{huu}/\kappa_{hdd}$, i.e. the alignment region. Below this energy scale, the alignment and wrong-sign scenario are still valid.

Now considering a more phenomenological point of view, numerous analyses have been performed to constrain both domains. The decay channels of the two additional neutral Higgs bosons, A, and, H, have been shown to be important in the wrong-sign limit of the model in Ref. [78]. In the case of the A the lack of observation through direct searches provides the first constraint via upper limits on the couplings between the $A, W^{\pm}$ and Z bosons. This then applies constraints on the value of $cos(\beta-\alpha)$. Then the lack of expected excesses in the SM-like Higgs in regions where it could have decayed from a non-SM Higgs Boson, should there be one to decay from provides further constraints, pushing up the minimum mass or pushing down the maximum coupling strengths for these Higgs particles.

Further constraints are placed on $cos(\beta-\alpha)$ by measurements of SM-like Higgs decays to either τ or μ pairs at the LHC. We find that:

$$|cos(\beta-\alpha)|\eta_l < \mathcal{O}(1) \tag{4.7}$$

where η_l is the Yukawa coupling between the SM-like Higgs and the leptons. It is the decay to the τ pairs that has the largest impact, constraining the value of $cos(\beta-\alpha)$

to low values.

Another measurement that constrains the parameter-space is that of the muon anomalous magnetic moment, a_μ. Within the SM this parameter has corrections due to loop effects from electroweak particles (it recieves them from hadronic loops also but the EW is of more relevance here ·). In the 2HDM we have more bosons in the EW sector and thus more particles to contribute loop corrections. If a model has non-zero, non-diagonal elements in the Yukawa matrices then we have Feynman diagrams that will contribute to Lepton Flavour Violating (LFV) decays, for example $\mu \to e\gamma$. A lack of observation of such processes places strong constraints on the flavour-changing Yukawa couplings, and on the masses within the loops affecting measurements of g_μ. The Yukawa couplings to the fermions are all trigometic functions involving either the weak mixing angle, α and/or the angle β from the ratio of the doublet vevs, as seen in Table 2.1, thus non-observation of these processes places constraints on allowed values of $cos(\beta - \alpha)$ and $tan\beta$. B-physics plays a role in constraining the up-type Yukawa coupling, in particular $b \to s\gamma$ and $B_s \to \mu^+\mu^-$ cross-section measurements are the most constraining.

4.1 Magellan: Global Scan For Bounds Extraction And Data Interpretation

`Magellan`, is designed with two different uses in mind. Firstly, it allows easy imports of any new experimental results making it possible to interpret these within a given model and derive bounds on the constituent parameter space. Secondly, `Magellan` is able to predict the regions of a parameter space that are be accessible in a given search with the actual luminosity and hence display the characteristics of new particles to be searched for (e.g., mass, width, decay rates, etc.) allowing for an improvement to the data analysis.

To illustrate this model exploration approach within `Magellan` we created an interactive website [144]. For this we took the 2HDM Type-II as an example. Within our framework, the limits on this model were derived in order to show how effective this method is.

When considering processes that the already discovered h enters it is logical to concentrate on the experimental observables. But when performing searches for BSM Higgs states, neutral or charged, simply choosing a specific theoretical model, can immediately force constraints on the parameter space. Therefore these should to be included too. In order to do this effectively we have used `HiggsBounds` 4 [24] (recall that this was the earlier version of `Magellan`, hence it uses `HiggsBounds` 4 and not 5); as mentioned in chapter 3 this software tests the provided model against the exclusion limits from the Higgs boson searches at LEP, Tevatron and LHC. We must also account for the inclusive weak radiative, B-meson Branching Ratio (BR) which

Parameter	min	max	step-size
Z_7	-10.0	10.0	0.2
m_H [GeV]	150	1000.0	20.0
$m_{H^\pm}$ [GeV]	500	1000.0	20.0
m_A [GeV]	100	1000.0	20.0
$\cos(\beta-\alpha)$	-1.0	1.0	0.03
$\tan\beta$	0.5	30.0	0.5

4.1 Range and step-size of the 6-dimensional 2HDM parameters used in the MCMC scan.

α	α_s	$\alpha_{\text{EM}} \equiv \alpha(Q^2=0)$	m_t [GeV]	m_h [GeV]
1/127.934	0.119	1/137.035997	172.5	125.09

4.2 Physical parameters kept fixed in our scans.

proceeds through the quark-level transition of, $b \to s\gamma$. A study we consulted [118], which used results from the Belle Collaboration, placed a, 95% C.L., lower bound on the charged Higgs mass of, $m_{H^\pm} > 580$ GeV, (C.L. standing for "confidence limits"). As such, we select points with charged mass above this value.

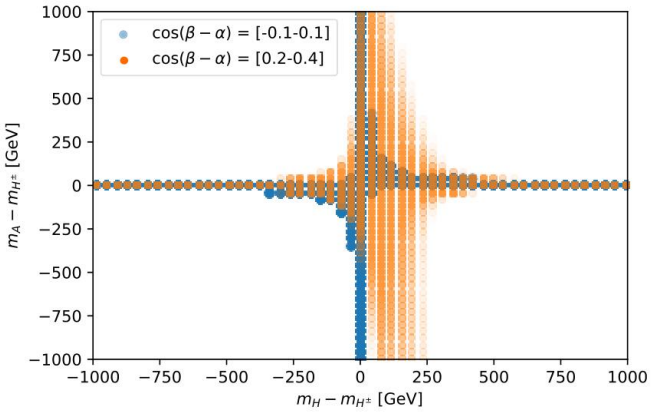
The MCMC scan is performed over the 6-dimensional parameter space (m_H, $m_{H^\pm}$, m_A, $\cos(\beta-\alpha)$, $\tan\beta$, Z_7). The ranges and step-size of each parameter can be found in Table 4.1. Other physical quantities are set to a constant value, these are listed in Table 4.2. To give the reader some idea of how computationally expensive the scan was, it is worth specifying what options were chosen for the scan: 400 independent chains were submitted, each for 20 hours on Dual 2.6 GHz Intel Xeon 8-core processor machines. With the given time limit, the setup yields an average chain length of O(10000) steps for each chain. A post-processing step is also performed where the production cross sections of the extra neutral scalars (A, H) are calculated using SusHi [99], along with their branching fractions into the various decay products using 2HDMC [73].

4.2 Bounds On The 2HDM Type-II

With our toolbox we can extract bounds on the six independent free parameters of the 2HDM Type-II while taking into account important features like Higgs coupling strengths, EWPOs as well as theoretical constraints.

4.2.1 Experimental constraints

The values of the EWPOs, S, T, and, U, in the 2HDM have been derived in [89; 88] and implemented in `2HDMC`. `2HDMC` is dependent on the squared masses of the neutral



4.3 Allowed parameter points with, $-0.04 \leq T \leq 0.24$, in the, $(m_H - m_{H^\pm},\ m_A - m_{H^\pm})$, plane for, $|\cos(\beta-\alpha)| < 0.1$, and, $0.2 < \cos(\beta-\alpha) < 0.4$.

Higgs bosons via the, F, function [146], which is commonly part of loop calculations:

$$F(x,y) = \frac{x+y}{2} - \frac{xy}{x-y}\ln\frac{x}{y}, \tag{4.8}$$

$F(x,y)$, is a non-negative function, it is zero for, $x = y$, and it increases as the difference between, x, and, y, does. In order to simplify notation, we use, $F(A,B)$, denoting, $F(m_A^2, m_B^2)$. The, T, parameter in the 2HDM can be expressed as:

$$\begin{aligned} T = c\Bigg\{ & \cos^2(\beta-\alpha)\Big[F(H^\pm,h) - F(A,h) - F(H^\pm,H) + F(H,A) \\ & + 3\left[F(Z,H) - F(W,H)\right] - 3\left[F(Z,h) - F(W,h)\right]\Big] \\ & + F(H^\pm,H) - F(H,A) + F(H^\pm,A)\Bigg\} \end{aligned} \tag{4.9}$$

where, c, is

$$c = \frac{1}{\alpha_{\text{EM}}}\frac{g^2}{64\pi^2 m_W^2} \tag{4.10}$$

In the alignment limit, where, $\cos(\beta-\alpha) \approx 0$, T, simplifies to:

$$T = c\Big[F(H^\pm,H) - F(H,A) + F(H^\pm,A)\Big] \tag{4.11}$$

This shows us that if there is a mass degeneracy between, A, or, H, and, $H^\pm$, then we have a vanishing, T, parameter. Typically, in the literature, either, m_H, or, m_A, are set as equal to, $m_{H^\pm}$, in order to satisfy the EWPO constraints for the 2HDM. This is not as important for the wrong-sign solution, however, at least where, $\cos(\beta-\alpha) > 0$.

It is useful to discuss briefly what the T parameter *is* and in what way it is relevant here. As discussed in section 3.2 the T parameter is one of the oblique parameters used

as measures of possible deviation from the SM in experimental measurements. The Feynman diagrams which contribute to the T parameter all contain either Z or $W^{\pm}$ vacuum polarization loops with novel Higgs particles in them. These introduce corrections to the propagators of those particles that are proportional to the masses of the Higgs particles appearing in the loops. As such, T can be thought of as a measure of the violation of the custodial $SU(2)_c$ symmetry mentioned in chapter 2, which leads to the $Z, W^{\pm}$ triplet splitting and the different masses we see for the neutral and charged bosons.

Taking only the leading bound on, T, into account, we can relax the mass degeneracy to a reasonably large extent. This can be seen in Figure 4.3, where the allowed mass differences between the Higgs states meeting the requirement, $-0.04 \leq T \leq 0.24$. The choice for the, T, value comes from the GFitter analysis of Ref. [18], where, U, is set to zero for the extraction of 95% C.L. bounds. The orange points in Figure 4.3, are for large, $\cos(\alpha - \beta)$, values in the wrong-sign domain, m_H, and, m_A, can differ from, $m^{\pm}$, by around 250 GeV (if not more). However large non-perturbative contributions arise when there are very large differences between the scalar masses and as such extreme cases are disfavoured.

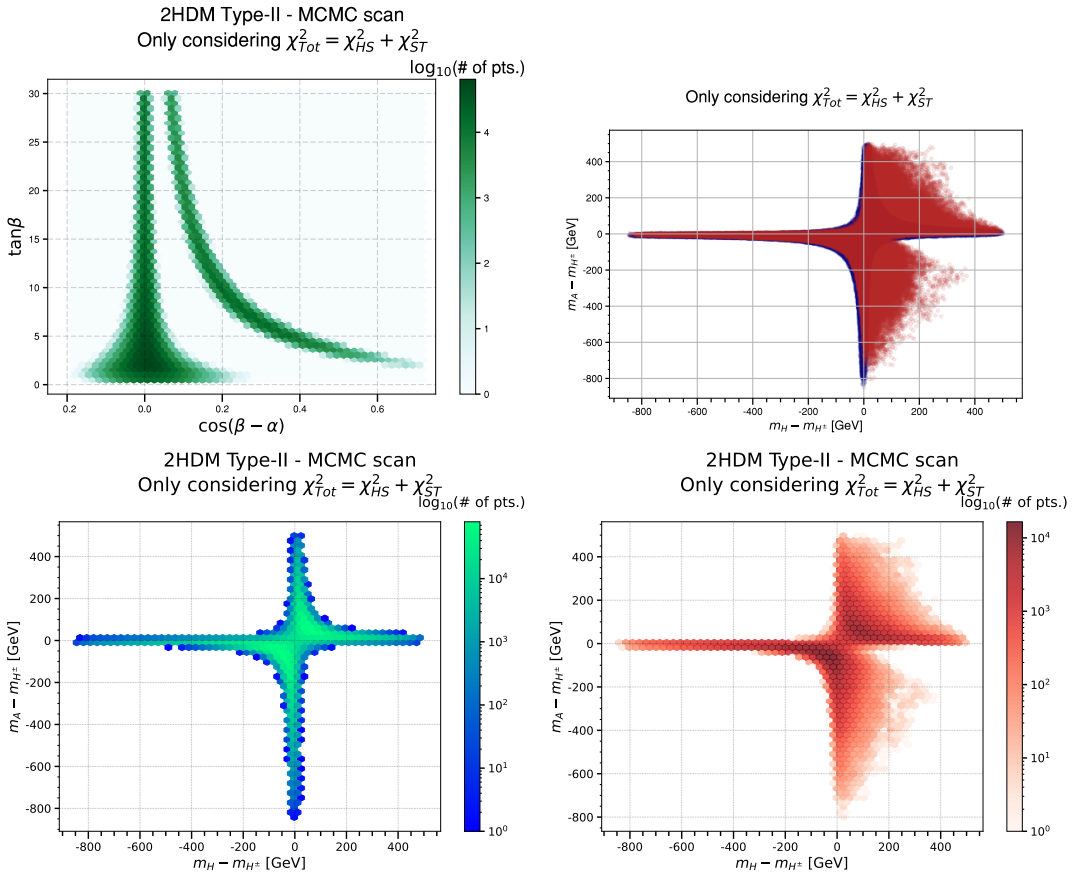
The result of including these constraints, for the MCMC scan performed by `Magellan`, from both the EWPOs and the SM-like Higgs boson measurements is visualised in Figure 4.4. There we have plotted the allowed points in two parameter planes: that of, $(\cos(\beta - \alpha), \tan\beta)$, (top-left) along with, $(m_H - m_{H^{\pm}}, m_A - m_{H^{\pm}})$, (top-right, and both bottom plots).

In the top LHS plot we have plotted the density of points, as indicated by the bar indicating how the colour of the point relates to the number of points found in a spot. For the top RHS we plot both the alignment and wrong-sign scenario points in blue and red respectively. The lower two plots show the points from the top RHS but split into the alignment (LHS) and wrong-sign (RHS) solutions separately. In addition they have been binned and are presented like the first plot with density of points rather than points. As before, they each have a colour bar indicating where the greatest density of points can be found. The top left plot in Figure 4.4 appears to have good agreement with the experimental fits shown in Figure 4.2, and so it passes the goodness-of-fit test for the adopted `HiggsSignals` link.

4.2.2 Theoretical constraints

Having covered the limits on our parameter space as a result of direct and indirect experimental searches, we now examine the effect of theoretical constraints. There are three main conditions which can be summarised as follows;

- Unitarity: the S matrix must be unitarity, this means that there is an upper bound on the eigenvalues L_i of the scattering matrix for all Goldstone and Higgs



4.4 Distribution (and concentration) of the parameter space points on the, $(\cos(\beta-\alpha), \tan\beta)$, plane (top-left) and, $(m_H - m_{H^\pm}, m_A - m_{H^\pm})$, plane (top-right) from the MCMC scan in the 2HDM Type-II. In the top RHS plot displaying the mass differences, the alignment region is represented in blue while the wrong-sign one is superimposed in red. To ensure all interesting data is displayed we also plot these separately: the blue-green points isolate the alignment limit scenario (bottom-left) while the red-orange ones isolate the wrong-sign configuration (bottom-right). The colour gauges measure the number of scan points plotted.

boson 2-to-2 channels [84; 103]. This limit is:

$$|L_i| \leq 16\pi, \tag{4.12}$$

- Perturbativity: in order to justify the perturbative nature of calculations the quartic Higgs couplings should be small

$$|\lambda_{H_i H_j H_k H_l}| \leq 8\pi, \tag{4.13}$$

- Stability of the potential: we require that the quartic Higgs potential terms are bounded from below, which implies that [68]

$$\lambda_1 > 0, \quad \lambda_2 > 0, \quad \lambda_3 + \sqrt{\lambda_1 \lambda_2} > 0, \quad \lambda_3 + \lambda_4 - |\lambda_5| + \sqrt{\lambda_1 \lambda_2} > 0, \tag{4.14}$$

General potential coefficients can be expressed using the masses and the angles, α, β, (the notation used is, c_X, s_X and t_X, to signify, $\cos X, \sin X$, and $\tan X$, respectively; where, $X = \alpha, \beta,$):

$$\lambda_1 = \frac{m_H^2 c_\alpha^2 + m_h^2 s_\alpha^2 - m_A^2 s_\beta^2}{v^2 c_\beta^2} - \lambda_5 t_\beta^2 - 2\lambda_6 t_\beta, \tag{4.15}$$

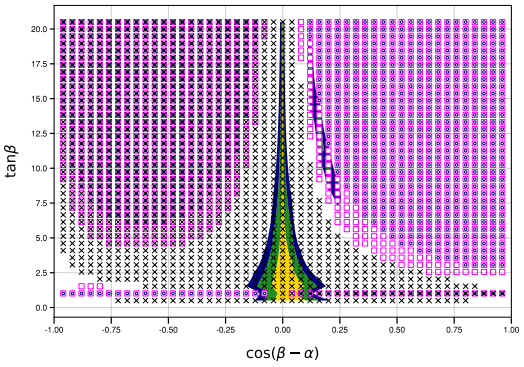
$$\lambda_2 = \frac{m_H^2 s_\alpha^2 + m_h^2 c_\alpha^2 - m_A^2 c_\beta^2}{v^2 s_\beta^2} - \lambda_5 t_\beta^{-2} - 2\lambda_7 t_\beta^{-1}, \tag{4.16}$$

$$\lambda_3 = \frac{(m_H^2 - m_h^2) s_\alpha c_\alpha + (2m_{H^\pm}^2 - m_A^2) s_\beta c_\beta}{v^2 s_\beta c_\beta} - \lambda_5 - \lambda_6 t_\beta^{-1} - \lambda_7 t_\beta, \tag{4.17}$$

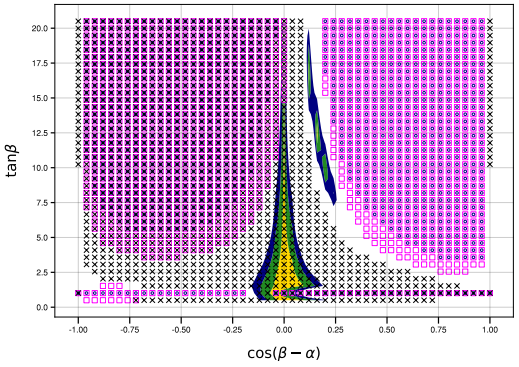
$$\lambda_4 = \frac{2(m_A^2 - m_{H^\pm}^2)}{v^2} + \lambda_5, \tag{4.18}$$

$$\lambda_4 - |\lambda_5| = \frac{2(m_A^2 - m_{H^\pm}^2)}{v^2} + \lambda_5 - |\lambda_5| = \begin{cases} \frac{2(m_A^2 - m_{H^\pm}^2)}{v^2}, & \text{if } \lambda_5 > 0, \\ \frac{2(m_A^2 - m_{H^\pm}^2)}{v^2} - 2|\lambda_5|, & \text{if } \lambda_5 < 0, \end{cases} \tag{4.19}$$

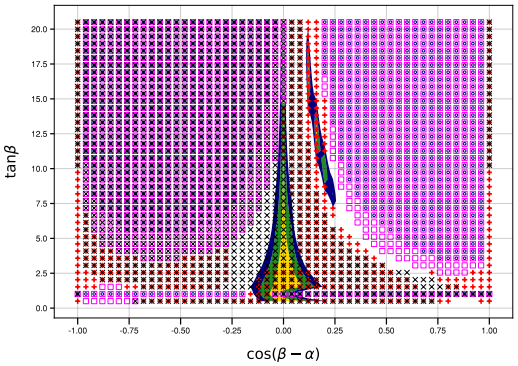
Amongst the theoretical conditions the most severe constraints on the parameter space come from the stability and perturbativity of the potential. We show an overview of the theoretical bounds in Figure 4.5. Here the 2HDM Type-II parameter space regions excluded by the different constraints are plotted. In order to illustrate this, we have fixed the masses of the 2HDM Higgs bosons to be, $m_{H^\pm} = m_H = 600$ GeV, and, $m_A =$ 300, and 400 GeV. Blue dots reflect the bounds arising from unitarity. We see that this has the largest impact in the medium-high $\tan\beta$ range as well as the for, $|\cos(\beta - \alpha)| \geq 0.1$. Positive(negative) values of Z_7 disfavour negative(positive) values of, $\cos(\beta - \alpha)$, this leads to the excluded region shifting to the right-(left-)hand side. Unitarity bounds do not impact the alignment and the wrong-sign domains. The regions allowed by both `HiggsSignals` and EWPO constraints are represented by the blue(green and yellow) region at the 95%(90% and 68%) C.L. . The magenta squares indicate regions excluded by the perturbativity constraint, which extends the region of excluded points down towards lower values of $\tan\beta$ and $|\cos(\beta - \alpha)|$ for a given value for the quartic Higgs coupling. We also see in the top two plots of Figure 4.5 that when $Z_7 = 0$ the wrong-sign contour becomes completely excluded by the perturbativity constraint (see left plot). Increasing this to $Z_7 = 0.6$, the excluded region shifts and the wrong-sign contour becomes visible (and not excluded) again (see right plot). While it is not explicitly shown the figure, there seems to be a trend in how the value of Z_7 affects the excluded region. Considering that we can deduce that a negative Z_7 value would tend to exclude the alignment region. Finally, the black crosses that represent the exclusion due to the stability of the potential, eliminates all negative values of $\cos(\beta - \alpha)$ and some of the positive values too in such a way that the alignment domain is almost completely suppressed.



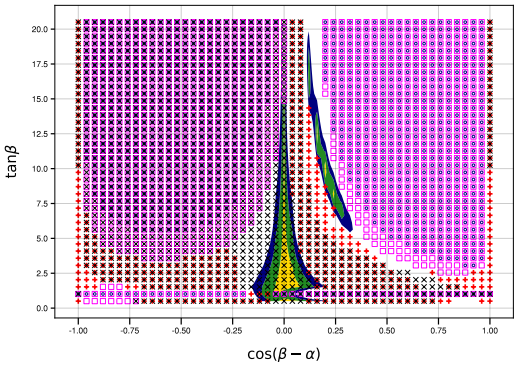
$m_A = 300$ GeV and, $Z_7 = 0$



$m_A = 300$ GeV and, $Z_7 = 0.6$



$m_A = 300$ GeV, and, $Z_7 = 0.6$



$m_A = 400$ GeV, and $Z_7 = 0.6$

4.5 parameter space points plotted on the, $(\cos(\beta - \alpha), \tan\beta)$, plane. Each plot shows the `HiggsSignals` + EWPO allowed regions in yellow, green and blue for compatibility of C.L. 1, 2 and 3σ respectively. Points excluded by the theoretical constraints of unitarity (blue hollow dots), perturbativity (magenta hollow squares) and stability (black crosses) in the 2HDM Type-II are also shown. In the top row: LHS, $m_A = 300$ GeV and, $Z_7 = 0$; RHS, $m_A = 300$ GeV and, $Z_7 = 0.6$. Bottom row: LHS, $m_A = 300$ GeV, and, $Z_7 = 0.6$; RHS, $m_A = 400$ GeV, and $Z_7 = 0.6$. Finally, points excluded by `HiggsBounds` are shown as red crosses.

To surmise the combined effect of the unitarity, perturbativity and stability constraints; negative values of Z_7 appear to be disfavoured. For our setup, there were no points lying in the alignment region; they are instead concentrated within the wrong-sign domain (see lower right plot). We also saw that generally, increasing the m_A value would result in both the alignment and the wrong-sign scenarios becoming populated once more. In the case of the alignment contour this happens at extremely low $\tan\beta$ values.

This exercise leads us to the conclusion that a lower bound is enforced on the pseudo-scalar mass m_A by the stability of the scalar potential when we are situated in the alignment part of the parameter space.

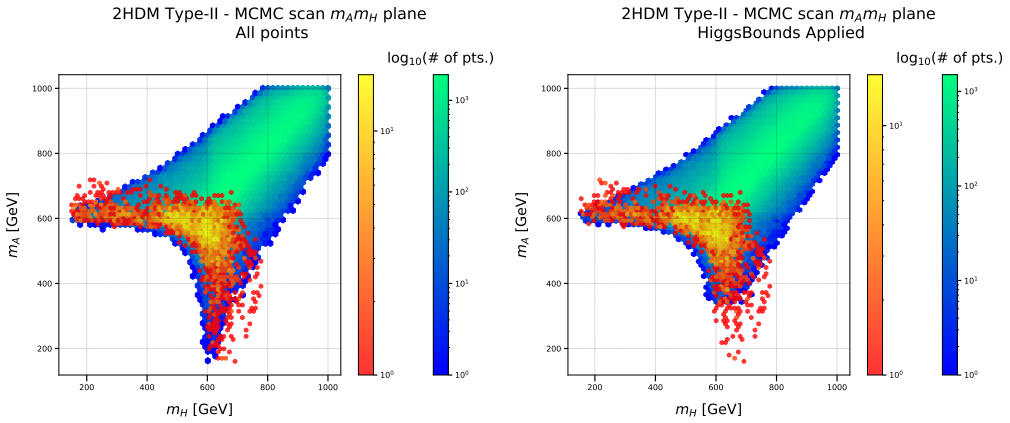
4.2.3 Importance of m_A

Now we investigate what conditions a stable scalar potential enforces and what effect these have on the two limits of our focal model: the alignment and wrong-sign domains. We take a set of points from our MCMC scan which pass the condition, $\Delta\chi^2_{\text{tot}} < (3\sigma$, C.L. upper limit), and we do not impose any further constraints at this time. The inequalities in (Equation 4.14) are implemented one at a time in order to be able to uniquely identify their effect on the parameter space. We note the following;

- Before we impose any stability conditions we have points in both the alignment and wrong-sign regions.
- Imposing, $\lambda_1, \lambda_2 > 0$, removes points from both regions irrespective of the m_A value. However there are surviving points in both regions afterwards.
- Imposing, $\lambda_3 + \sqrt{\lambda_1\lambda_2} > 0$, eliminates a large number of points solely from the alignment limit where the m_A value is high. However it does not exclude any additional points for low m_A values.
- Imposing the final constraint, $\lambda_3 + \sqrt{\lambda_1\lambda_2} + \lambda_4 - |\lambda_5] > 0$, We find that points from the alignment solution are disfavoured again with the value of m_A having no impact. Crucially this serves to exclude most of the remaining points from the alignment region for the low-intermediate m_A range; barring a handful of points with low $\tan\beta$. On the other end of the m_A value range we find surviving points in both regions even after we have imposed all conditions.

This result is presented in the hexabin plots of Figure 4.6, where we show the, (m_H, m_A), parameter space. The blue-green gradient dots are the alignment region and the red-yellow gradient ones are the from wrong-sign scenario (be sure to note that the scales of the alignment and wrong-sign colour-bars differ, this is due to the difference in the amount of remaining points after enforcing constraints). These plots show what remains of our data after we enforce experimental bounds coming from `HiggsSignals`, plus EWPOs and the theoretical constraints mentioned above. In addition, we set the lower bound on $m_H^{\pm}$ to 600 GeV. Looking at the LHS plot, it is easy to see that there are few points left for the alignment solution for low m_A. The few that do remain are characterised by having very low $\tan\beta$ values, as was previously discussed. If we then add the `HiggsBounds` limits, (see the RHS plot), we find that even these remaining points disappear.

The overall picture can be seen in the RHS plot of Figure 4.6. For the alignment limit of the 2HDM we can see that the pseudo-scalar state is required to be fairly heavy: $m_A \geq 350$ GeV. In wrong-sign scenario, however, it can conceivably have a mass as small as, $m_A \simeq 150$ GeV, (see red-yellow dots), if Z_7 is positive and quite large, as was shown in Figure 4.5. This final feature results from the effects of perturbativity enforcement. This is dependent on whether or not the lower limit on the charged Higgs

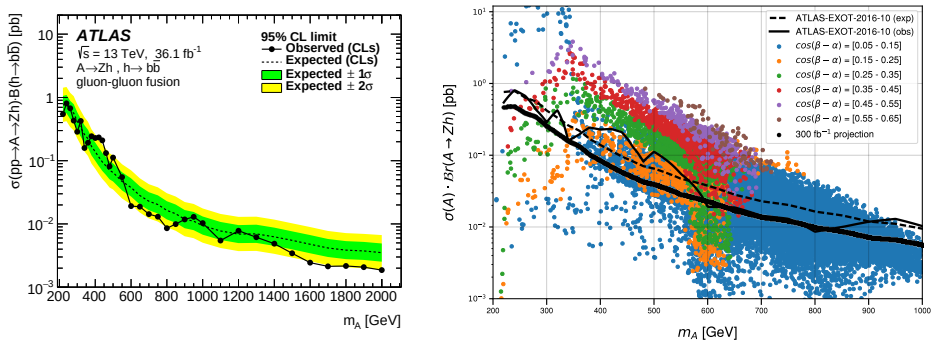


4.6 Distribution of the parameter space points on the, (m_A, m_H), plane allowed by the theoretical constraints in the 2HDM Type-II. The bound on the charged Higgs mass has been implemented as, $m_{H^\pm} \geq 600$ GeV. In the left plot, the `HiggsSignals`, EWPOs and theoretical constraints are enforced. In the right plot, `HiggsBounds` limits are also enforced. The blue/green dots are the alignment region while the red/yellow ones refer to the wrong-sign scenario. Both plots have a pair of colour bars shown next to them, which give a logarithmic scale count to the colours used for the plot.

mass increases significantly. As the $m_{H^\pm}$ limit increases it simultaneously pushes up the lower bound on m_A for the alignment scenario. Interestingly, in the wrong-sign domain it is still possible to have light *CP*-odd Higgs boson masses at the cost of warping Z_7 towards larger positive values, with, $Z_7 \geq 1$, typically. This agrees with the findings of Ref. [30]. However, we have been able to add a more comprehensive analysis of effects coming from each individual constraint. In particular we have been able to highlight how the stability requirement for the scalar potential sets a lower bound on the *CP*-odd Higgs mass specifically for the alignment scenario.

4.3 Data Interpretation

Here we apply the methodology of the global scan tool, `Magellan`, to interpret the LHC data within the 2HDM Type-II. During an MCMC scan within `Magellan`, many experimental and theoretical properties linked to each parameter space point are computed and saved. The retention of this information allows us to examine different aspects of the model with the same dataset. We can then translate any new unfolded experimental results to direct bounds on the parameter space of the 2HDM Type-II. The experimental results corresponding to a given observable, typically the 95% C.L. exclusion bound on the cross-section times BR, can be projected onto any two-dimensional sections of the full parameter space, thus allowing the extraction of limits on different parameters of the theory. The observables, i.e., cross-sections and BRs used for comparison, are computed by with the use of `SusHi` and `2HDMC`.



4.7 Left plot: 95% C.L. upper bound on the cross-section times BRs, $\sigma(pp \to A \to Zh \to Zb\bar{b})$, as a function of the *CP*-odd Higgs boson mass, extracted by ATLAS at the 13 TeV LHC [14]. Right-plot: Theoretical predictions for the same process, $pp \to A \to Zh \to Zb\bar{b}$, within the 2HDM Type-II (here, $\sigma_A \equiv \sigma(pp \to A)$). The different colours of the points in the scatter plot represent different values of $\cos(\beta - \alpha)$. Superimposed, there is the ATLAS observed (expected) cross-section times BR given by the black solid (dashed) line. Finally, the heavy black curve shows the projection of the expect limit curve of the ATLAS analysis to a luminosity of, $L = 300$ fb^{-1}.

4.3.1 Applying a New Analysis with `Magellan`

As an example we now consider (what was at the time of this investigation) the most recent ATLAS analysis[14] for the process

$$pp \to A \to Zh \to Zb\bar{b}, \tag{4.20}$$

and apply this onto our dataset with `Magellan`. The analysis is on the search for a heavy *CP*-odd Higgs boson, A, within a 2HDM, decaying to a Z boson plus the 125 GeV SM-like Higgs. This is performed by studying only final states which contain either a pair of opposite-sign charged leptons, (l^+l^- with, $l = e, \mu$), *or* pair of neutrinos ($\nu\bar{\nu}$) with two b-jets at the 13 TeV LHC having a total integrated luminosity of, $L = 36.1$, fb^{-1}.

Figure 4.7 displays the 95% C.L. upper bound on the cross-section multiplied by the BR as a function of m_A. In this figure, the hypothetical signal is taken to have come as a result of pure gluon-gluon fusion; in our own analysis, however, we have also accounted for the quark-antiquark annihilation production mode as it is typically sub-leading by comparison. The RHS plot of the same figure shows the theoretical cross-section times BR, computed for a subset of that same range in m_A. We have colour coded the points in the plot so that each one corresponds to a set value range for $\cos(\alpha - \beta)$, the legend indicates these value ranges. The product of the cross-section and BR depends on this *cosine* parameter, as can be clearly seen in the RHS figure. $\cos(\alpha - \beta)$ is a key factor in the values of the couplings of; the *CP*-odd Higgs boson, A, the heavy quarks present in the production subprocess as well as subsequently with the Z and h bosons in the decay chain.

We have superimposed the observed and expected curves from the ATLAS analysis (see LHS plot) onto the RHS plot[2] . Clearly we see that the excluded range of the *CP*-odd Higgs boson mass appears related to the value of, $\cos(\alpha - \beta)$.

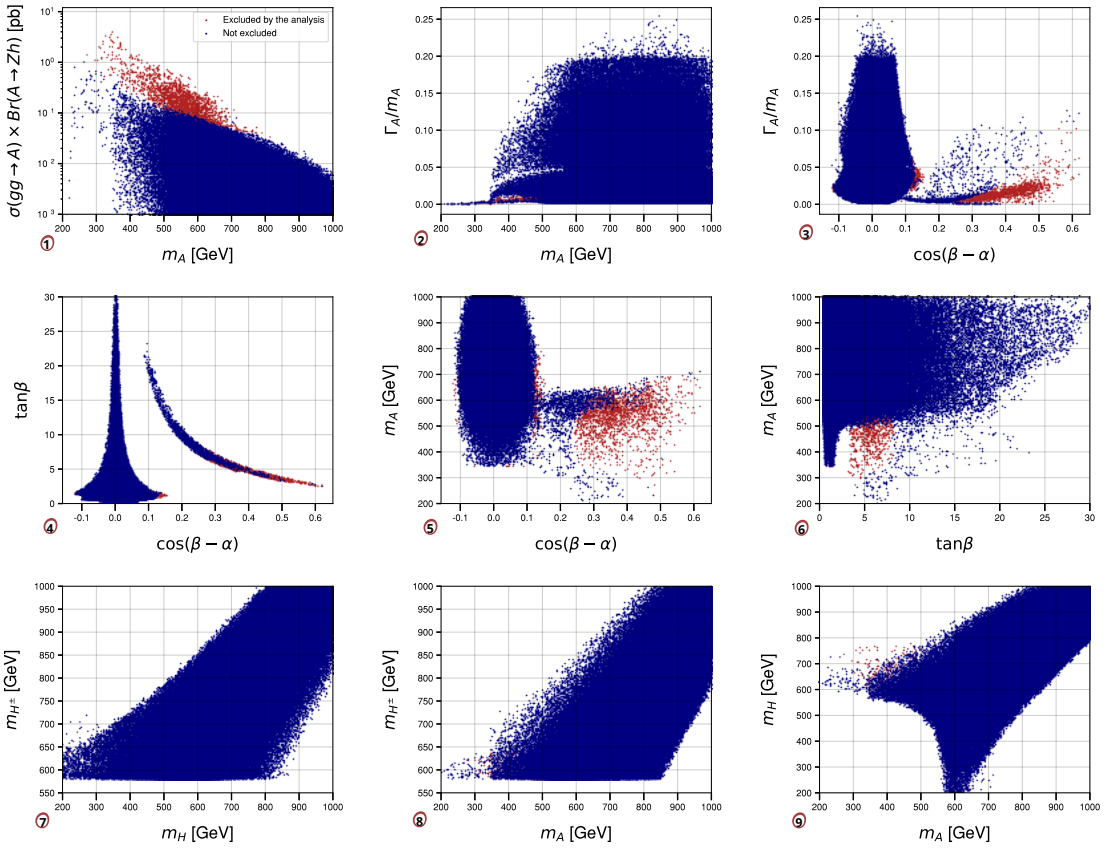
`Magellan` allows the user to access a diverse variety of variables at once, rather than having to take the usual approach of holding some parameters at a constant value etc. . The toolbox benefits from the `DataFrame` class of `pandas`, which allows for many variables associated with a single point to be recorded in a convenient way. This makes applying a custom selection on the set of points relatively easy. Points excluded (or allowed) by a particular theoretical constraint or experimental bound can then be projected onto any other plane, i.e. we can select some interesting points from one bi-dimensional plane and quickly project onto a second one with two different parameters, but still showing the same points as before. Specifically considering the case mentioned above, we are able to select points that are above the 95% C.L. upper bound seen in the form of the solid black line on the RHS plot of Figure 4.7; we can then project these points to see what effect that specific model-independent measurement has on all the free parameters of the 2HDM Type-II. It is worth noting that the limits coming from the experimental analyses entered on HEPData (`https://www.hepdata.net/`) are dependent on the assumption made on the width of the hypothetical 2HDM Higgs bosons. The width of the (pseudo)scalar states is similarly taken into account when bounds are extracted on the parameter space [3].

We created a series of bi-dimensional plots to demonstrate this feature in Figure 4.8. There are nine different projections onto different planes from the parameters/observables of the 2HDM. We first plot points excluded by the ATLAS analysis (shown in red), then the "allowed" points (blue); which is done without taking into consideration other parameter values that are not displayed. Naturally this could be done in the reverse order instead, which would be effective to highlight the region of the parameter space tested by the particular experimental measurement being considered. We have implemented both of these options and placed them on the `Magellan` interactive webpage [144] should there be interest.

Even looking at just this subgroup of possible parameter spaces, we can conclude that it is the low $\tan\beta$ region that the ATLAS 13 TeV analysis has tested at the LHC [14]. i.e., where $\tan\beta \leq 5$ (c.f. the ($\cos(\beta-\alpha)$, $\tan\beta$) plane, labelled 4). $\cos(\beta-\alpha) \geq 0.5$ is nearly excluded for all possible m_A masses (c.f. the ($\cos(\beta-\alpha)$, m_A) projection, labelled 5). Thus, even if a colourless scatter plot was initially created for the, $pp \to A \to Zh \to Zb\bar{b}$, rate as a function of m_A, purposefully hiding information on the $\cos(\beta-\alpha)$ value of individual points from ourselves, this projection feature could still allow us to learn about the range of $\cos(\beta-\alpha)$ and $\tan\beta$ that is being tested. It is of course possible that higher luminosities could be sensitive to larger values of $\tan\beta$ and

[2] note that the apparent difference in the shape of these curves is simply a result of the different ranges of m_A, with the original analysis extending to, $m_A = 2000GeV$, while our own data extends to $1000MeV$ only

[3] Experimental limits are available up to, $\Gamma_A/m_A < 11\%$.



4.8 A series of different projections of the 2HDM Type-II parameters and observables. Blue points are allowed by `HiggsSignals`, EWPOs and theoretical constraints. Red points are those excluded by the ATLAS analysis of, $pp \to A \to Zh \to Zb\bar{b}$, at a luminosity, $L = 36.1$ fb^{-1}. (Note that we do **not** enforce `HiggsBounds` constraints at this stage.) Points which fail against one or more of the limits from `HiggsSignals`, EWPOs or theoretical constraints have been removed and thus the white background corresponds to this region of parameter space.

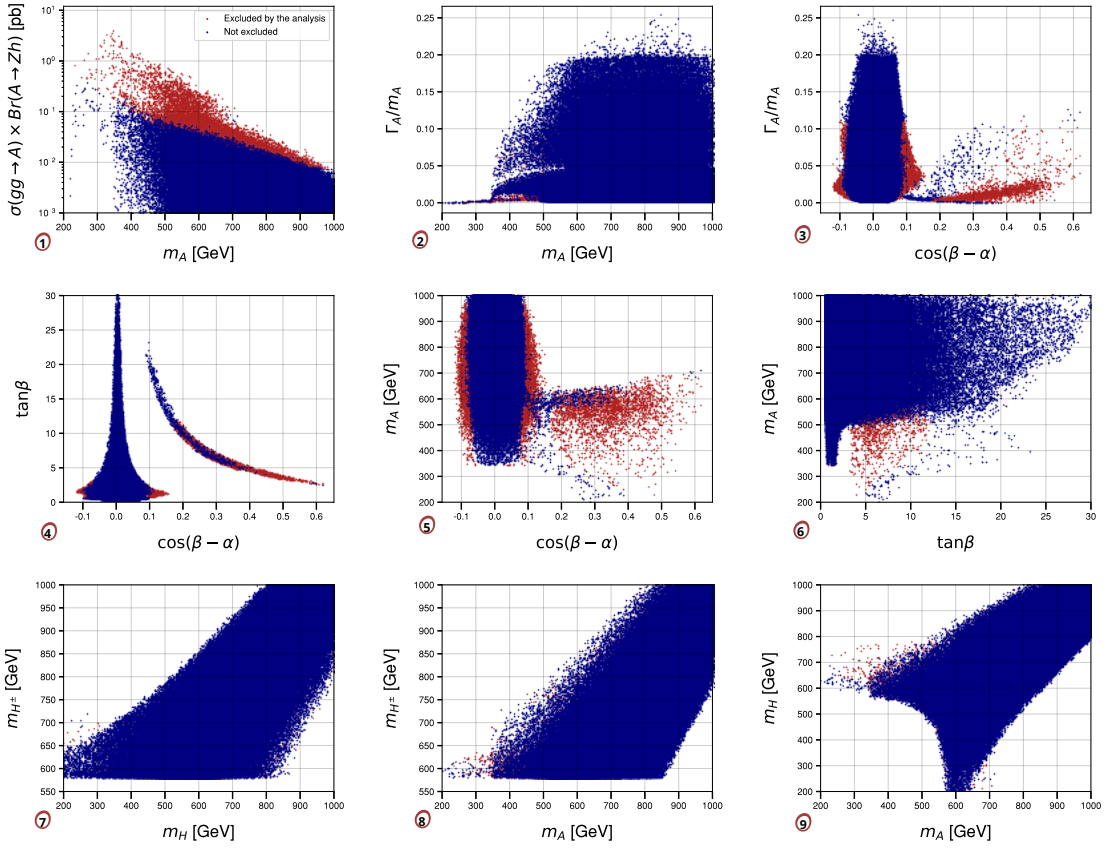
equally to smaller values for $\cos(\beta-\alpha)$, and so would extend the search of new physics, particularly in the wrong-sign region.

We also see in the top-right plot,(labelled 3), which shows the width of the A divided by the corresponding mass as a function of $\cos(\beta-\alpha)$, that the analysis covers the parameter space up to, $\Gamma_A/m_A \leq 11\%$. However our dataset indicates that there are clearly possible values for this ratio, extending up to, $\Gamma_A/m_A \simeq 25\%$. This implies that it may be beneficial for future experimental analyses to rely less on the narrow width approximation and try to include wider resonances in the search.

We can take this comparison further by accounting for the limit on the cross-section times BR expected in the near future[4] with a luminosity set to, $L = 300$ fb^{-1}. Looking at the projected exclusion bounds on $\cos(\alpha-\beta)$ for this increased luminosity indicates a reasonable improvement, dropping the bound down and capturing a greater range of m_A for different values of $\cos(\alpha-\beta)$.

[4] from the perspective of when this project was carried out

In Figure 4.9 we plot the same series of planes as in Figure 4.8 but now use the projection for the expected limit on the production cross-section times BR for the process, $pp \to A \to Zh$, with an integrated luminosity of, $L = 300$ fb^{-1}. The excluded points are again shown in red, we see that quite a significant section of the parameter space is ruled out by this increased luminosity. The region, $\cos(\beta - \alpha) \geq 0.4$, can be all but completely excluded, as shown by the plots numbered 3, 4 and 5. In addition points begin to be excluded from the alignment region. This alone starts to paint a picture of what we can expect to see when the luminosity of the LHC reaches 300 fb^{-1}.



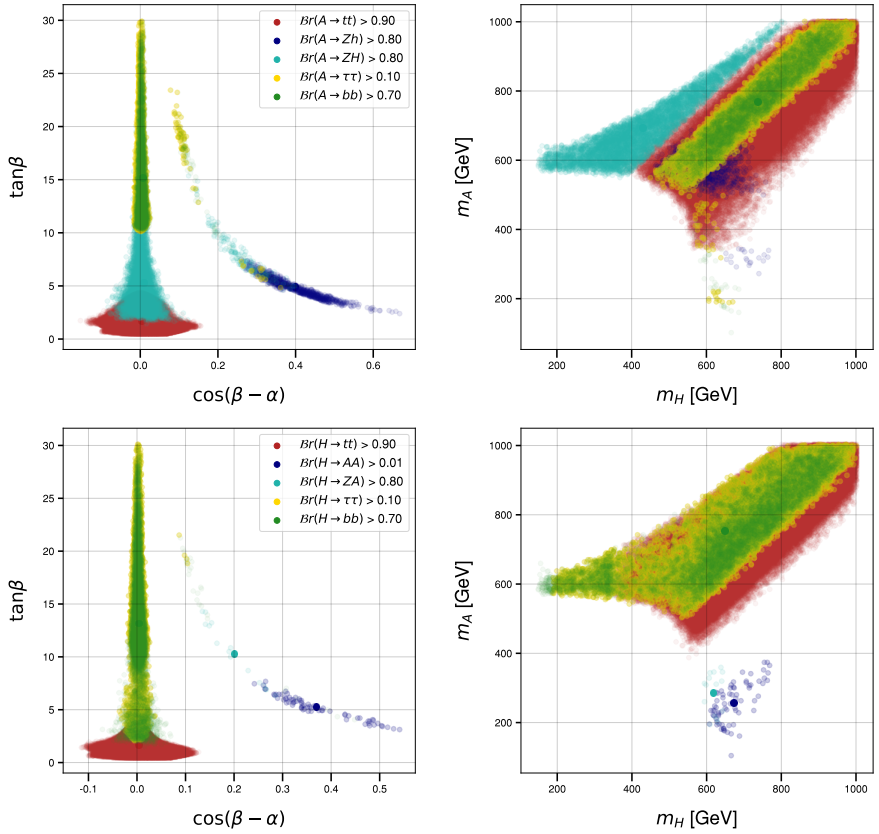
4.9 Blue points are allowed by `HiggsSignals`, EWPOs and theoretical constraints. The red ones are those excluded by the ATLAS analysis of the process, $pp \to A \to Zh \to Zb\bar{b}$, projected to a luminosity of, $L =$ 300 fb^{-1}. (Note that we do not enforce `HiggsBounds` constraints at this stage.) Points which fail against one or more of the limits from `HiggsSignals`, EWPOs or theoretical constraints have been removed and thus the white background corresponds to this region of parameter space.

We find that interpreting the model-independent experimental data for a given model is considerably more flexible and thorough than the use of more standard procedures adopted in the literature. With reference to the, $pp \to A \to Zh$, search performed by ATLAS [14] that we have examined, the interpretation of the limits on the cross-section times BR within the 2HDM, the masses of the additional Higgs bosons are assumed to be degenerate. Our analysis, however, allows the three masses to differ by 250 GeV or more, as can be seen in subsection 4.2.1. In general the visualisation of

the limits at 95% C.L. on the 2HDM parameters as presented in Ref. [14] is constrained and therefore can only be partial. We see bounds displayed at a fixed resonance mass m_A on the, $(\tan\beta, \cos(\beta-\alpha))$, plane, as well as on the, $(\tan\beta, m_A)$, plane, for a fixed value of $\cos(\alpha-\beta)$. Instead with `Magellan` have been able to create a more global scan, presenting full limits on any desired 2D plane. This naturally offers access to a richer variety of information.

4.3.2 2HDM sensitivity of different measurements at the LHC

We now move to analyse several possible measurements that can be made at the LHC, our aim being to demonstrate that they have sensitivity to specific model parameters of the 2HDM Type-II. First we will cover the relevance of the different channels, which may one or more Higgs bosons as an intermediate state, in covering the parameter space of the A and the H via study of their BRs. Some of these overlap to a degree while others do not at all, as can be seen in Figure 4.10.

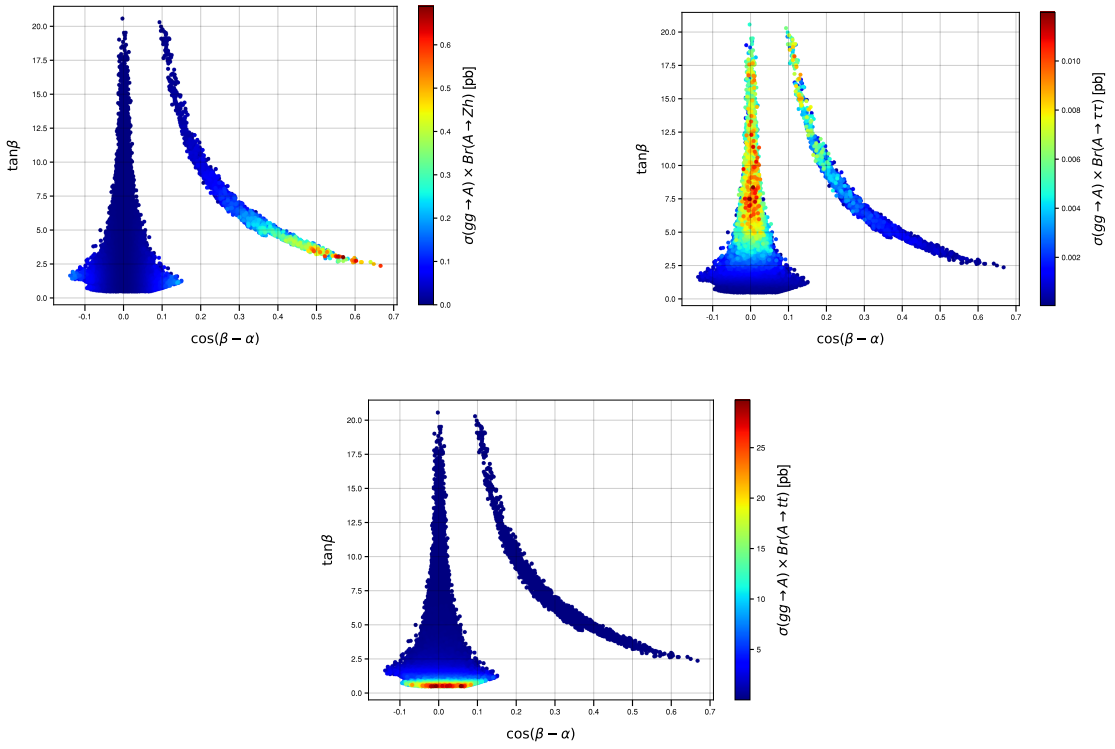


4.10 Top plots: show areas of the two-dimensional parameter spaces with high BRs of the *CP*-odd Higgs boson A. The LHS plot shows the different BR channels in the legend. Bottom plots: show the same regions as before but for the heavy *CP*-even Higgs boson H, again, the LHS plot has a legend indicating the decay channels.

In the top two plots of Figure 4.10, the complementary coverage of, $A \to t\bar{t}$, (red) and, $A \to ZH$, (light blue) channels is apparent. The first of these processes has sensitivity to low $\tan\beta$ values (see top LHS plot) and covers a large range of the mass spectrum

when the A and H masses are separated by no more than 200 GeV (see top RHS plot).When the $\tan\beta$ values are medium-high there is an enhancement to the number of A decaying to down type particles, b-quarks or τ-leptons, this can be seen in the green/yellow points of the top LHS plot. Lastly, there is particular sensitivity in the, $A \to Zh$, mode to large values of $\cos(\beta - \alpha)$ with low-medium $\tan\beta$.

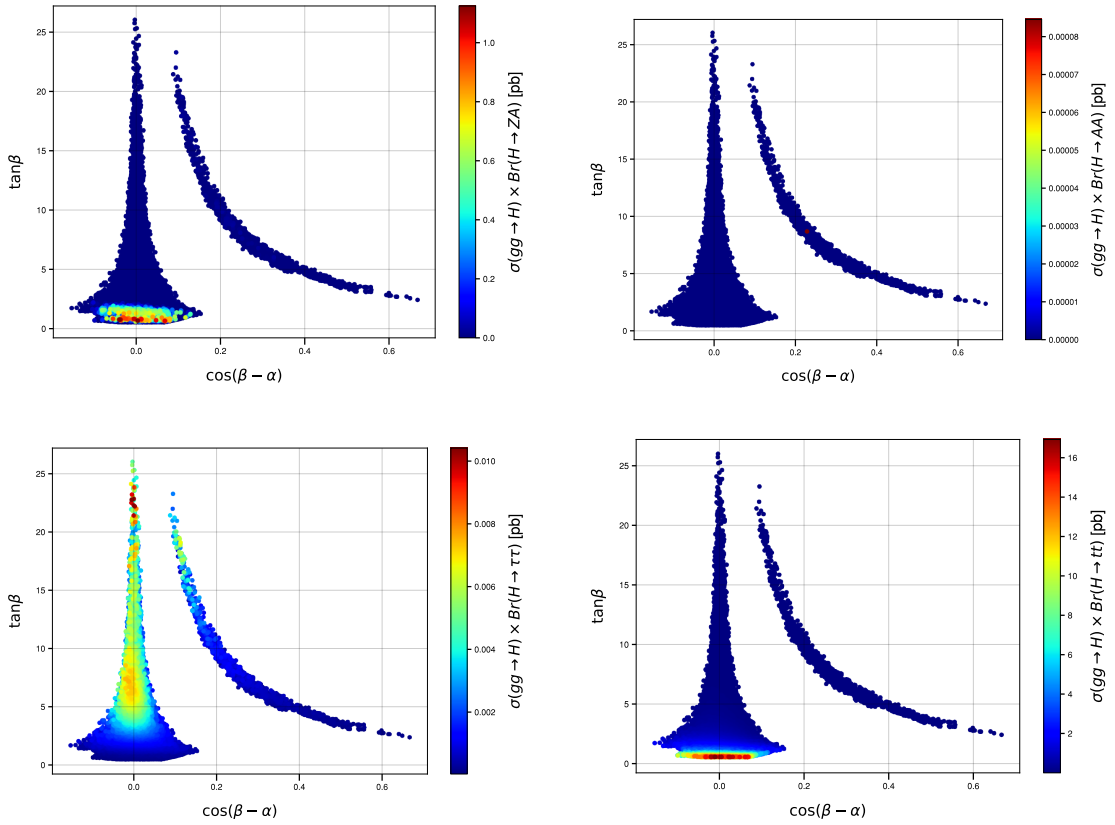
Continuing on to examine the H decay modes (the bottom row of Figure 4.10). The high $\tan\beta$ region shows us one dominated by decays into $b\bar{b}$ and $\tau^+\tau^-$; while at low $\tan\beta$ this instead is replaced by $t\bar{t}$, ZA. The majority of our decays are concentrated within the alignment region. From this we can conclude that processes mediated by H do not display sensitivity to the region with large $\cos(\beta - \alpha)$. In order to investigate or indeed exclude this section of the parameter space, i.e. the wrong-sign scenario, we must instead lean on processes which are mediated by the A state, specifically, $A \to Zh$. Observe that there may be some layering in Figure 4.10 whenever two or more criteria are met, however, we have attempted to minimise overlapping regions to avoid any impact on the overall picture given.



4.11 Total cross-section multiplied by BRs for three different, CP-odd Higgs mediated processes in the, $(\cos(\alpha - \beta), \tan\beta)$, plane. The top LHS is, $pp \to A \to Zh$, the top RHS is, $pp \to A \to \tau^-\tau^+$, and the bottom plot is, $pp \to A \to t\bar{t}$.

Naturally these decay modes can only provide an incomplete image of how much sensitivity experimental searches have to the free parameters of the 2HDM. For a more complete view we must take into account the actual production rate, i.e. , the production cross-section multiples by the BR/s. As such we have plotted this in Figure 4.11 for the *CP*-odd Higgs state and in Figure 4.12 for the *CP*-even Higgs state.

Each subplot is given on the bi-dimensional, $(\cos(\beta-\alpha), \tan\beta)$, plane. The magnitude of each subplot's total cross-section is indicated by the colour bar on the RHS of its respective plot. We see a range for the A mediated processes from only a few fb corresponding to the $\tau^+\tau^-$ channel, to 30 pb, corresponding to, $pp \to A \to t\bar{t}$. Analogous results are found for the H mediated processes.



4.12 Total cross-section multiples BR for four different processes mediated by the heavy CP-even Higgs boson, in the, $(\cos(\alpha-\beta), \tan\beta)$, plane . Going clockwise from the top-left to the bottom-left we have: $pp \to H \to ZA$, $pp \to H \to AA$, $pp \to H \to t\bar{t}$, and, $pp \to H \to \tau^-\tau^+$.

Chapter 5

Seeking A Detectable Cross-section In The 2HDM

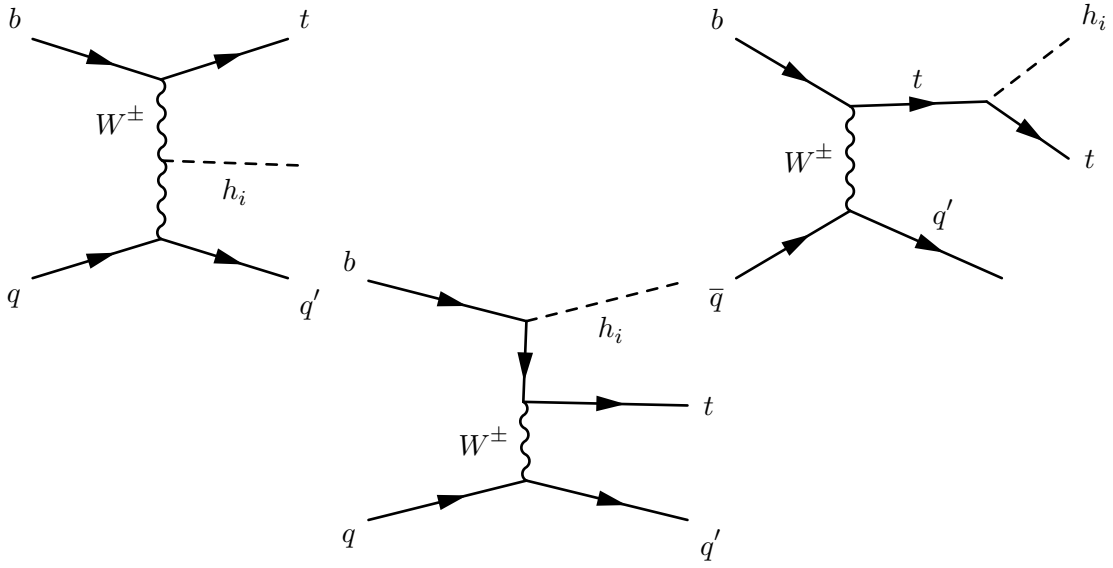
Having already shown that `Magellan` showed promise we endeavoured to upgrade several of the constituent HEP software alongside gathering data for our next project with it.

The Higgs boson discovered in 2012 at the LHC has had its physical properties extensively measured and in all respects it has come out as being consistent with the Higgs we expect from SM predictions. An aspect that has proven difficult to obtain a direct measurement for is the couplings to the weak bosons and the, t, b, c, τ, μ, fermions. Within the SM we are mainly sensitive to the modulus of these couplings rather than to the couplings themselves directly. This is due to the processes of production and decay used in our measurements; they do not exhibit a large amount of interference between one another or with others. SM access to examine the signs of couplings specifically occurs mainly in, $h \to \gamma\gamma$, and, $h \to Z\gamma$, decays, which, (unfortunately), are affected by the appearance of, $t, \bar{t}$, quarks and, W^+, W^-, at loop level, making the background too large to distinguish our signal. Importantly for our purposes, we are also missing access to the sign of the Higgs boson to bottom (anti)quark coupling. As discussed in section 2.1 the SM-like Higgs has Yukawa-type couplings to fermions, which implies that the strength of $ht\bar{t}$ is much greater than that of $hb\bar{b}$, thus in the aforementioned decay processes the latter plays a negligible role when compared to that of the former, (this is also true in the production process $gg \to h$)[1].

However, if we instead consider the 2HDM and resulting interactions then our access changes, the $hb\bar{b}$ coupling can take the opposite sign to that of the SM coupling. This would have no impact on current measurements, due to the sensitivity being to the modulus as mentioned, but it could result in a significant boost to the cross-section of some alternative ones. This may mean that if we look to alternative measurements

[1] A recent review of the current LHC status on the nature of the SM-like Higgs boson can be found in [104].

from those currently being performed at the LHC, we may have an important avenue to establishing whether or not the coupling is SM like. In order to do this it is necessary to consider other h boson production channels alongside the conventional ones. For example: gluon-gluon fusion, ($gg \to h$), vector-boson fusion, ($qq \to q'q'h$),[2] and associated production with weak gauge bosons, ($q\bar{q}^{'R} \to Zh(W^{\pm}h)$), or top (anti)quark pairs, ($q\bar{q}, gg \to t\bar{t}h$), (see Ref. [111] for a review). In this instance we focus on SM-like h boson production in association with a single top (anti)quark. The leading subprocesses for this are (in order of SM magnitude) the following: $bq \to tq'h$ + c.c. (hereafter, bq), $bg \to tW^{-}h$ + c.c. (hereafter, bg) and $q\bar{q}' \to t\bar{b}h$ + c.c. (hereafter, qq), see Figure 5.1–Figure 5.2.

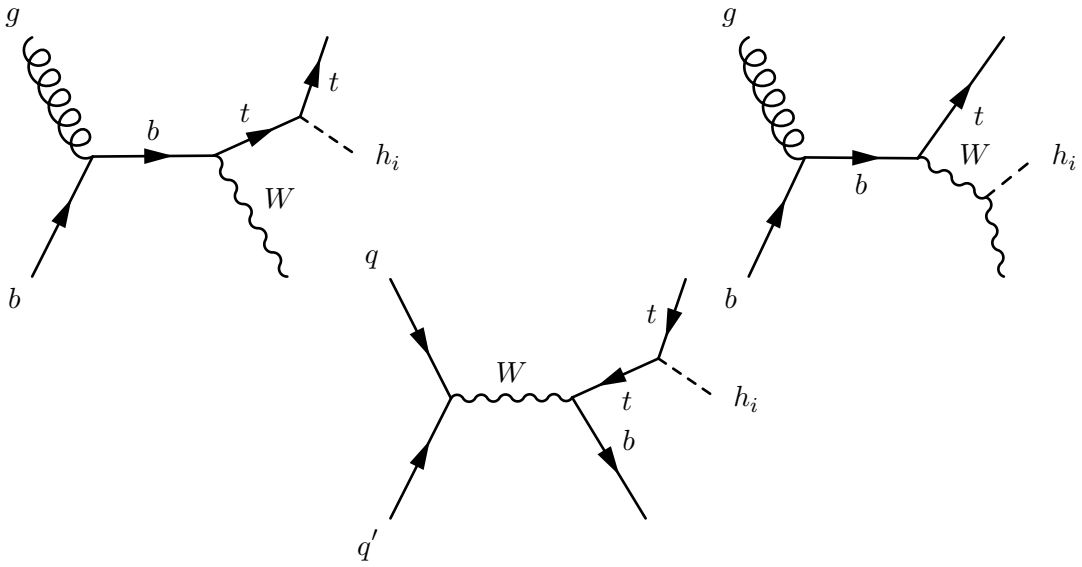


5.1 Feynman diagrams for the bq subprocess, assuming time flowing rightwards, wherein we ignore the contribution of a charged Higgs boson ($H^{\pm}$), which we take heavy enough so as to give a negligible correction. Notice that same diagrams appear in the qq subprocess, when time is flowing upwards.

The SM cross sections for these processes can be seen in Table 5.1, as calculated by `MadGraph5@NLO`. When summed their combined production cross-section is smaller than that of, $q\bar{q}, gg \to t\bar{t}h$, but of the same order. This indicates that some sensitivity already exists for this additional h production mechanism[55; 133; 52]. Presently, in all such analyses, the bq, bg and qq channels are treated inclusively. The three subprocesses contributing to the thj process at tree level, in order of predicted size in the SM, are;

As it currently stands, searches are only able to exclude cross-sections for the SM-like Higgs production in association with a single $t(\bar{t})$ with cross-sections greater than the SM predicts. This is largely as a result of cancellations between some of the topologies involved. Specifically between those depicted in Figure 5.1–Figure 5.2. This is as a result of the fact that, unlike any of the previously mentioned production processes,

[2] Hereafter, 'q'' denotes a light quark, either, d, u, s or c.



5.2 Feynman diagrams for the bg subprocess along with one for the qq' process (the remainder of which are the same as the diagrams in Figure 5.1 but with time flowing upwards), assuming time flowing rightwards, wherein we ignore the contribution of a charged Higgs boson ($H^{\pm}$), which we take heavy enough so as to give a negligible correction.

	σ(bq) (pb)	σ(bg) (pb)	σ(qq) (pb)	σ(total) (pb)
SM	0.036	0.011	0.0023	0.049[55]

5.1 The tree-level cross-sections for the bq, bg and qq subprocesses of SM-like Higgs boson production in association with a single top (anti)quark at the LHC with 14 TeV of Centre-of-Mass (CM) energy. (These values have been calculated by `MadGraph5@NLO-3.1.0` [121] for the default SM implementation that comes with the package.)

the h couplings to $W^{\pm}$ bosons, as well as to both t and b quarks all enter simultaneously at amplitude level, allowing interference between them. Consequently, when the sign of these couplings changes so too might the size of the cross sections. It is possible that these cross sections could be accessible by the LHC during run 3. In addition to this, we may see differing kinematics in the final states in a BSM scenario compared to the equivalent in the SM scenario. Note that the couplings for the charged Higgs bosons of the 2HDM were omitted in this search due to the high mass constraints on the charged Higgs.

Hence, we seek to investigate this potential for constructive interference, studying the possibility of such a phenomenology being realised in the simplest extension of the SM Higgs sector. The extension uses an additional doublet field (akin to the pre-existing one within the SM), i.e., the one embedded in a generic 2HDM [36]. We consider the $h \to b\bar{b}$ decay channel, this being the dominant one within the SM, but, despite this dominance it is poorly measured. This is as a consequence of the formidable background accompanying it whenever the standard four production processes are used

for searches[3] would be needed to extract a signal for SM-like Higgs boson production in association with a single top (anti)quark. An integrated luminosity of this magnitude would only achievable at the High-Luminosity LHC (HL-LHC). We intend to show that within the 2HDM type-II a significant increase of the cross-section is possible, for the bg subprocess specifically. We take particular interest in solutions that occur when the $hb\bar{b}$ coupling changes sign with respect to the SM, i.e., the so-called 'wrong-sign solution', (to the current SM-like Higgs boson measurements). We also investigate the kinematics emerging as a result of the coupling change between the SM and 2HDM type-II cases. We hope this will provide suitable motivation for the experimental pursuit of this additional h production channel, with a twofold purpose; to prove that a BSM Higgs sector exists and secondly, to show that (at least for the 2HDM type-II) it should be possible to establish the basal structure of it .

As before, the hybrid parametrisation (with the recasting of Z_4 and Z_5 as listed previously in Equation 4.4)

The couplings of the neutral Higgses to fermions, normalised to the corresponding SM value (m_f/v, henceforth, denoted by κ_{hqq} or simply κ_{qq} for the case of the SM-like Higgs state coupling to a quark q, where $q = d, u$), can be found in Table 2.1.

We chose to focus on the simpler realisations of the 2HDM, Type-I and Type-II.

[3]Indeed, current estimates suggest that a minimum of 1500 fb^{-1} of integrated luminosity in the $h \to b\bar{b}$ channel [107][75]

5.1 Parameter-point Generation

Here, we intend to detail how the constraints on the 2HDM Types I and II parameter spaces are normally drawn and how `Magellan` was used to scan the available parameter space and calculate cross-sections for the generated points.

5.1.1 Tools

As before (in chapter 4), there are 6 parameters that make up our parameter space. These are:

$$Z_7 \ , \ m_H \ , \ m_{Hc} \ , \ m_A \ , \ \cos(\beta - \alpha) \ , \ \tan(\beta)$$

As `Magellan` was being upgraded in parallel to this work and thus incorporated slightly out of date versions of these three programs, it was thus deemed important to check the output from `Magellan` externally using the most up-to-date versions of each . At this point there was a significant loss of points as more recent observations have since been incorporated into `HiggsBounds` and `HiggsSignals`, squeezing the available parameter space, in particular for the type-II (This effect can be seen by comparing to Figure 4.4 in page 39 to Figure 5.3 & Figure 5.4).

Following this `Magellan`'s pipeline to scan CSVs of these points through `MadGraph5@NLO` (v3.1.0) [6][121] was used, which then calculated the cross-section for the three contributing subprocesses individually. This was done for both the type-I and type-II models, as these presented areas of greatest interest. For the type-I we used the model 'THDM_type1_UFO', created by Souad Semlali [3] using `FeynRules` [67][5], and for the type-II we used '2HDMtII_NLO' [70] [4] We used a default PDF that comes with `MadGraph5@NLO`, the NNPDF 2.3 LO PDF[57].

5.1.2 Constraints

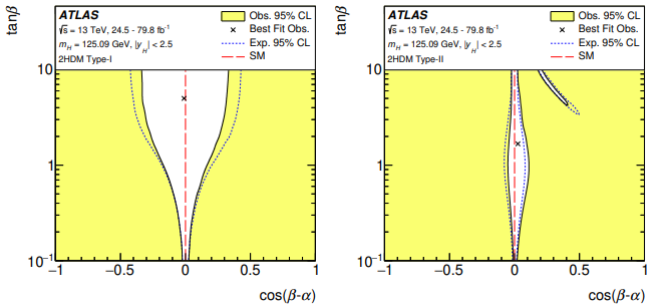
In Figure 5.3 we can clearly see that the constraints for the Type-II model are much tighter than those of Type-I, thus there is a larger parameter space to be scanned for potentially high cross-sections. We see a very similar picture when we look at equivalent plots from CMS in Figure 5.4

[1] `HiggsBounds`-5.1.0beta, `HiggsSignals`-2.2.3, `2HDMC`-1.7.0

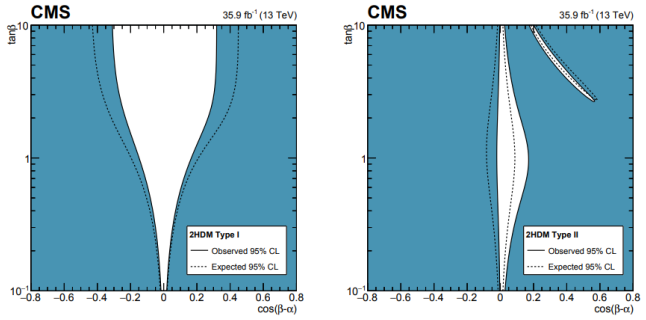
[2] `HiggsBounds`-5.10.1, `HiggsSignals`-2.6.2, `2HDMC`-1.8.0

[3] At time of writing Dr Semlali is a research fellow as part of NExT and can be reached via souad.semlali@soton.ac.uk

[4] Note that while the model itself contains 'NLO' in the name it was only used for tree-level calculations. The model simply allowed access to the α & β parameters needed for our work.



5.3 Allowed regions of $\cos(\beta - \alpha)$ & $\tan\beta$ parameters in 2HDM models Type-I and Type-II, on the left and right respectively for the observations made by ATLAS. These are obtained by comparing the observed couplings of the 125 GeV boson, when taken to be the light, CP-even Higgs boson, h of the 2HDM. Constraints are seen to be tighter on the Type-II model than the Type-I in this space. Plots are taken directly from [142]

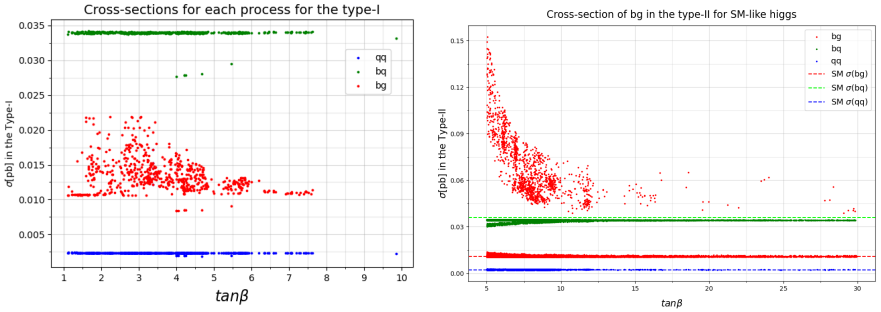


5.4 Allowed regions of $\cos(\beta - \alpha)$ & $\tan\beta$ parameters in 2HDM models Type-I and Type-II, on the left and right respectively given by CMS observations. These are obtained by comparing the observed couplings of the 125 GeV boson, when taken to be the light, CP-even Higgs boson, h of the 2HDM. Constraints are seen to be tighter on the Type-II model than the Type-I in this space. Plots are taken directly from [143].

It has been shown that for a wrong-sign solution we require that, $\sin(\beta - \alpha) > 0$, [98], as the available parameter space for a negative value has been ruled out. This occurred in-between the previous investigation and this one, so it is interesting to see how much our parameter space will be restricted by. As before we use the lower bound for the charged Higgs mass in the type-II of, $m_{H^\pm} > 580\ GeV$, found in Ref. [28]. This bound does not strictly apply to the type-I, however as we wished to remove the Feynman diagrams with a virtual charged Higgs this seemed a suitable bound to apply there as well. The bounds here are closely tied to the value of $\tan(\beta)$ we consider them at. A general limit on 2HDMs with Z_2 symmetry is given in Ref. [130] as, $\tan(\beta) \geq 1$, this is the limit used on the type-I model. For the type-II there is a stronger restriction on $\tan(\beta)$ from the mass of the charged Higgs itself, this was shown in Ref. [28] (c.f. figure 4, a plot of the, $m_{H^\pm}$ - $\tan(\beta)$). A lower limit for the type-II was chosen as, $\tan(\beta) \geq 5$.

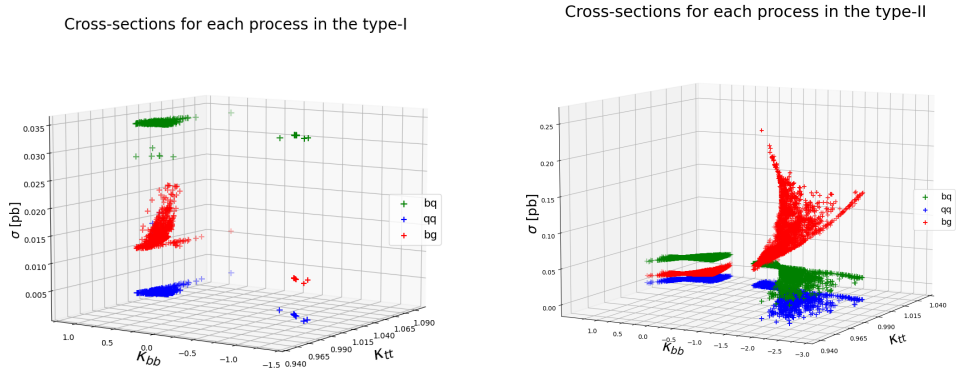
5.1.3 Initial Output

Figure 5.5 shows the initial results output by MadGraph for our points.



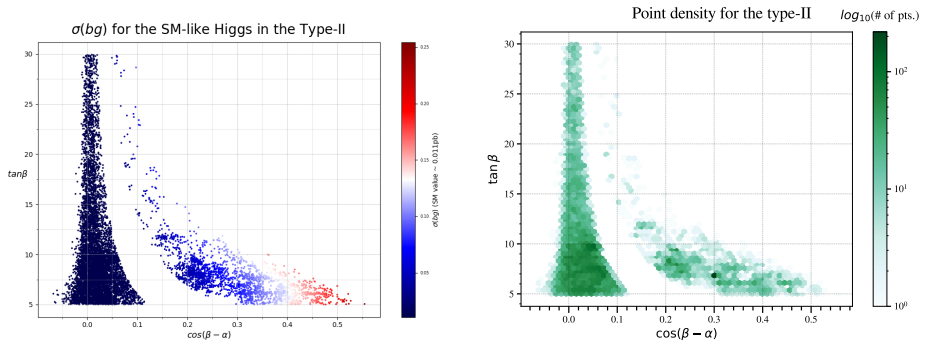
5.5 Cross-sections of points obtained in our scans of the parameter space for Types I and II, plotted against the value of $\tan(\beta)$. Note that these two plots are **not** to the same scale as the highest cross-sections in the Type-II are considerably larger than in the Type-I. For an idea of the difference at a glance, the green (bq) and blue (qq) points roughly occupy a line at the SM value for their respective processes.

We can see in Figure 5.5 that the Type-I points appear to behave much like the SM predicts, with some points having a lower cross-section even. While there is a slight increase in the *bg* process this does not seem to show anything which would be distinguishable from the SM at the LHC. The Type-II, however, behaves very differently. While the *bq* and *qq* processes for the Type-II do behave like the SM, showing similar sizes and the expected hierarchy between themselves; the *bg* process does not adhere to this. Instead we have the *bg* process becoming dominant over the expected leading order *bq*, in some places this is by a large margin. The highest point in the Type-II has a cross-section for *bg* that is over twice the size of the *bq*, presenting us with some candidates to study further in the hopes of them surviving selection at the LHC should such a signal exist.



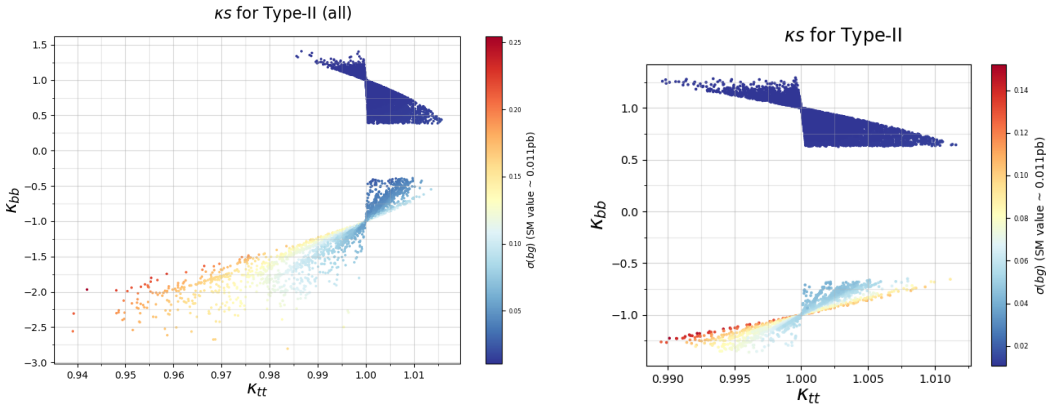
5.6 Cross-sections found for points in the Type I and II models for the SM-like Higgs, shown plotted against the κ_{bb} and κ_{tt}. We can see that, as in Figure 5.5, for Type I the wrong-sign solution does **not** result in a higher cross-section for bq or qq, however there is a meagre increase in the bg process. In the Type II it appears that both wrong-sign and alignment points offer an increase in the bg cross-section. We can see multiple cross-sections for the bg process with considerably larger cross-section than that of the leading process (bq) in the SM, for both alignment and wrong-sign solutions.

Looking now at Figure 5.6 we establish that the points found for the Type-I do not display a significant variation in cross-section size between the alignment and wrong-sign for the processes, $bq \to tqh$, and, $qq \to tbh$. There is a clear boost in the bg cross-section, but this occurs in the alignment region and is not as significant as the boost we see in the Type-II. Turning to the plot for the Type-II we have a large number of wrong-sign solution points for which the bg cross-section dwarfs the SM value of 0.011pb and, additionally, the SM value of the leading process (0.036pb). In addition there are alignment solutions with notable cross-section, comparable to those of the wrong-sign solutions. we will examine both these and the high points for the wrong-sign solution.



5.7 The figure on the LHS shows our points by $\cos(\beta - \alpha)$ and $\tan(\beta)$ value, with colour assigned according to cross-section size for, $bg \to h$. We see the highest points occur at medium tan values, and are concentrated around high cos values. The figure on the RHS is a hexabin plot, the darker the colour the greater the number of points found in that region. Again we see the points are mainly in medium values of tan.

In Figure 5.7 we examine the $\cos(\beta-\alpha)$, $\tan(\beta)$ plane. In the LHS plot the colour of the point indicates how large the cross-section found for it is, with the associated colour-bar for scale. It is important to remember when looking at it that the SM value is 0.011pb, i.e. only the very darkest points correspond to this, the vast majority are of greater magnitude to some degree. Turning to the RHS the colour now indicates the number of points that have been binned into each coloured hex. The scale is logarithmic and shown on the right of the plot. We see that medium values of tan are favoured, in particular between 5 and 10. There is also a very concentrated region of points around, $\tan(\beta)$ 9, $0 \leq \cos(\beta-\alpha) \leq 0.07$.

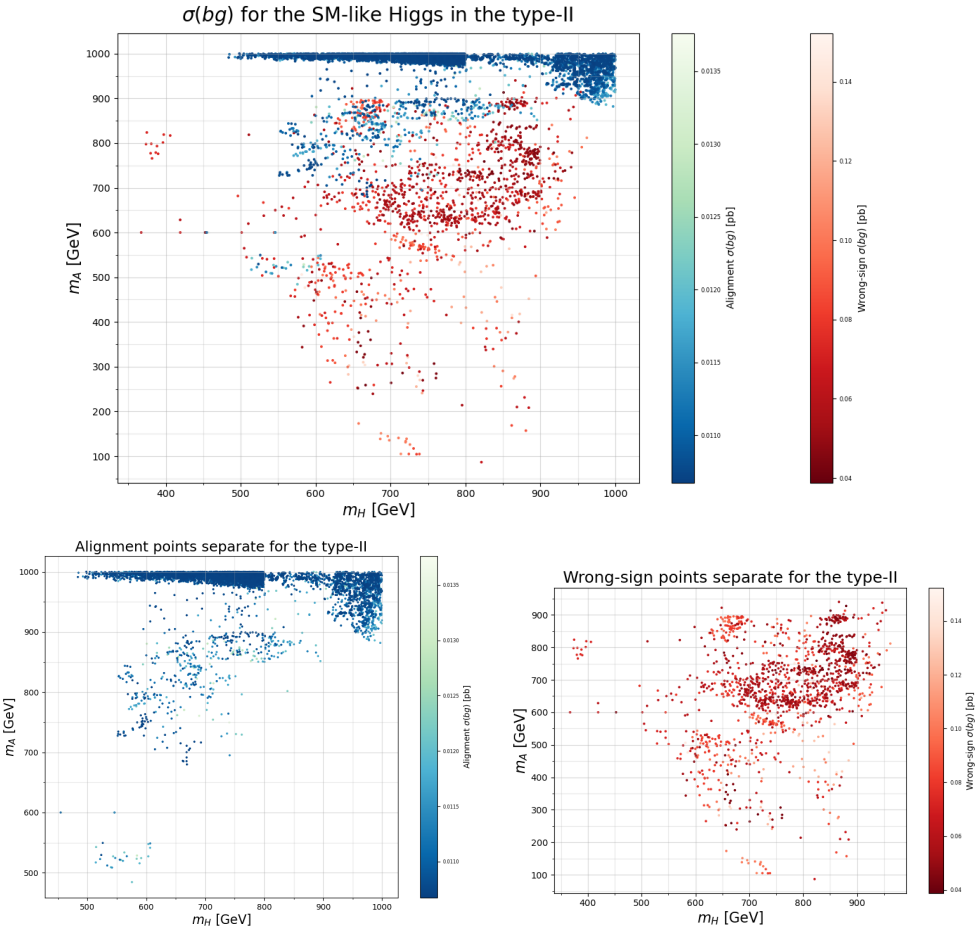


5.8 Cross-sections of points plotted in the, (κ_{tt}-κ_{bb}) plane, the colour-bar on the right indicates size of cross-section. The LHS plot shows all points found in the scan, the RHS shows only those points that had, $\chi^2 \leq 120$; that being roughly within 6σ of the median value of χ^2_{Tot} as defined in Equation 3.3

In Figure 5.8 we plot the points in the space of κ_{bb} and κ_{tt}, again we use a colour gradient to indicate the cross-section of each point. In the LHS plot we can see that the vast majority of points are SM-like in κ_{tt}, though there is a small cluster of 'doubly wrong-sign' points on the left side of the plot. Here 'doubly wrong-sign' refers to points where the κ_{bb} and κ_{tt} are point wrong-sign. The highest points are found when κ_{bb} is SM-like, we zoom in on this region the the RHS plot.

Something that has become clear in checking constraints and studying the resulting plots of our data is that while κ_{tt} is constrained into quite a small region; the same cannot be said of κ_{bb}. Points for κ_{bb} were found roughly in the region, $-3 \leq \kappa_{bb} \leq 1.5$. The highest alignment points were around, 0.75, and the highest wrong-sign ones, -1.

In Figure 5.9 we examine the, (m_A,m_H), plane. The LHS shows a large number of the highest cross-sections clustered in a region of high mass for both particles, though it also shows surprisingly low cross-section for many points with a high mass for the A and medium-high mass for the H. The RHS plot helps us to make some sense of this, these mysterious points are all alignment solution points. There seems to be a great deal more variation in the cross-section of alignment points in the 2HDM than the wrong-sign points; which consistently sit at a higher than SM cross-section.



5.9 Plots on the (m_H-m_A) plane. The top plot is split into two colour palettes, the blue-green is only alignment points, while the red-orange are wrong-sign. Both sets are coloured in a gradient shown in their respective colour bars. The lower plots are coloured according to the size of the, $bg \to tWh$, cross-section, as indicated by the associated colour bars. We see the high values are quite spread out throughout the region along with the lower values.

5.2 Analysis

The next step was to investigate the possibility of detecting a 'wrong-sign solution' cross-section at the HL(High Luminosity)-LHC for our bg sub-process (i.e. the production of a SM-like h in association with a single-top and a $W^{\pm}$). This required an in-depth MC analysis, simulating the decay of the h into a, $b\bar{b}$, pair in order to show that this phenomenology would be easily seen at the detector level. Specifically, we sought to determine if the 'wrong-sign solution' leads to any difference in the mass, p_T and η distributions for the final state particles when compared to SM configuration. We pursued this from a theoretical perspective; considering the main backgrounds to our process, applying a cutflow and calculating the signal to background significances.[4]

[4]Full detector level analysis and detailed background study is beyond the scope of this work.

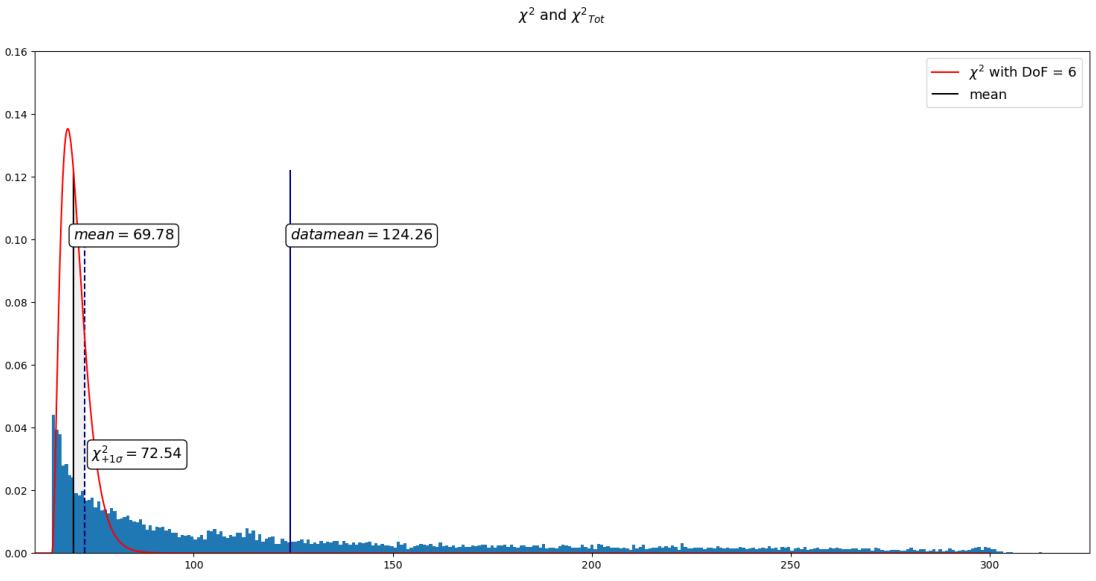
5.2.1 Simulation Details

Event samples were generated with, $\sqrt{s} = 13.6$ TeV, an integrated luminosity of, 3000 fb^{-1}, which is expected to be achievable at the HL-LHC in the near future. The SM events were generated along with the events for a Benchmark Point (BP) in the 2HDM type-II. Here we assume that the lighter, CP-even Higgs is the SM-like Higgs with, $m_h = 125$ GeV. The BP was chosen from the points generated by `Magellan`, we naturally sought a point with a large cross-section that would maximise our ability to detect a distinctive signature compared to the SM; however it was important to consider how likely those points were too. `Magellan` stores the χ^2 values for each point which gave us a way of determining whether a given point fits the real data well. While there were numerous points with cross-sections $> 0.2pb$, these all had χ^2_{Tot} values around, 200-300.

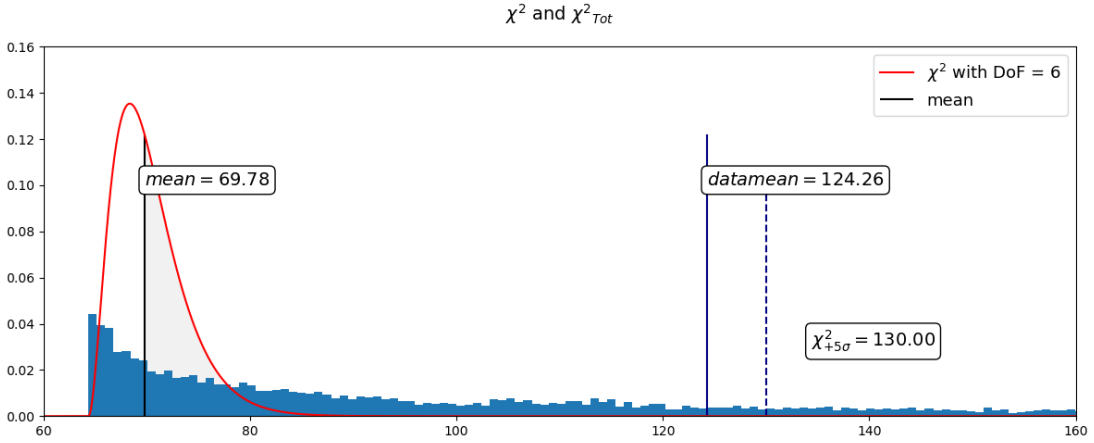
The highest point to fall within 6σ of the expected mean for a χ^2 distribution with 6 degrees of freedom (made of the degrees of freedom from our parameter space minus 1, 5, plus the degrees of freedom from our S-T fit minus 1, 1), shifted to start at the lowest χ^2_{HS} in order to take the errors involved in their calculation into account. This led us to a benchmark point that fell within 5σ of the mean that is shown below:

Having been generated by `Magellan`, this BP has passed checks on theoretical and experimental constraints as described in chapter 3.

The focus was on events containing, $h(SM) \to b\bar{b}$, decays along with the associated (anti)top quark decaying into leptons + b-jet and with the W boson in the subprocess



5.10 A histogram in dark blue showing the values of $\chi^2{}_{Tot}$ for the points found by `Magellan`, with bin height normalised such that the total adds to 1. In red is a standard $\chi^2{}_{Tot}$ with 6 d.o.f., shifted to start at the minimum $\chi^2{}_{Tot}$ value. Mean indicates the line along the mean of the shifted χ^2 distribution, the dashed line indicates this mean with 1σ added, the filled region in-between highlights the area within 1σ. 'datamean' is the mean value of the points in the `Magellan` scan



5.11 A histogram in dark blue showing the values of $\chi^2{}_Tot$ for the points found by `Magellan`, with bin height normalised so that the total adds to 1. In red is a standard $\chi^2{}_{Tot}$ with 6 d.o.f., shifted to start at the minimum $\chi^2{}_{Tot}$ value. Mean indicates the line along the mean of the shifted χ^2 distribution, the dashed line indicates this mean with 5σ added, the filled region in-between highlights the area within 5σ (note that much of this is not visible at this scale). 'datamean' is the mean value of the points in the `Magellan` scan

m_h (GeV)	m_H (GeV)	m_A (GeV)	$\tan\beta$	$\sin(\beta-\alpha)$	m_{12}^2 (GeV^2)	$\chi^2{}_{Tot}$	σ(pb)
125	781	701.594	5.06049	0.906849	113436	119.028	0.15228

5.2 Key values for the 2HDM Type-II Benchmark point, with the cross-section being that to LO of the bg process.

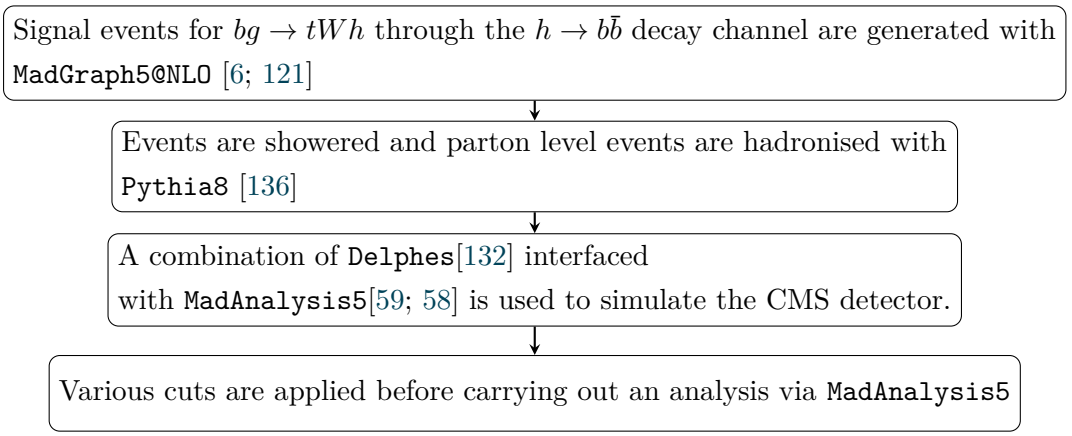
decaying leptonically. The Parton Distribution Function (PDF) we used was, `NNPDF23_lo_as_0130_qed`. Figure 5.12 depicts the methods and tools used to carry out our MC simulation.

The input parameters and production cross-section (to Leading Order (LO)) for our chosen benchmark point and process in the 2HDM are shown in Table 5.2. The corresponding SM cross-section value at LO level is 0.000187 pb. The settings for the tools were left at the default value beyond the necessary input of model parameters[5].

BR($h \to b\bar{b}$)	σ(pb)
6.85754×10^{-1}	0.00244

5.3 Values for the 2HDM Type-II BP cross-section to LO for the, $h \to b\bar{b}$, decay channel of the bg process.

[5] i.e. `Pythia8` is set to use 'simple' showers, the name being a comparison to some of the other options within `Pythia8`. The exact nature of these is quite involved hence we direct the reader to the `Pythia8` manual [136]



5.12 Diagram showing the combination and sequence of physics software used in the generation and analysis MC events for both the SM and our BP in the 2HDM.

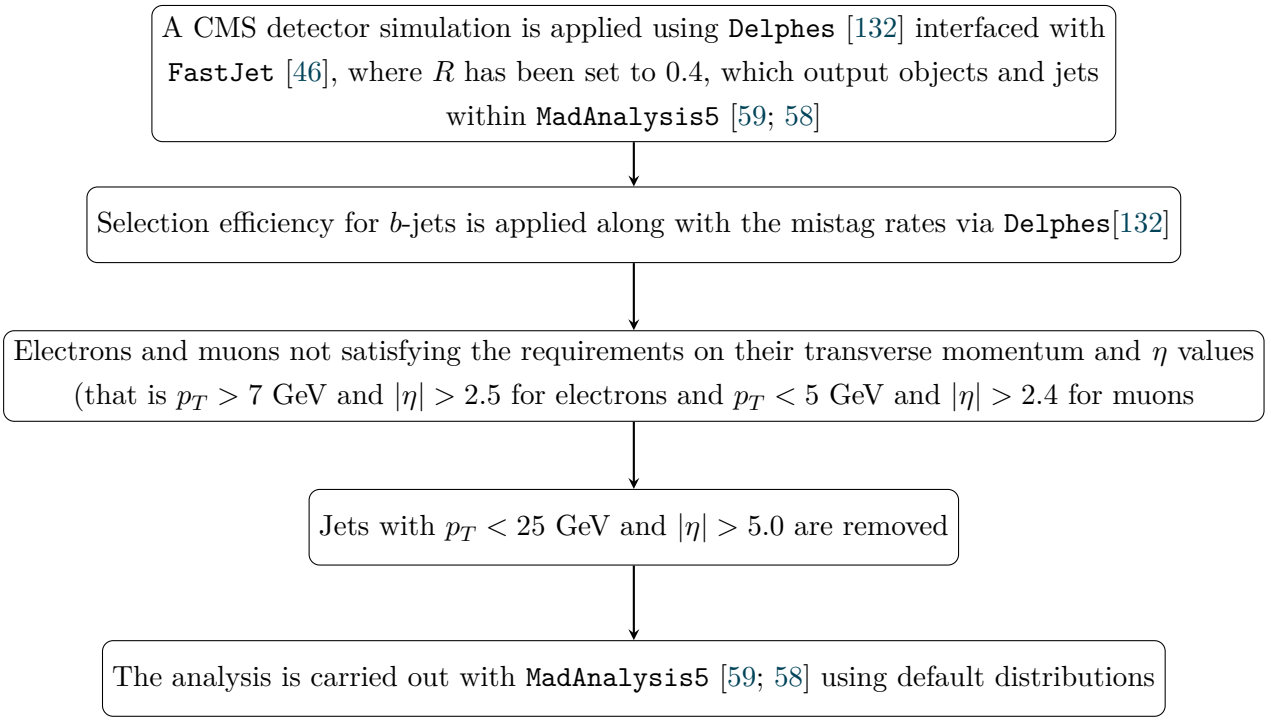
5.2.2 Cutflow

Various algorithms are applied by the CMS collaboration to identify individual particles and jets in the detector and then reconstruct them, we attempt to replicate this in our MC simulation in order to get the most accurate representation. There are few papers that look at $pp \to th$ processes, let alone $bg \to tWh$ by itself, due to the $t\bar{t}h$ washing it out. We thus follow the treatment of th processes as described in [55] to mimic event reconstruction and jet clustering. We do so with the *loose* tagging selection efficiency.

For the background, we considered the following SM processes:

- $gg, q\bar{q} \to t\bar{t}$,
- $gg, q\bar{q} \to t\bar{t}h$,
- $gg, q\bar{q} \to t\bar{t}b\bar{b}$,
- $gg, q\bar{q} \to t\bar{t}t\bar{t}$,
- $q\bar{q} \to W^+W^-h$, $q\bar{q} \to ZZh$,
- $q\bar{q} \to ZW^+W^-$.

Figure 5.13 gives an overview of the initial cutflow.



5.13 Flowchart showing the initial procedure used for event reconstruction and jet clustering.

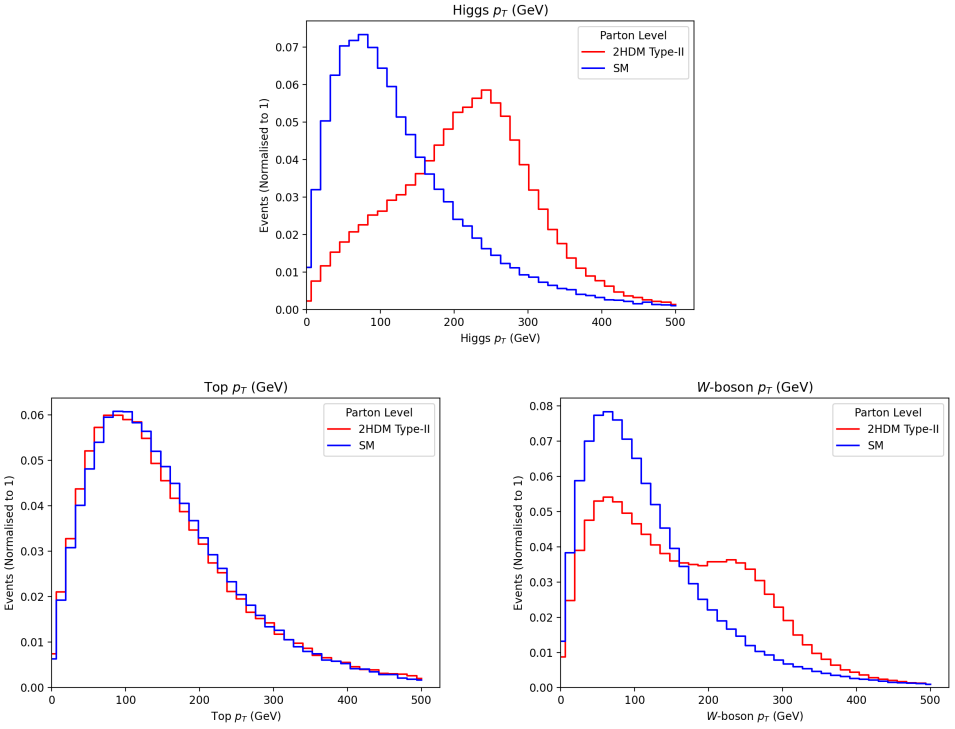
- In order to match [56] we require that in our simulation, all our reconstructed electrons meet the requirement $p_T > 7$ GeV and $|\eta| < 2.5$ to.
- Reconstructed muons were required to have $p_T > 5$ GeV and $|\eta| < 2.4$, naturally the difference between these requirements and those for the electron is down to the increased mass.
- Jets were clustered using the 'anti-k_t' algorithm[45], this is somewhat similar to an idealised cone algorithm. A constant value for R of $R = 0.4$ is used for the cone size in the case of each jet. To be considered in our analysis the jets are then required to have $p_T > 25$ GeV and $|\eta| < 5.0$. b-jets are considered to be those with $|\eta| < 2.4$ [55]. In addition they may not be allowed overlap with any electron, muon or hadronic τ.
- The 'loose b-jet selection efficiency' we adopt from [55] is 84%.
- Mistag rates for c-jets and light quarks are 1.1 and 11% respectively; in line with [55].[6]

[6]The difference between these mistagging rates may appear somewhat illogical, with the c-jets having a rate a factor of ten lower than that of the light-quarks; this largely comes as a result of the high rate for light quark production combined with the effectiveness of the tagging algorithm for the c-jets specifically.

5.2.3 Analyses

We perform our analysis on two different levels, first we examine the parton level and then we zoom in further to the hadron level. As intimated in Figure 5.12, we applied numerous cuts to the MC events to both reduce the backgrounds and hone in on the hypothetical signal in addition to those that aim to make our simulation more realistic. The specific cuts we applied are as follows;

5.2.3.1 Parton Level Results



5.14 The plots above show the normalised transverse momenta distributions for our BP in the 2HDM type-II (red) along with those of the SM (blue). The top plot shows the distribution for the SM-like Higgs boson. The bottom left is of the top quark at parton level while the RHS plot is the distribution for the W boson, again, at parton level.

For the parton level analysis we compare the kinematic distributions for the 2HDM type-II to that of the SM. In Figure 5.14 there are several plots showing the p_T distributions for different particles in both models.

Looking to Figure 5.14 there is a clear difference between the transverse momentum distributions for the most massive objects in the final state; the h, t and W in the two models. This arises due to the relative kinematics of these models. We can see that the signal enhancement for the 2HDM type-II has translated into a distribution shift to the right, i.e. with higher average transverse momentum for both the h and W bosons. The W boson distribution has also been warped, almost displaying dual peaks. We

have phenemonology displayed in these two diagrams that shows the two different possible solutions of the type-II 2HDM, right and wrong-sign. The red line in the top plot seems to be an amalgamation of our high pT wrong-sign scenario and the alignment points which peak in roughly the same spot as seen in the blue, SM line. Together they form this shifted and somewhat indistinct peak. Because the wrong-sign scenario creates addition between previously cancelling Feynman diagrams we have additional kinematics contributing to the shape. It is perhaps more obvious that this is what is occuring when we look at the bottom right plot of Figure 5.14. There are two very clear bumps in the red 2HDM plot that clearly look like the result of two peaks being added together. One of these peaks is similar to the SM peak, while the other, much like the Higgs peak, has been shifted upwards to a higher pT value.

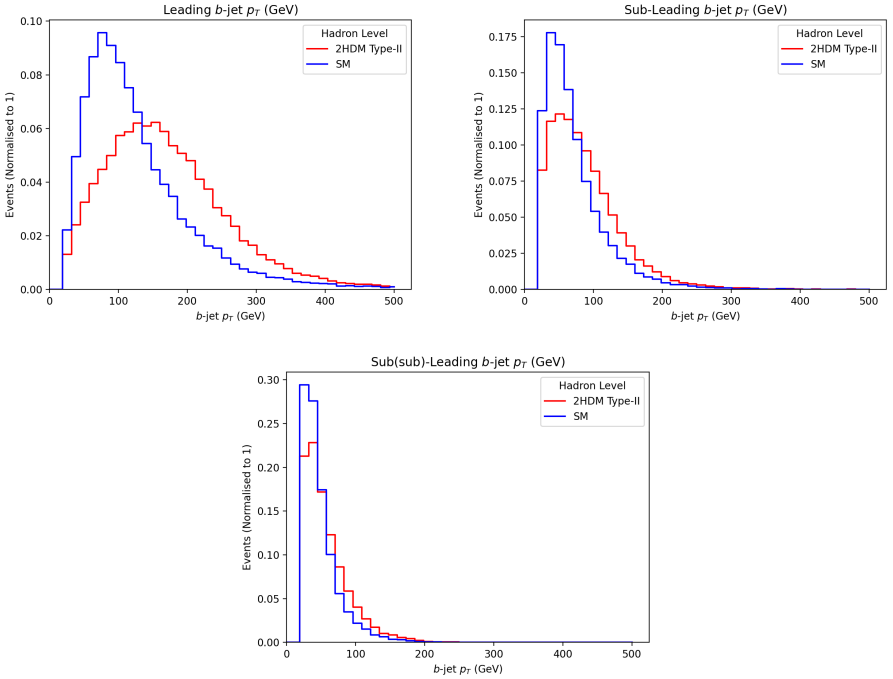
When looking to the plot concerning the top quark there is little change to the p_T distribution. The differences seen here can be pursued at the hadron level with the goal of characterising a signal as being BSM in origin, providing additional evidence beyond previously noted difference in the integrated cross-section yields of the two scenarios. Bearing this in mind, we will only assess the (identically) normalised distributions to extract the shapes.

Given the clear difference in the W distributions we would expect that, in the case of prompt W bosons, the p_T would transfer efficiently into the lepton spectra at the detector level; allowing them to potentially be identified as separate from the leptons emerging from leptonic decays of the (anti)top via (secondary) $W^{\pm}$'s.

Clearly these differences amplify the potential of our results; this increase in cross-section for the process when used with 2HDM type-II framework can lead to some interesting physics with which to probe the LHC for the tH process. In order to solidify the evidence of our study, we will look at the hadron level analysis, with a view to investigate the kinematic distributions of the final state particles.

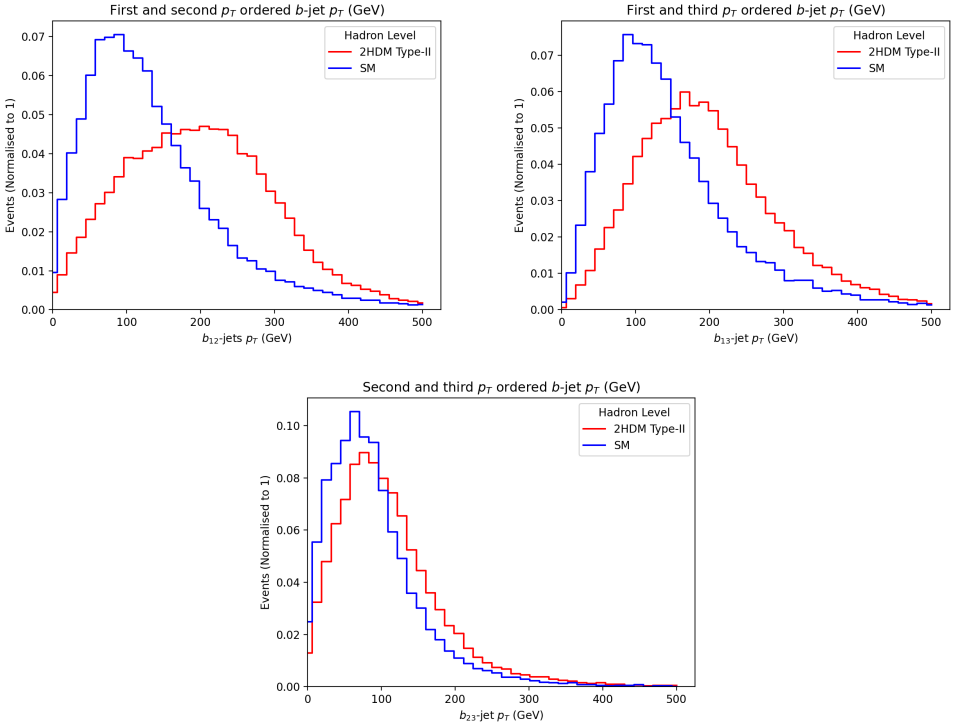
5.2.3.2 Hadron Level Analysis

In this section we consider a hadron level analysis of the events, using hadronised parton showers that have been run through detector simulation and clustered into reconstructed particles and jets.

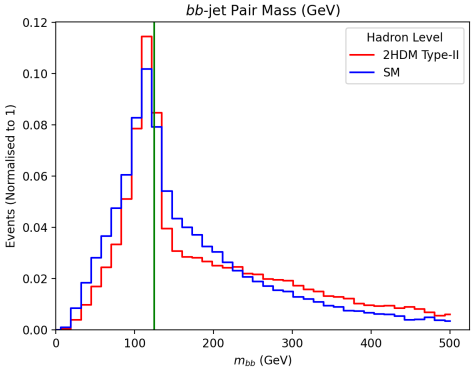


5.15 All three plots depict the normalised transverse momentum distributions of b-jets for both the SM and 2HDM type-II. The top plots are of the leading and sub-leading b-jets. On the bottom we have the transverse momentum of sub to sub leading b-jet.

We start by comparing the kinematic distributions and resonances for 2HDM type-II model and SM. The p_T ordered b-jets transverse momentum distributions are shown in Figure 5.15, we expect three of these jets in total. The kinematic differences have propagated through to the distributions with a clear widening and squashing down of the distributions in the top two plots. This effect can be seen in the lowest plot too but it is much reduced by comparison. It is apparent that the leading jet in particular is considerably harder than its' equivalent in the SM, this is a result of the large discrepancy in p_T that we saw in Figure 5.14 between the BSM scenario and the SM for the h itself. the b-(anti)quarks produced in h decays



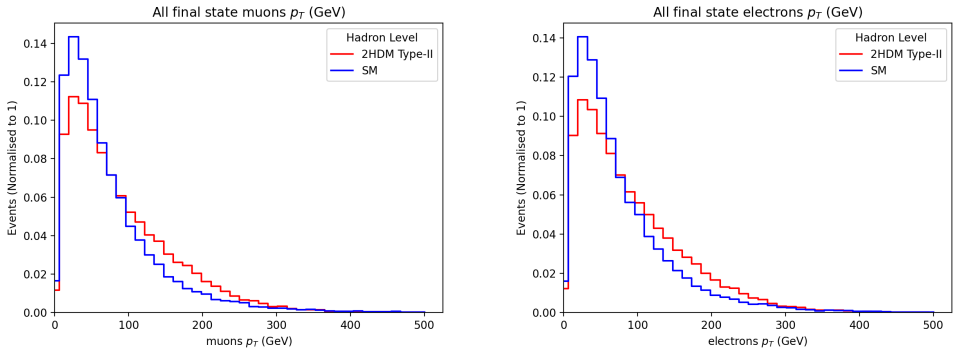
5.16 Upper panel: Combined Transverse momentum distribution of leading and sub-leading b-jets and combined Transverse momentum distribution of leading and sub to sub-leading b-jets. Lower panel: Combined Transverse momentum distribution of sub-leading and sub to sub-leading b-jets.



5.17 The b-dijet invariant masses for both 2HDM type-II and SM.

Next, we look at the invariant mass of all pairs of clustered b-jets, m_{bb}, in order to reconstruct and identify the SM-like Higgs mass resonance. Remarkably, from Figure 5.17, we can see that in the 2HDM type-II framework, the mass reconstruction is somewhat sharper than in the SM. This indicates that the effect of the combinatorics is milder in the former case than in the latter. However, both

distributions are somewhat misaligned w.r.t. the true MC value of the corresponding Higgs boson resonance.



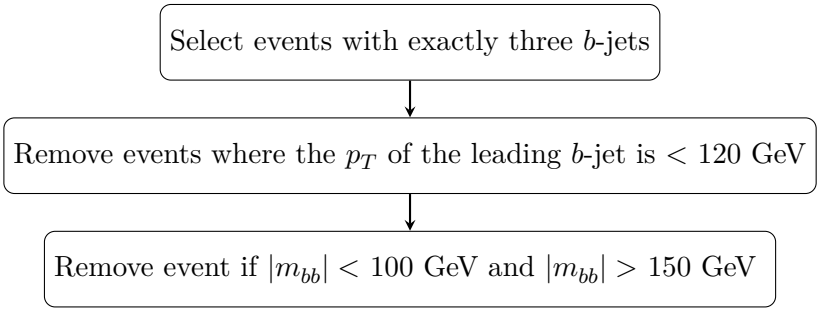
5.18 Left panel: Transverse momentum distribution of all reconstructed muons. Right panel: Transverse momentum distribution of all reconstructed electrons.

Finally, we look at the transverse momentum distributions for reconstructed muons and electrons. The p_T distributions for muon and electron are shown in Figure 5.18. There is a small change in the kinematic distributions for both models.

As we can see that there is an enhancement in the signal properties at the hadron level; we consider leading backgrounds and calculate the signal-to-backgrounds rates to provide a conclusive reasoning for the above mentioned results.

5.2.3.3 Signal To Background Analysis

As a final exercise, we consider the three leading backgrounds $pp \to t\bar{t}$, $pp \to t\bar{t}h$ and $pp \to ZWW$ to calculate signal-to-background significance rates for both the models. We apply a selection criteria to our events samples described in Figure 5.19.



5.19 Event selection used before computing the signal-to-background rates.

Next, we calculate the event rates to get significance rates for one value of integrated luminosity $\mathcal{L} = 200$ fb^{-1}:

$$N = \sigma \times \mathcal{L}, \tag{5.1}$$

The event rates for signal (both models) and backgrounds are given in Table 5.4. The significance, Σ, is then calculated and is given (as a function of signal S and respective

background B rates) by:

$$\Sigma = \frac{N(S)}{\sqrt{N(B_{t\bar{t}}) + N(B_{t\bar{t}h}) + N(B_{zww})}}, \tag{5.2}$$

Process	
bg (2HDM)	279.624
bg (SM)	7.663
$pp \to t\bar{t}$	9252.179
$pp \to t\bar{t}h$	110.127
$pp \to t\bar{t}b\bar{b}$	1013.760
$pp \to t\bar{t}t\bar{t}$	0.182
$pp \to W^+W^-h$	0.530
$pp \to ZZh$	0.062
$pp \to ZW^+W^-$	0.221

5.4 Event rates of signal (in both models) and backgrounds for $\mathcal{L} = 3000$ fb^{-1} upon enforcing all cuts.

Upon inspection, Table 5.5 makes it clear that that the dominant background by far is $gg, q\bar{q} \to t\bar{t}$. This is followed by $gg, q\bar{q} \to t\bar{t}b\bar{b}$. $gg, q\bar{q} \to t\bar{t}h$ provides a small contribution but the remaining backgrounds, including $gg, q\bar{q} \to t\bar{t}t\bar{t}$, $q\bar{q} \to W^+W^-h$, $q\bar{q} \to ZZh$ and $q\bar{q} \to ZW^+W^-$ are negligible.

Process	2HDM	SM
$bg \to twh$	11.594	1.374
$pp \to t\bar{t}$	12297.601	12297.601
$pp \to t\bar{t}h$	55.046	55.046
$pp \to ZWW$	0.069	0.069

5.5 Event rates of signal and backgrounds for $\mathcal{L} = 200$ fb^{-1} upon enforcing the cuts.

Table 5.6 shows the final Σ values, the signal from the 2HDM type-II framework results in far better significance than the SM case. Should the 2HDM type-II be true then the HL-LHC would detect this long before it could detect the SM equivalent (should the SM prove the winning theory). As we have also seen large differences in partonic behaviours, which then manifest at detector level, it will potentially be possible to identify the hypothetical signal as being from the 2HDM type-II explicitly.

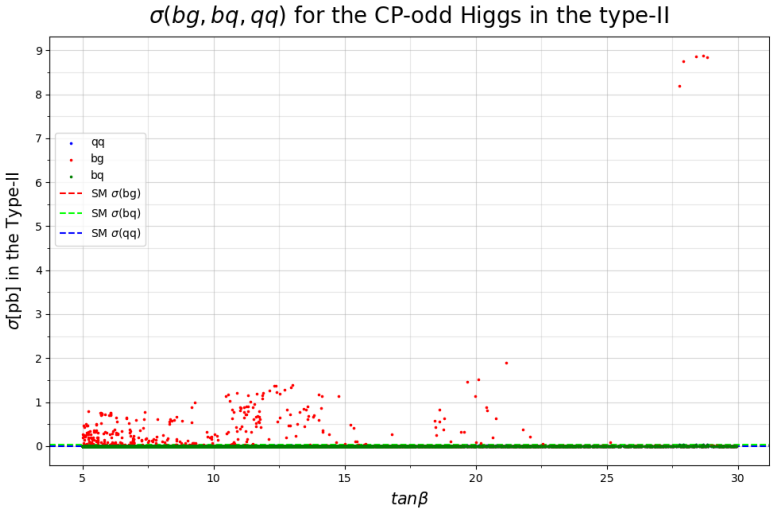
	2HDM	SM
$\mathcal{L} = 3000$ fb^{-1}	2.744	0.075

5.6 Final Σ values calculated for $\mathcal{L} = 3000$ fb^{-1} after enforcing all cuts.

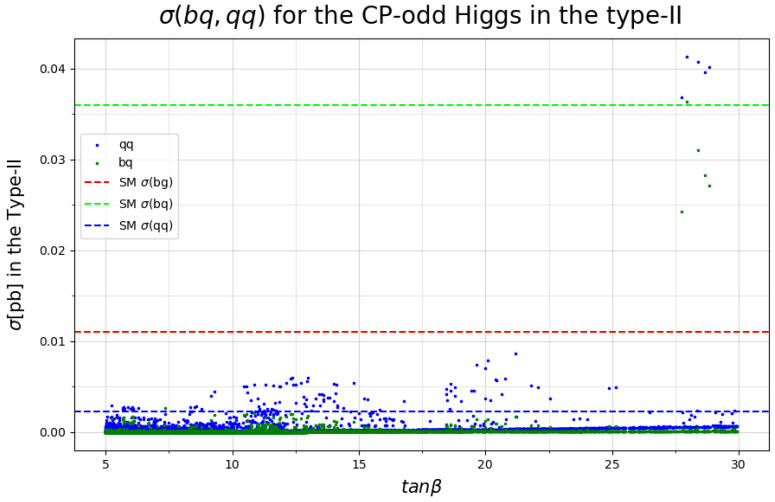
It is clear from Table 5.6 that signal within 2HDM type-II framework provides better significances compared to SM configuration. This means that signal with 2HDM type-II framework will be much more visible than the SM.

5.3 BSM Higgs Bosons

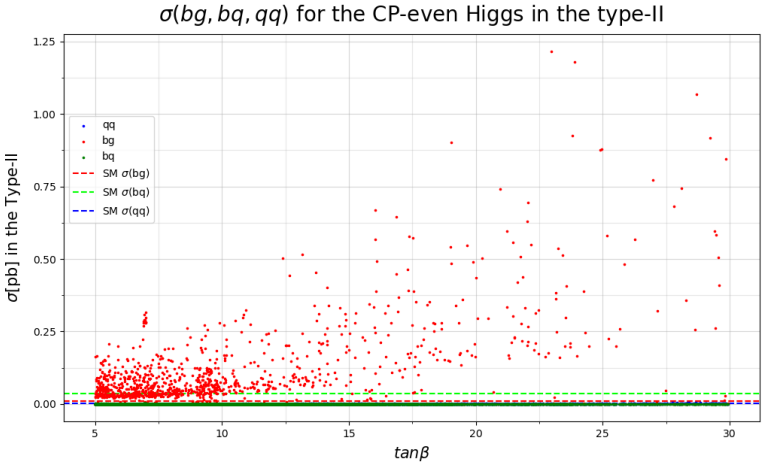
While an analysis of the signals for the BSM Higgs bosons in our model was deemed slightly beyond the scope as well as difficult due to the lack of pre-existing measurements/analyses relating to them; the cross-sections were still calculated for them. These cross-sections may warrant further investigation as signals that may be detectable in the near future, should the 2HDM type-II prove correct.



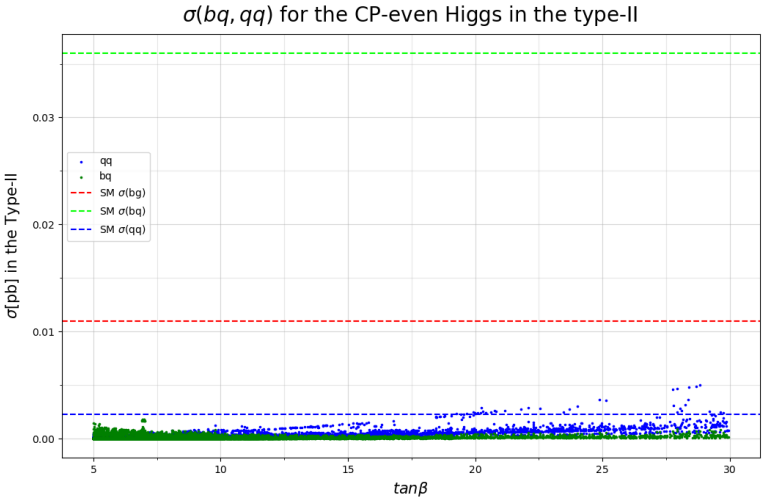
5.20 A plot showing the cross-section values for the CP-odd Higgs boson in the type-II model. Dashed lines are plotted to show the SM values for each process, however these are not visible due to the presence of much larger cross sections. We can see that there are far more points at lower values of $\tan\beta$, but that the bg process cross-section rises rapidly at the higher values of $\tan\beta$. Generally the bg values are orders of magnitude larger than the SM equivalent, further, it is considerably larger than the dominant process in the SM too. The size of the bg cross section makes it difficult to see what is happening to the cross-sections of the two other processes.



5.21 A plot of the cross-section values for the CP-odd Higgs boson in the type-II model. In this plot we remove the overwhelming presence of the bg process cross sections and instead focus on those of bq and qq. Dashed lines are plotted to show the SM values for each process for comparative purposes. The values for qq do not appear to vary notably with $\tan\beta$, however the bq display a handful of points at high $\tan\beta$ with markedly increased cross-section. The points are almost all below the cross section value for the SM sub-leading process, bg. Interestingly the SM dominant bq process is now consistently the smallest of the cross sections, with the majority of values being lower than that of the smallest SM process, qq. Clearly these are lower than both the bg predicted for our type-II model *and* and the expected cross-section for the leading SM process (bar a handful of points which are fractionally higher than this); thus we would not expect this to be detected before either of these signals.



5.22 Above is a plot of the cross-section values for the additional CP-even Higgs boson in the type-II model. Dashed lines are plotted to show the SM values for each process, however these almost invisible due to the presence of much larger cross sections. Akin to the case of the CP-odd Higgs in Figure 5.20 the bg process cross-section rises rapidly for the highest values of $\tan\beta$. There is a more consistent, gradual increase in this case also, at lower $\tan\beta$ values. The highest cross-section we see only reaches around a quarter of the size of the highest CP-odd Higgs bg cross section, and yet this is still considerably above the SM equivalent as well as the SM dominant process. The size of this cross section makes it difficult to see what is happening to the cross-sections of the two other processes.



5.23 Plot of the cross-section values for the CP-even Higgs boson in the type-II model. In this plot we remove the much larger bg process cross sections and instead focus on those of bq and qq. Dashed lines are plotted to show the SM values for each process for comparative purposes. The values for qq show a clear positively correlated relation to the value of $\tan\beta$, and past around, $\tan\beta = 10$, we see that the qq completely supersedes the bq process that dominants in the SM. We see that for the majority of points, both the qq and bq processes give a cross section that is lower than the SM equivalent. In particular the bq cross sections are around a tenth of the SM values. Clearly these two processes would not be fruitful to follow up.

As we can see in Figures 5.20 and 5.22 the bg process has a much greater magnitude in the 2HDM type-II for the neutral Higgs bosons when compared to the SM equivalent. This is in contrast to the cross sections found for the type-I, which was not found to produce significantly larger cross sections. Of particular interest is the presence of large cross sections for both the 'wrong-sign solution' (of the bottom (anti)quark) and the 'alignment limit'. These cross sections are considerably larger than the equivalent bg sub-process in the SM. Furthermore, there are potentially very, very large signals to be found for the 2HDM H and A Higgs particles. However, when considering these cross sections, there are a great deal more unknowns compared to the SM-like Higgs cross section. We expect the SM-like Higgs to behave as, just that, SM-like. This means we have strong confidence in the mass we expect it to have, on the couplings to many particles[7] and even where there is uncertainly in such values it is highly constrained. For the additional neutral Higgs particles introduced by the 2HDM we do not have this same luxury. We are not sure of the masses or couplings we might see for these particles as they are entirely new should they be discovered. While we may have values for many of these parameters coming purely from the theory side, we do not have pinpoint accuracy. There will be a range into which each of these falls in reality, (if they should exist), compared to the precise values the scanner has identified.

Further to this we must consider the branching ratios for each Higgs in these scenarios, where they are more massive than the SM Higgs we should see these additional Higgs particles decaying into the SM Higgs; this needs to be accounted for in the current measurements of Higgs particles to ensure the model fits the data. The branching ratios would also be altered from the SM-like Higgs to the quarks due to the change in coupling constants from the SM to the 2HDM, this would affect what would be expected at the detectors of the LHC again, which must be separated from the background. Finally, the additional Higgs bosons would also decay into numerous particles that would add to the detected levels, this must be accounted for within previous data and new cuts and backgrounds would need to be devised.

Clearly the analysis we performed here is intended to be preliminary, aiming to highlight the potential of the bg sub-process triggering 'SM-like Higgs boson in association with a single-top' as a test for a possible non-standard nature of EWSB. A test that serves this purpose far better than the alternative two channels, (bq and qq), can provide, as they have production cross sections that essentially do not differ between the two theoretical scenarios. Thus far the only approaches that have been used in pursuit of this signature have been inclusive, i.e., one where all three of the sub-processes are captured simultaneously. This approach does not appear to be optimal, at least if one wishes to test for a 'wrong-sign solution' of the 2HDM type-II; which, we have seen, survives the current experimental scrutiny of the SM-like Higgs boson signals thus far.

[7] Clearly we do not include the couplings we are actively investigating here, rather to other SM particles such as the weak bosons, leptons etc.

We also note that recent improvements in jet clustering algorithms provide a potential route to detecting a clear BSM signal at the LHC. Our colleagues recently published two papers regarding this, both taking the 2HDM type-II as an initial case study. The first, considered the decay channel, $gg, q\bar{q} \to H \to hh$, eventually decaying further to a four b-jet signature; it was found that alteration to the standard p_T cuts and the introduction of a variable-R algorithm had a marked impact on the efficiency of particle reconstruction [50]. The second compared different jet-clustering algorithms in establishing fully hadronic final states, this is done in the context of a 2HDM type-II once again and the decay chain in question is a heavy Higgs state into a pair of the SM-like Higgs boson. Fat b-jets are used to reconstruct the lighter Higgs particles and a double b-tagging applied. In addition, two jet clustering algorithms are compared, one with variable-R and one with fixed R. It is found that the variable-R is notably more efficient [51].

These improvements may allow us to detect far more of the potential signals we have found for the parameter space points in the 2HDM type-II, as a lower cross section should become detectable with these techniques. This would allow us to consider some of the points with far better ${\chi^2}_{Tot}$ values than our BP. For example, the following point has a predicted cross section still considerably greater than the SM dominant process:

m_h (GeV)	m_H (GeV)	m_A (GeV)	$\tan\beta$	$\sin(\beta-\alpha)$	m_{12}^2 (GeV^2)	${\chi^2}_{Tot}$	σ(pb)
125	723.3	397.119	5.23348	0.928653	94658.9	74.779	0.116438293

5.7 Key values for the 2HDM Type-II possible *future* BP, with the cross-section being that to LO of the bg process. Comparing to Table 5.2 we have a point with a considerably lower value for ${\chi^2}_{Tot}$, in fact it is within 2σ C.L. (51.5067 - 77.2045)

but this cross section may not be large enough to be detected with current methods. Perhaps a future investigation could be carried out into the potential detectability, using these improvements, of the points found here with `Magellan`.

Chapter 6

Summary And Outlook

This thesis can largely be split into two subtopics; the development and running of `Magellan` and the investigation into detectable 2HDM signals at the LHC. As such, we break the conclusions down into these.

6.1 The Future Of Magellan

The intention moving forward would be to first upgrade the `Python` used to `Python` 3. A clear benefit to this would be losing the reliance on depreciated code, many of the modules used in `Magellan` are versions which are no longer supported due to the developers having moved on to `Python` 3 versions, this is far from ideal. This would also make it easier for users who prefer `PIP` over `Anaconda`, some versions of the modules used are elusive to track down and it can be complicated to ensure these are all compatible[1]; as `Python` 3 is fully supported this is considerably less of an issue. Another very useful opportunity with the switch to `Python` 3 is that `MadGraph5@NLO` could be upgraded alongside it. The `MadGraph5@NLO` group no longer ensures compatibility with `Python` 2, this means that we are missing out on numerous `MadGraph5@NLO` improvements in using our current version, in particular there have been upgrades to the running speed of the software that ideally we would like to take advantage of.

Adopting `Python` 3 would involve either altering and upgrading `T3PS` to `Python` 3 so that it may remain as the MCMC engine, *or* choose a suitable replacement e.g. a modified version of `STAN`[137]. An alternative might be `BAT`[47], (the Bayesian Analysis Toolkit), either in the older C++ version or the more recent `BAT.jl`[131] which is written in `Julia` and is the current focus of the creators. Both versions have clear benefits, as such it may come down to the coding languages known by future `Magellan` developers.

[1] `Anaconda` helpfully takes care of this for the user, but with `PIP` it must be done themselves to ensure functionality.

It is also important to progress HiggsBounds, HiggsSignals and 2HDMC into the authors newer, replacement toolbox - HiggsTools[19]. Instead of having separate tools the authors have created a combined program for ease of use. We are uncertain what percentage of the older code has been used, but this is certainly something that would need to be investigated before deciding on which of the two available interfaces to use. There is both a Python (3) interface as well as a C++ one. Currently we would recommend using the C++ as this language is what is currently used to run all three of 2HDMC, HiggsBounds and HiggsSignals so it should be the simplest way to make the change.

Hypothetically it is possible to use Magellan for any theory with an extended Higgs sector with little alteration, or any model with a UFO file and alternative to HiggsBounds, HiggsSignals and 2HDMC for checking constraints and outputting a usable likelihood to T3PS. As is detailed in chapter 3, Magellan calls 2HDMC with HiggsBounds and HiggsSignals in order to calculate the likelihood for T3PS; it should be relatively straightforward to alter the 'ParameterPointProcessor" to remove 2HDMC from this script and rewrite so that the focus is on HiggsSignals. Doing this would cause us to lose the theoretical constraints provided by 2HDMC, of course, and so it would be best if an alternative method was integrated for for any supported models; this could be done using a pre-existing package or from scratch depending on future developers. But this would allow the user to explore a much greater variety of models and should be given due focus in development.

It should also be possible to add flavour physics constraints into the toolbox either using novel code or incorporating some existing software e.g. SuperIso[120]. SuperIso is written in a mixture of C and Fortran, which should be taken into account when determining whether or not to use it in Magellan. There is also a relatively new program, EOS [145], which uses a similar approach to Magellan in using an MCMC method in a mixture of Python and C++ , as a result it seems that it would have good compatibility. Other alternatives include;

- flavio[139], which is written entirely in
  python and includes likelihood calculations
- HEPfit [65], which is written in C++ and uses the BAT MCMC library mentioned previously, as such a likelihood is calculated. This is an obvious choice should BAT be used to replace T3PS.
- FlavBit [29], which is a module for GAMBIT[9] which itself is a global fitting tool with a
  python interface.

It is worth noting that while some of these physics tools mentioned above may contain some similar aspects to Magellan, the focus of them seems to be on finding the excluded regions without considering what the model *itself* excludes of that region.

`Magellan` does this by allowing the user to examine plots of the parameter space under a Bayesian statistical framework, running a high number of or very long (statistically there is no difference thanks to the Markov property) chains then plotting the allowed points provides a good approximation of what regions are likely and how likely they are with respect to others in the plot. The likelihood takes into account our constraints but also the physics of the model itself so we clearly see when a region that is not ruled out by exclusion limits is still not populated. Besides this, we also have the ability to apply a new analysis on just two parameters (say) onto our points in that same parameter space and then see how that impacts the rest of our parameters immediately. Again, the Bayesian framework allows us to make statements about what is excluded in these other parameters as a result of that initial analysis that has been applied.

When considering the MCMC functionality itself, regardless of whether a new program is adopted it may be worthwhile investigating if a cap on the maximum allowed χ^2 for point acceptance has an impact on the convergence speed, the rate at which mean acceptance ratio changes and the maximum that it reaches during a run. This would not be some absolute cut, but rather a modification on the current acceptance conditions. Rather than having an 'if' statement which checks if the likelihood increases and accepts if it has but if it has not, accepts based on a probability; we would suggest incorporate a probability on the acceptance of points with χ^2 above some cap in a similar fashion. This would hopefully push the algorithm further in the direction of higher likelihood generally rather than just relative to itself. This would need to be tested to see what effect it has in reality, and should probably be an optional feature that can be turned off. The cap on the χ^2 would, of course, vary depending on the degrees of freedom for the system and should be input by the user; as clearly the cap on a parameter space with 3 d.o.f. will be far lower than that of one with 10 d.o.f.[2].

The option of adding a heating function is already offered in the config file for MCMC runs, however it may be helpful if we could add some example templates, perhaps with some explanation around what each different heating function is suitable for. This goes hand and hand with another suggestion that we add basic tools for assessing how the MCMC is performing. This would simply involve tidying-up the code we used ourselves to examine various features such as the acceptance ratio, convergence etc.; and ensuring we add thorough commentary therein. A `Python` `Jupyter Notebook` would be useful for this purpose, allowing the user to quickly visualise difference aspects of diagnostic data. It should be possible to add coloured text indicating to the user whether the data indicates something positive or negative about their MCMC; e.g. for convergence we could have it indicate when convergence is achieved or whether or not it is actually achieved at all.

[2] Assuming both distributions are starting at the same point, which is dependent on χ^2 values coming from sources such as `HiggsBounds` and `HiggsSignals` in the examples we have seen.

It could also be beneficial to consider implementing more complex methods of tempering, such as Parallel tempering (also known as Metropolis-coupled MCMC)[71]. This method involves running multiple chains simultaneously and adding the option that two chains may swap their positions for their next step subject to acceptance criteria. This can improve sampling due to the difference in their priors leading to different likelihood ratios for the steps immediately after the swap.

Foundations have also been laid towards an interactive web interface for `Magellan` using `Jupyter Notebook`, `Bokeh`[35], `Pandas`[149] and `holoviews`[138]. This can be seen with an example dataset at the `Magellan` website.

We are also in the process of creating a `Singularity container` which would make distribution much easier. `Singularity` is a container platform, it allows one to combine their packages, environments etc. into a neat, portable and reproducible package that can be used on another machine. A particular boon of this is that, as it is self-contained, once in singularity form `Magellan` could be run on platforms other than Linux; this widens the audience for the tool immediately. This also removes the need for the user to install anything or to check versions of compilers etc. are all compatible. This task in particular is to be completed before the end of Summer 2023 for the initial publishing of `Magellan`.

Along with the Singularity container, we have begun the process of creating some straightforward `Python` modules and `Jupyter Notebook`s to provide the user with easy-to-use tools for analysing their Markov chains. It is important that users can make traceplots, examine the statistics of their chains and establish whether or convergence is happening, how much autocorrelation there is etc. As we have also recommended that they alter the likelihood functionality within `T3PS` in chapter 3 should there be issues with convergence or time taken between points we would like to add sections to compare these things in different set-ups. It would also be beneficial to add a switch to the `MCMC_Jobs` that allows a user to make these changes without having to directly edit the python code themselves - however this may be too ambitious to finish without new students joining the collaboration.

6.1.1 The 2HDM Type-II Signals

The shortfalls of the SM and lack of sufficient evidence for new physics at the LHC calls for ways to identify suitable regions for investigation based on any given model, we developed `Magellan` to help with this. The model chosen to test this tool on was the 2HDM Type-II, a scan of the parameter space was performed, retaining numerous properties for each point, taking into account both experimental and theoretical constraints. Then after the calculation of the cross-sections for the processes, $pp \to A \to Zh$,$pp \to H \to ZA$, a new analysis from ATLAS regarding, $A \to Zh$, was applied onto one bi-plane of the parameter space then projected onto others to show how `Magellan` can be used to place constraints on additional parameters with the use

of a Bayesian statistical framework. It was suggested that a low/mid mass CP-odd Higgs boson, ($m_A \leq 400GeV$), might exclude the alignment limit of the 2HDM Type-II.

Following this and the upgrade of `Magellan` with many additional experimental constraints the reduced parameter space was mapped, (though this is not claimed to be a complete mapping), for both the Type-I and Type-II 2HDMs. Cross sections were calculated for the processes, $bg \to twh$, $bq \to tqh$, and $qq \to tbh$. These processes provide direct access to the values of Yukawa couplings for the top and bottom quarks. In the wrong sign solution of the 2HDM these had the potential to produce significantly larger cross-sections than those seen in the SM. It was subsequently found that the 2HDM Type-I did not display an significantly large cross-sections for these processes, but that the Type-II did, despite the large reduction in the size of the available parameter space. Many points had cross-sections considerably larger than those of the SM, `Magellan` allowed for a gauge on the probability of each point so that this could be factored into the choice of which point to focus on in the analysis stage.

The next stage was to simulate a benchmark point via `MadGraph5@NLO`, `Pythia8` and `MadAnalysis5`. This allowed for the CMS detector to be simulated for the process after which cuts were applied for the analysis. It was shown that the 2HDM Type-II produced markedly different distributions compared to that of the SM for the p_T at the parton level as well as in that of the corresponding b-jet p_Ts at the hadron level. This provides strong motivation for the experimental community to seek such a signal, given that it should be very clear that it is the 2HDM Type-II Higgs being detected as opposed to the SM.

Appendix A

Magellan: Under The Hood

A.1 Dependencies And Installation

Currently Magellan is only available on Linux based systems, in the future we intend to create a Docker or Singularity container that can be used to run it on windows and mac in addition to this.

Magellan can be directly downloaded or cloned via git from the github repository at *github.com/Ceebs93/THDM_T3PS_scanner*.
There are several pre-requisites required to compile Magellan:

- A C++ compiler, for example GCC 11.1.0 .
- A Fortran compiler such as Gfortran
- Common build utilities such as make and Cmake
- The GNU Scientific Library (GSL); this is a C library for numerical computations. Users on a Linux operating system will find that most package managers will have a stable version of this. Magellan is compatible with v2.6 or later. It can be found on the GSL website.
- Python 2, the authors recommend using v2.7 or later but explicitly **NOT** Python 3. We also recommend the use of either Anaconda or Miniconda (a 'minimal installer for conda') to manage and install Python packages. Using one of these will ensure the correct versions of packages for the Python environment (version) are installed whenever a new one is added. Additionally, Magellan contains a file "THDM-env.txt" which can be used to create an environment inside Anaconda/Miniconda containing all the necessary Python packages with a single command, saving the user time and effort. This will be shown later in this chapter.
- Pandas; this is a python library that allows the use of mutable dataframes, tables that can be manipulated easily with python and written to CSV or HD5f format.

Pandas can easily be obtained through the use of conda with the command "conda install pandas" while in an activate conda/miniconda environment.

- NUMPY; this is a python library for computations. When downloading ensure that the version is one that is compatible with Python 2 rather than Python 3. The authors note that installing Pandas will install NUMPY too automatically; thus the recommendation is to simply do this.

The toolbox was successfully compiled and run with GCC 11.1.0 and GSL 2.6 . The authors are reluctant to recommend using versions released much before these as the nature of the toolbox means that versions must satisfy all contained packages to ensure functionality.

The easiest way to obtain the toolbox is to use a GIT clone directly into the directory for installation of Magellan. Alternatively it can be downloaded directly, moved to the installation directory and uncompressed. In either case the next step is to enter the main directory, "THDM_T3PS_scanner", (which will henceforth simply be referred to as 'the top directory' for the sake of simplicity), where the "env_local.sh" (Figure A.1) file can be edited to set up the environment for running (after an initial setup is performed).

```
#·Load·Python·environment
module·unload·--all
module·load·conda
#conda·activate·THDM
source·activate·THDM

#·Load·libraries·and·compiler
module·load·gcc/11.1.0
module·load·gsl/2.6

#·Print·THDM_T3PS_SCANNER_DIR·path·to·console·for·user·to·check
echo·"Scanner·Dir:·${THDM_T3PS_SCANNER_DIR}"
```

A.1 env_local.sh used by authors on the computing cluster IRIDIS5

Here is where any conda environments you may wish to use should be activated, and if using Magellan on a cluster it is where modules should be loaded. If using on a local machine, with only Python 2 installed then it is not, in theory, necessary to use "env_local.sh" in which case the user may remove the file and continue the setup as normal. A .yml file, ("THDM.yml") is provided in *utils* which can be used to replicate the authors' conda environment. If the user wishes for the conda environment to be saved to a specific place they can make a small edit to provide the path to this location where they wish to put this conda env, otherwise it will simply go to the default path. This can be done on the final line of the .yml file and looks like this:

```
prefix:  /home/user/.conda/envs/THDM
```

Then to create the conda environment run the following:

```
conda env create -f THDM.yml
```

This command will install all necessary python packages to the environment. This environment can then be activated with:

```
conda activate THDM
```

When not using `Magellan` the environment can be easily deactivated with:

```
conda deactivate
```

A.1.1 Automatic Installation

It is recommended that the automatic installation option is used. Once the user has activated their environment so that the correct versions of all dependencies will be active they can find and run the auto-install file from the top directory, like so:

```
bash setup_auto.sh
```

If this command is executed successfully, the code in the toolbox should have been edited to contain the individual paths relevant to the user's installation. If this is not the case then the user will need to go through the toolbox and replace all instances of "`top_dir_`" with this path manually. In case of such an eventuality a file is provided with the locations of occurrences listed (or check "`/utils/set_home_dir.sh`" to see where this is being changed.). This file is named "`Absolute_paths.md`", given as an md to make it easy to keep track of what has and has not been changed.

A.1.2 Manual installation

If an issue occurs and the auto-install file that is called on by "`setup_auto.sh`" does not work then the toolbox can be compiled manually as detailed below.

- While in the top directory run: `export THDM_T3PS_SCANNER_DIR=$PWD`
- then run: "`bash utils/set_home_dir.sh`"
- if on a cluster load the following modules (if not just ensure they are installed on your computer):
  - `module load gcc/11.1.0`
  - `module load gsl/2.6`
  - `module load cmake/3.22.0`
  - `module load conda`

- Next, run `source env.sh` to activate the THDM conda environment (or replace this with personal equivalent); this is important to do as much of the other code relies on this path being present and the toolbox will **not** function without doing this.
- The external physics packages are installed next by running `bash packages/install.sh` [1]
- Now symbolic links can be set up with `bash utils/setup_links.sh`
- Finally, enter the "*ParameterPointProcessor*" directory and run `make`
- If "`set_home_dir.sh`" has run correctly, all the code in the toolbox should have been edited to contain the individual paths relevant to the user's installation. If this is not the case then the user will need to go through the toolbox and replace "`top_dir_`" with this path manually. In case of such an eventuality a file is provided with the locations of procurances listed (or check "`/utils/set_home_dir.sh`" to see where this is meant to be changed). This file is named "`Absolute_paths.md`", given as an md to make it easy to keep track of what has been changed.

- The first external package to be installed is `HiggsBounds`, this needs to be done first as `HiggsSignals` and `2HDMC` need it to be compiled already in order to access its' library. This is done as instructed by the authors of the packages themselves [2]

  [3], and as detailed below. [4]
- `cd /packages/HiggsBounds-5.10.1/`
- `mkdir build & cd build`
- `cmake ..`
- `make`

Once this is done we move on to the compilation of `HiggsSignals` (assuming you are continuing on from the compilation of `HiggsBounds` and thus in the build folder for that):

- `cd ../../HiggsSignals-2.6.2`

[1] If this step fails please refer to later in this section for a guide to individual installation
[2] `HiggsBounds` and `HiggsSignals` manuals can be found here.
[3] The `2HDMC` manual can be found here.
[4] The `LHAPDF` manual can be found here.

- The compilation is identical to that of HiggsBounds[5], `mkdir build & cd build`
- `cmake ..`
- `make`

Now the last of the packages to install is 2HDMC;

- `cd ../2HDMC-1.8.0`

In order to connect HiggsBounds and HiggsSignals to 2HDMC we must copy their libraries over to the `lib` directory in 2HDMC as follows;

- `cp ../HiggsBounds-5.10.1/build/lib/libHB.a lib/libHB.a`
- `cp ../HiggsSignals-2.6.2/build/lib/libHS.a lib/libHS.a`

Finally 2HDMC can be compiled with `make`

If there is an issue with any of these installations it is recommended you check the guides written by each of the developers, these are hyperlinked in the footnotes of the previous page or alternatively these are easily found by putting the package name and 'install' into a search engine.

Next is MadGraph5@NLO [7], as this is written in Python there is little installation to do. The recommended version is provided in the "packages" directory, this is version MadGraph5@NLO-V-3.1.0 , but any version compatible with Python 2 may be used.

- This link is to the official MadGraph5@NLO launchpad where all available versions of MadGraph5@NLO can be found and downloaded should they be required: MadGraph5@NLO launchpad.
- In order to complete the installation, run MadGraph5@NLO with `./bin/mg5_aMC` or `python2 bin/mg5_aMC`[6]
- Once MadGraph5@NLO has started the user can install several other programs using the interface, such as LHAPDF, Boost, Pythia as needed. Entering "`install -help`" will display all available programs.

MadGraph5@NLO comes with the SM and several BSMs already present, but the user may wish to add a different one or indeed their own. This can be done with any UFO model by simply copying it into "`MadGraph5@NLO /models/`".

[5] HiggsSignals should detect the library of HiggsBounds automatically, however if you find you are having difficulties running the toolbox this is something to check in the relevant Makefiles

[6] The authors have found that unless Python 2 is the only installation of Python on your system it is best to specify this

Finally there are some links that must be set-up from the top directory in order to compile ParameterPointProcessor. This is done by simply running

```
bash setup_links.sh
```

The user should be sure that they check the paths are correct before running this, or in the event that there are issues compiling the ParameterPointProcessor.

```
#!/usr/bin/env bash

ln -f -s ${THDM_T3PS_SCANNER_DIR}/packages/2HDMC-1.8.0/src/ ${THDM_T3PS_SCANNER_DIR}/links/inc/2HDMC
ln -f -s ${THDM_T3PS_SCANNER_DIR}/packages/2HDMC-1.8.0/lib/lib2HDMC.a ${THDM_T3PS_SCANNER_DIR}/links/lib/lib2HDMC.a
ln -f -s ${THDM_T3PS_SCANNER_DIR}/packages/2HDMC-1.8.0/lib_HBHS/lib2HDMC_HBHS.a ${THDM_T3PS_SCANNER_DIR}/links/lib/lib2HDMC_HBHS.a
ln -f -s ${THDM_T3PS_SCANNER_DIR}/packages/HiggsBounds-5.10.1/build/lib/libHB.a ${THDM_T3PS_SCANNER_DIR}/links/lib/libHB-5.10.1.a
ln -f -s ${THDM_T3PS_SCANNER_DIR}/packages/HiggsSignals-2.6.2/build/lib/libHS.a ${THDM_T3PS_SCANNER_DIR}/links/lib/libHS-2.6.2.a
```

A.2 setup_links.sh used by authors on the computing cluster IRIDIS5

Following this the ParameterPointProcessor needs to be compiled. Provided everything prior to this has run smoothly, this should be straightforward;

- cd out of your 2HDMC directory back to the top directory of Magellan
- cd into the ParameterPointProcessor directory
- Run make

A.2 Running The Package

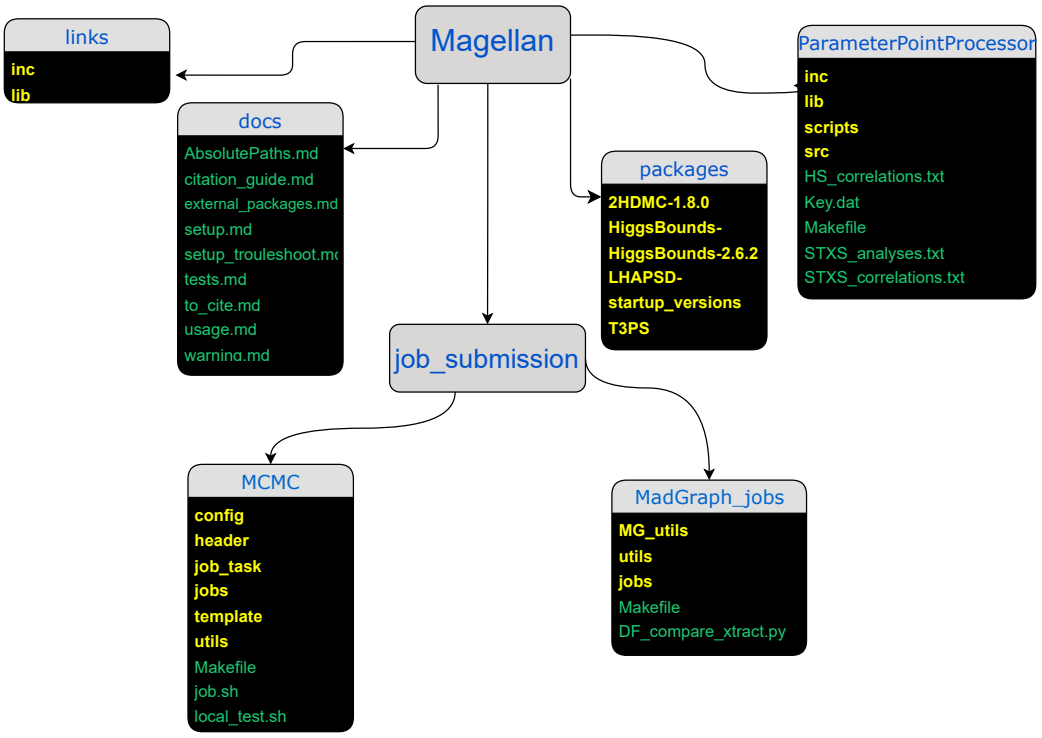
There are two modes within Magellan that can be run, with the parameter-scanner alone or with MadGraph5@NLO (and hence, calculation of cross sections for the parameter space points). This second mode can be done using the output of the scanner directly or with an externally sourced datafile. Regardless of which mode is being used it is **imperative** to source env.sh, which can be done from any directory[7]. If this is step is skipped the toolbox will not work, as it relies on additions to the bash $PATH made in this step.

A.2.1 The parameter scanner

The Makefile located in the MCMC directory is central to use of the scanner, it is split into four parts, these can be seen in detail in subsection A.2.5. However, for ease of use a bash script has been created in the same directory, called 'control_file.sh'. When run this file will query the user about what they want to do, create/submit/merge jobs, which basis they want to use etc. This information is saved to 'utils/settings.mk' and then read by the makefile when it is called[8]. This method

[7] Though it should be noted that sourcing it will change the current position to that of the top directory

[8] Which will be done by the "control_file.sh" as part of it running.

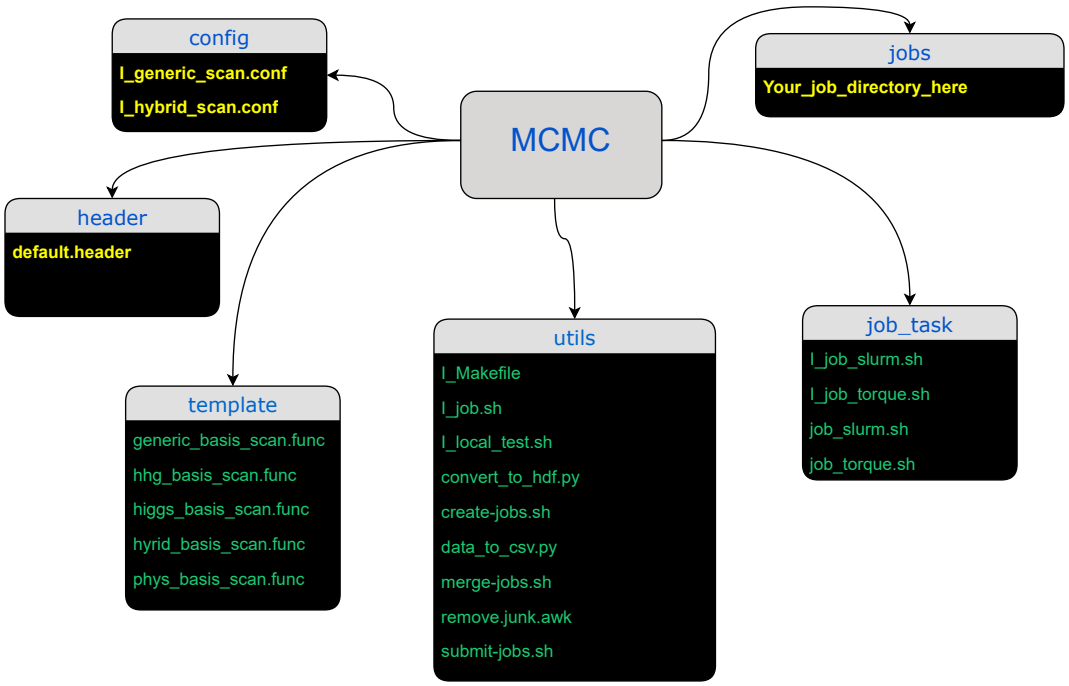


A.3 Diagram of the layout for the top directory of `Magellan`. The yellow names indicate directories, green names are text files and blue names are directories that are expanded in this diagram. For example the blue "Magellan" is the expanded directory which contains everything the arrows from it point to. Then the blue "Packages" is a directory within "Magellan" contain the yellow directories in the black box below it. Note that there are text files directly in the top directory as well, but these were not included for visual clarity.

means that the user does not have to worry about editing the makefile or ensuring they have changed the template and config files to match the basis they want to use. It is still possible to edit the Makefile directly should the user wish to, for instruction on this please see subsection A.2.5).

Once the env.sh file has been sourced, `cd` into `job_submission/MCMC`, this is the directory jobs are created in. What follows is a summary and explanation of the contents;

- **The `config` directory**. Inside is the configuration file needed for jobs. Below are descriptions of the different sections therein
  - The `[setup]` section of the config file is shown in Figure A.5. Here we indicate which "`point_processor`" to use and which template. If running via the "`control_file`" there is no need to manually edit this section, it will be done automatically.
  - The "`point_processor`" is the script that `T3PS` will implement to run, as well as interpret the `ParameterPointProcessor` program which itself



A.4 Directory layout for "`MCMC`" section of `Magellan`. Yellow names indicate directories, green names are text files and blue names are directories that are expanded in this diagram. For example the blue "MCMC" is the expanded directory which contains everything the arrows from it point to. Then the blue "config" is a directory inside "MCMC" which contains the yellow directories in the black box below it. Note that there are individual files in the "MCMC" directory such as the Makefile and "`control_file`", but these were not included for visual clarity.

```
##########################################
[setup]

mode = mcmc

point_processor = /path/to/THDM_T3PS_scanner/packages/T3PS/processors/SimpleProcessor_modified_OG.py

template = job.template

# - Length of the Markov chain
unit_length = chain_length_

concurrent_processors = nCores_
```

A.5 "`[setup]`" in configuration file for the Hybrid basis. Here the mode, chosen point processor and template can be set. When running "`make create-jobs`" this template will be copied into the new job directory where "`chain_length_`" and "`nCores_`" will be replaced with the appropriate value from the makefile. If using the `control_file` all these variables will be set automatically. If running via the Makefile directly, the user needs to set the "`point_processor`" path themselves (though running the setup file correctly when installing will replace "`/path/to/THDM...`" with the top directory path and so only the specific processor needs to be added).

incorporates `2HDMC`, `HiggsBounds`, `HiggsSignals`. Fundamentally it will run whichever ParameterPointProcessor is specified by the Makefile and

read the output from it. As such it is vital that it is edited appropriately to only keep the useful bits of information. This will be addressed later.

- The parameter section shown in the LHS of Figure A.6 allows the user to define their parameters for input. This is done by first declaring them in "`par_names =`" then declaring their values by setting "`par_yourvariable =`" to the required values. This can be done in several ways,
  - A finite list of numeric values e.g. "1, 3, 5, 7" (these need not be evenly spaced)
  - A randomly generated value in a normal distribution (this uses the random module in python to generate) i.e. "normalvariate(μ, σ)"
  - A numerical constant

  This is **not** automatically edited, so be sure to check it matches what you want before running.

```
# -- Parameters -- #

par_names = Z7, mH, mHc, mA, cba, tb

par_tb  = interval( 5, 11.5) with mcmc_stepsize = 0.5
par_Z7  = interval( -10, 10) with mcmc_stepsize = 0.2
par_mH  = interval( 150, 1000) with mcmc_stepsize = 20
par_mHc = interval( 580, 1000) with mcmc_stepsize = 20
par_mA  = interval( 150, 1000) with mcmc_stepsize = 20
par_cba = interval(0.36, 0.4)  with mcmc_stepsize = 0.01
```

```
# -- Bounds

data_names = Z7_in,
  → mH_in,
  → mHc_in,
  → mA_in,
  → cba_in,
  → tb_in,
  → sinba,
  → Z4_c,
  → Z5_c,
  → m12_2,
```

A.6 `[params]` in configuration file of the Hybrid basis. Here the user defines their input parameters and the values they wish to use for them. This can be done in several ways; a simple list of numeric values, an interval i.e. "interval(1,2) with count=5" or in the form of a normal distribution, "normalvariate(μ, σ)" `[bounddata]` in configuration file. Output variable names are given, in order of output, here.

- In the next section of the config file, shown in the RHS of Figure A.6 the "`data_names`" are defined. The ParameterPointProcessor will output a long string of variable values and here the program is told what name to assign each of them to, in order. As such, it is very important to ensure they are listed in the correct order! This is **not** automatically edited, the user should be sure to check it matches the desired order before running.

- The final part of the config file is the "`[algorithm]`", this is the expression the user wishes to use to calculate the likelihood. This is a critical part as the MCMC relies on this to determine which steps to take. By default this is set to a Gaussian function, with ${\chi_{tot}}^2$ as defined in Equation 3.10. It can be defined with any variables that have **already** been declared within the config file, however, we

do not recommend altering this unless the user has some familiarity with MCMC likelihoods or good reason to believe a Gaussian is not appropriate here.

```
##############################
[algorithm]

likelihood = exp(-0.5 * data.chi2_Tot_gfitter )
```

A.7 `[algorithm]` in configuration file. Here the expression for the likelihood is given, by default this is Gaussian.

- The next aspect is the "template" file. This defines the model type (of the 2HDM) in order to run `2HDMC` inside the ParameterPointProcessor. The variables should match those defined in the "Parameters" section of the corresponding config file. There are multiple template files, these correspond to the different ways `2HDMC` can accept input e.g. the generic basis, (HiggsHuntersGuide)hhg basis, Higgs basis, hybrid basis and physical basis. See Figure A.8. Note that the first value in each of these is always a number, this number indicates the type of 2HDM to use. When using the "control_file" the user will be asked to indicate which model type to use which will change the template file to match automatically. If, instead, the user runs via the makefile then they will need to replace "`Y_` with the appropriate number for the model type they want.

A.2.2 Running MCMC Jobs

A.2.2.1 Creating, Submitting and Merging

The following instructions make the assumption that the user is running via the `control_file`, instructions to run via the Makefile directly can be found in subsection A.2.5

- From the top directory run "`source env.sh`"
- `cd` into `job_submission/MCMC/`
- Find the config file that matches the desired parametrisation in the `config` directory, and open this to edit the parameter ranges you will be scanning over. If unsure of how to input this, check the description for Figure A.6 for details of accepted formats.

```
2 $Z7 $mH $mHc $mA $cba $tb
```

```
1 $mH $mHc $mA $mHc $sba $L6 $L7 $m12_sq $tb
```

A.8 On the left we see the template for the hybrid basis with a Type-II model while on the right is the template for the physical basis with a Type-I model.

- Run `bash control_file.sh` and answer each of the questions as they appear. The merging step after the MCMC has run will need to be done separately, but again it is just a case of using the control_file.sh.

Figure A.9 shows the start of an instance of "`control_file.sh`" being used to create and run an MCMC job. This shows how straightforward questions are used to prompt the user in order to avoid ambiguity.

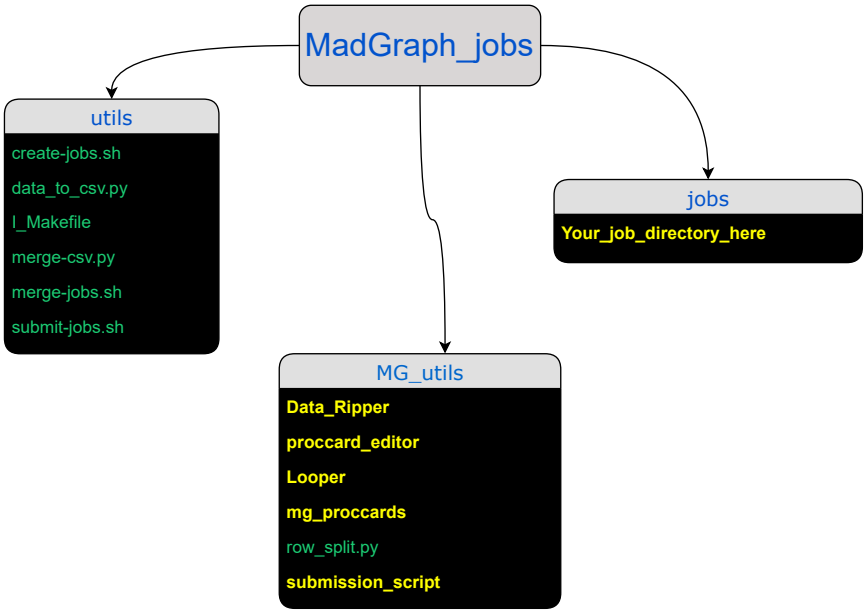
```
(THDM) [cb27g11@cyan53 MCMC]$ bash control_file.sh
Starting setup for MCMC jobs...
Enter 1 to create a job and submit/run it, enter 2 to create ONLY, enter 3 to submit ONLY or enter 4
to merge previous job output.
1
Enter job_name
Test_Job
Create and submit job selected
Setting up create-jobs
Enter chosen parameterisation basis: higgs, hhg (higgs hunters guide), generic, phys (physical) or hy
brid
generic
Do you want to run a cluster job (enter yes for cluster, no for local)?
yes
Enter number of cores to use
1
Enter number of jobs to run
2
Enter chain length
100000
./utils/create-jobs.sh called

####################
NAME:         Test_Job
CONFIG:     config/generic_scan.conf
CLUSTER:    yes
nJobs:      2
TEMPLATE:   template/generic_basis_scan.func
ROOT_DIR:   /scratch/cb27g11/THDM_T3PS_scanner/job_submission/MCMC
####################
```

A.9 A snapshot of the "`control_file.sh`" being run. In the snapshot the user has selected the options for creating and submitting jobs, using the generic basis, running on a cluster(HPC), and using a single core with 2 jobs running in parallel. Both of the jobs are set to run to a chain length of 100000. This does not demonstrate the complete set-up, further questions will be asked to establish other necessary variables, but this shows what one should expect to see when running.

A.2.3 Running MadGraph Jobs

```
(THDM) [cb27g11@cyan52 Data_Files]$ pwd
/scratch/cb27g11/THDM_T3PS_scanner/job_submission/MadGraph/jobs/NC_bgh1/job_002/Data_Files
(THDM) [cb27g11@cyan52 Data_Files]$ ls
bg_twh1.csv                    bg_twh1_7.925020_0.968783.csv
bg_twh1_10.528300_0.999493.csv bg_twh1_7.931510_0.970997.csv
bg_twh1_11.798900_0.992435.csv bg_twh1_7.935210_0.974078.csv
bg_twh1_11.884400_0.999075.csv bg_twh1_7.935360_0.975384.csv
bg_twh1_12.107100_0.998873.csv bg_twh1_7.941800_0.974979.csv
```

A.10 Figure shows location of an example "Data_Files" directory for a job named "NC_bgh1" which had the "CREATE_JOB_PROCESS_NAME" variable of the Makefile set to "bg_twh1"



A.11 Directory layout for the "MadGraph_jobs" section of Magellan. Yellow names indicate directories, green names are text files and blue names are directories that are expanded in this diagram. For example the blue "MadGraph_jobs" is the expanded directory which contains everything the arrows from it point to. Then the blue "MG_utils" is a directory inside "MadGraph_jobs" which contains the yellow directories in the black box below it. Note that there are files directly in the "MadGraph_jobs" directory in addition to those shown, such as the Makefile, but these were not included for visual clarity

This part of Magellan aims to take a CSV (or HDF5) file of parameter points, usually these would be output from the scanner section, but it is possible to use externally sourced data[9] and run MadGraph5@NLO on each point sequentially before combining the

[9]Note that it is important the data has a header that matches the output coming from the MCMC section.

input CSV with the newly calculated cross sections added on. It is important to note that this will need to be set up for any new models introduced. Setting up for a new model, new variables of interest and/or new requirements for the MadGraph5@NLO simulation can be somewhat involved; however, once it is done a job can be created and left to run over thousands of points without having to interact with the code for each individual one.[10]

A.2.3.1 Setup for New Processes and Models

While Magellan can run any MadGraph5@NLO process over a large amount of points, it is necessary to first run a single instance directly in order to generate a "proc_card".

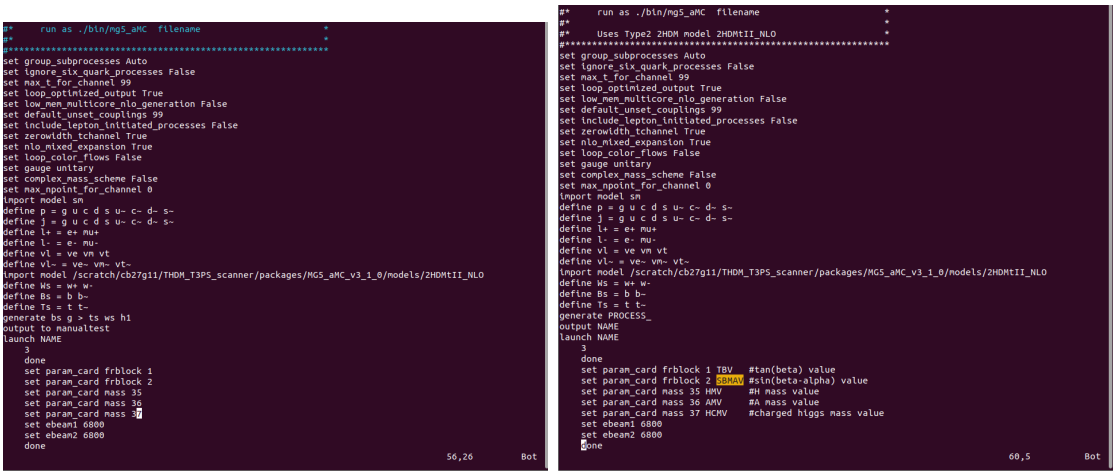
- First, the user should place their chosen UFO model in "packages/MG5_aMC_v3_1_0/models", which is located in the 'top_dir'.
- Then the user should start MadGraph5@NLO ,**without** answering yes to an update of MadGraph5@NLO[11].
- Import the model they wish to use, define any new particle groupings they need, (e.g. setting, $Ws = w^+w^-$), and edit the parameter card when prompted for each of the masses as well as any other variables that are changed across their points.
- It is also possible to do this by altering one of the example "proc_card"s which are located in "MadGraph_jobs/MG_utils/mg_proc_cards". Even so, it is easier to avoid mistakes by using MadGraph5@NLO to create a new one.
- Once a new proc_card has been created, copy this to "MadGraph_jobs/MG_utils/mg_proc_cards".
- Please refer to subsubsection A.2.6.1 for details on how to create a proc_card using MadGraph5@NLO this if you are unsure, or see the MadGraph5@NLO wiki page, which also has a link to the official MadGraph5@NLO question and answer page should further help be required.

Now the proc_card will need a few small edits to work within the MadGraph5@NLO job structure.

In Figure A.12 we can see a comparison between an original MadGraph5@NLO "proc_card" and a template that has been created from it. The majority of the text is unchanged across the two, but we do see the following differences;

[10] It is very much recommended that this is done on a computing cluster as MadGraph5@NLO can be quite computationally intense. The greatest benefit of this section will be when the user spreads the points across numerous nodes, possible to an extent with a powerful computer but considerably easier on a cluster.

[11] The update would make MadGraph5@NLO incompatible with the rest of the toolbox as Magellan runs on python2 and upgrading will switch MadGraph5@NLO to python3



A.12 On the left is an unaltered `MadGraph5@NLO` "proc_card", it loads a model for the 2HDM type-II, runs a process and alters the values of some variables. On the right we have the template constructed from that proc_card. In the altered version we see the process being generated has been replaced with a placeholder so that *any* can be inserted. Similarly a placeholder has been added to the variable changes so that different values can be inserted. Various comments have also been added to make using the template easier.

- In place of an actual process after "generate" there is a placeholder variable, "PROCESS_"
- Placeholders have similarly been added to allow `Magellan` to substitute in values for variables. The substitution is performed by the "`proccard_editor.py`" located in "`MadGraph_jobs/MG_utils/proccard_editor/`.
- Comments have been added to make it easier to recall what each param_card change is altering, this is of course optional, but can be very helpful - the PDG particle numbering scheme was used for this and can be found on the particle data group website or instead one can find the SM particles here.

We move on to discuss the 'looper' script, this will iterate over each point in a given datafile. Several examples of these can be found inside "`MG_utils/Looper/`". These scripts will read through a CSV file, line by line, take the value in column 'i' and read it into a corresponding variable. For this reason a new 'looper' script is needed for different runs; the columns it reads from, the number of them and the variables they are being read into might vary depending on your process. This is easy to alter:

Below are instructions for how to create a Looper script for a new process, starting from a template script:

- First, make a copy of one of the example Looper scripts to create a template for your new one.
- On line 11 of the new file replace/remove any unwanted variables (ensure that this is carried through the full file).

```
while IFS="," read -r mH mHc mA Itb Isinba } < <(cut -d "," -f2,3,4,6,7 CSV_NAME_.csv)
```

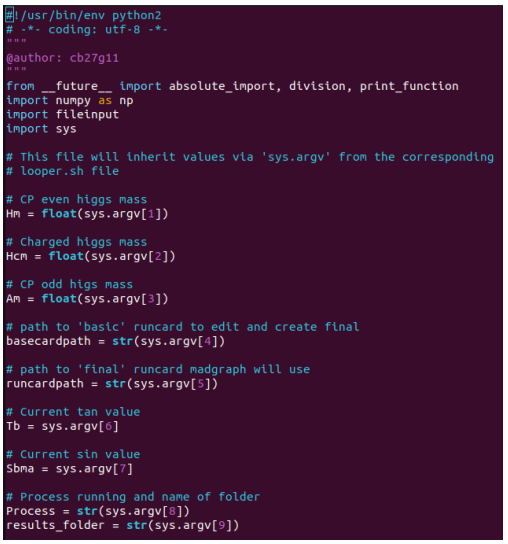
A.13 On the LHS we have line 11 of a Looper example script, "mH", "mHc", "mA", "Itb" and "Isinba" are the variables to be read into. These can be altered or added to them as needed. The RHS shows the *final* line of the same file (line 49). The important part here for the purposes of changing variables is "-f2,3,4,6,7" this indicates that the 2nd column is read into the first variable, the 3rd to the 2nd variable, 4th to the 3rd variable and so on. It is important to check that these pairings corresponds to the values to be read in along with their intended variable, and that something is being read in for each variable in line 11.

- If the user wishes to add additional variables, add these to the end of line 11 too.
- The user may note that in the examples "$tb" and "$sinba" are treated slightly differently. This is because in lines 25-30 they are being used to define directory names for the output, specifically for a given point the values of these variables is used to name it. This allows for distinct names as well as allowing the names to provide information on the point. Which variables are chosen to fulfil this role is arbitrary, but **they are needed for the naming convention**. Alternatively, the user needs little coding knowledge to incorporate a different naming convention that they prefer.
- On the final line, the numbers following "-f" need to be in the same order as the variable names were given in line 11. I.e. regardless of what column "mH" occupies in the datafile, if it is given as the 5th variable on line 11 then it must be supplied as the 5th number following "-f".

Afterwards the user needs to create a copy of an example "`proccard_editor.py`" file, (any will suffice), which are located in "`MG_utils/proccard_editor/`".

This file receives values from the '`Looper.sh`' file provided by 'argv'. Figure A.14 shows an example "`proccard_editor`"; lines 15-37 show the values from the corresponding "`Looper`" file being assigned to variables. The user needs only to confirm that the variables correspond correctly to those in both their 'Looper.sh' script and their input CSV file, not forgetting that the ordering is important! The 'make_input' function in this file replaces the 'placeholder' names in the proc_card template with variable values. The user should ensure these correspond correctly with their proc_card template.

Similarly a "`Data_Ripper.py`" script needs to be copied and edited, an example can be seen in Figure A.15. This script extracts data from the `MadGraph5@NLO` output, usually the cross section and 'comparison' variables. The comparison variables being the ones chosen to ensure points are correctly matched up after `MadGraph5@NLO` is run.
Required edits are as follows;



```
#!/usr/bin/env python2
# -*- coding: utf-8 -*-
"""
@author: cb27g11
"""
from __future__ import absolute_import, division, print_function
import numpy as np
import fileinput
import sys

# This file will inherit values via 'sys.argv' from the corresponding
# looper.sh file

# CP even higgs mass
Hm = float(sys.argv[1])

# Charged higgs mass
Hcm = float(sys.argv[2])

# CP odd higs mass
Am = float(sys.argv[3])

# path to 'basic' runcard to edit and create final
basecardpath = str(sys.argv[4])

# path to 'final' runcard madgraph will use
runcardpath = str(sys.argv[5])

# Current tan value
Tb = sys.argv[6]

# Current sin value
Sbma = sys.argv[7]

# Process running and name of folder
Process = str(sys.argv[8])
results_folder = str(sys.argv[9])
```

A.14 Image shows an example of the start of a "`proccard_editor.py`" file. We can see how the built-in function "`argv`" is assigning values passed from the '`Looper.sh`'

- Replicate the code that calls '`Xtracted_values`' for each of the desired variables/values. This can be done by simply copy and pasting the existing lines as many times as needed then altering the variable names.
- When the code calls '`Xtracted_values`' it reads each line into a list broken up by the spaces. The user must enter a 'search string' that can be used to identify the correct line for the variable value to be extracted for each of the required variables.
- Following this, the user should add the position in the line they are at, e.g. in the sentence "I want 5" the number '5' is at position 2 and 'I' is at position 0; so we would put "I want" as the search string, and '2' as the position to indicate we want the line containing that string and we want the part in position '2'[12].

Figure A.16 shows a subsection of the MadGraph job Makefile '`create-jobs`' - this section controls the setup for a job, whether this is on a cluster or locally. Most variable names are intentionally quite intuitive, the '`CREATE_JOB_`' prefix for each variable name will be stripped so this is ignored when interpreting the variable names. This means that '`CREATE_JOB_INPUT_DATA`' will be '`INPUT_DATA`' in all code it is passed to. This variable in particular will be the data you wish to run `MadGraph5@NLO` on.

- Once there is a version of all the files necessary for your run/process you can start by opening the Makefile and editing the variables in the '`create-jobs`' section. Please refer to Fig.Figure A.16 to see what each one is for. Once this has

[12] When choosing a 'search string' it is best to find one that is actually unique to the value required. Looking at Figure A.15 we can see that if we just used '`# mh`' for the third and fourth '`Xtracted_values`' calls then we would erroneously be assigning **both** the values of mh2 and mh3, which would cause problems later.

```
# The strings here will be used to find the values for each variable, make sure
# to change them if you are looking for something else
MGVAR1_LABEL_ = Xtrctd_values(File_List, '# tanbeta', 1, Multi=False)
MGVAR2_LABEL_ = Xtrctd_values(File_List, '# sinbma', 1, Multi=False)
MGVAR3_LABEL_ = Xtrctd_values(File_List, "# mh2", 1, Multi=False)
MGVAR4_LABEL_ = Xtrctd_values(File_List, "# mh3", 1, Multi=False)
X_sections = Xsection_values(File_List, '#  Integrated weight (pb)  :', 5,
             Multi=False)

# Changing our variable lists to arrays
MGVAR1_LABEL_ = np.array(MGVAR1_LABEL_)
MGVAR2_LABEL_ = np.array(MGVAR2_LABEL_)
MGVAR3_LABEL_ = np.array(MGVAR3_LABEL_)
MGVAR4_LABEL_ = np.array(MGVAR4_LABEL_)
X_sections = np.array(X_sections)

# Ensuring all values have been correctly read in as floats not strings
MGVAR1_LABEL_ = MGVAR1_LABEL_.astype(float)
MGVAR2_LABEL_ = MGVAR2_LABEL_.astype(float)
MGVAR3_LABEL_ = MGVAR3_LABEL_.astype(float)
MGVAR4_LABEL_ = MGVAR4_LABEL_.astype(float)
X_sections = X_sections.astype(float)

# Create a pandas dataframe to contain data
nw_df = pd.DataFrame({'MGVAR1_LABEL_':MGVAR1_LABEL_, 'MGVAR2_LABEL_':MGVAR2_LABEL_,
                      'MGVAR3_LABEL_':MGVAR3_LABEL_, 'MGVAR4_LABEL_':MGVAR4_LABEL_,
                      'X_sections':X_sections)
```

```
CREATE_JOB_MGVAR1_LABEL  = "tb"
CREATE_JOB_MGVAR2_LABEL  = "sinba"
CREATE_JOB_MGVAR3_LABEL  = "mH"
CREATE_JOB_MGVAR4_LABEL  = "mA"
```

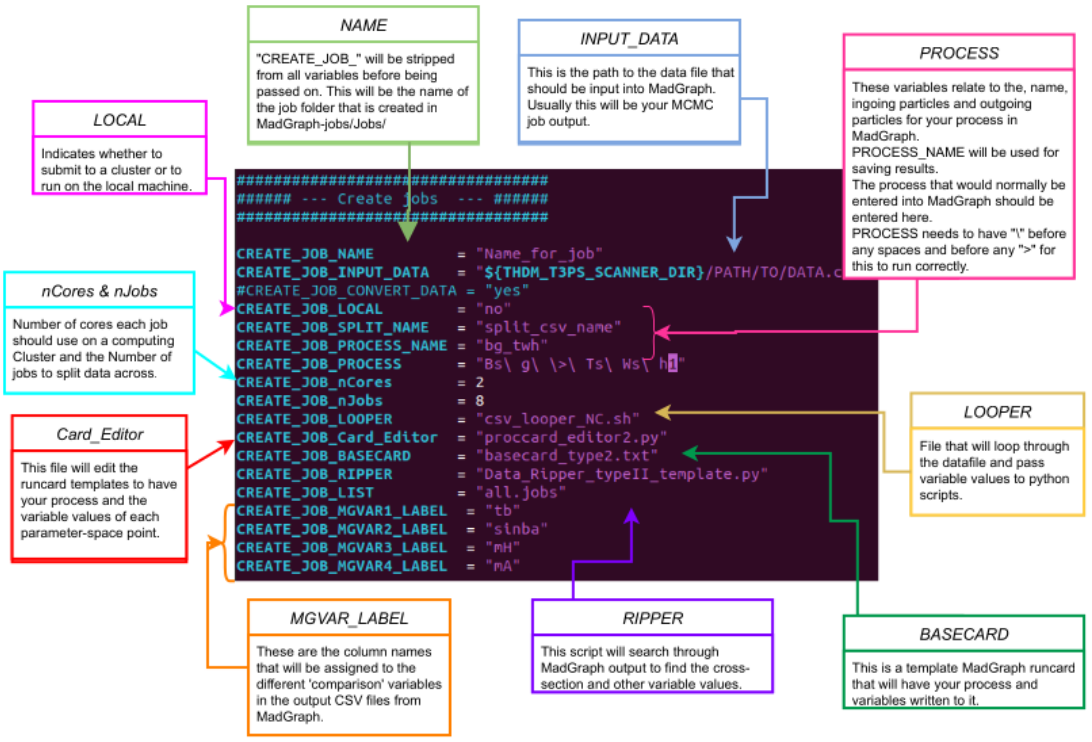
A.15 The top image is a section of an example "Data_Ripper" file; where it extracts values for each of the variables. In the lower image we see a section from the Makefile; recall that 'CREATE_JOB_' will be stripped from the variable name before it is inserted into the "Data_Ripper" file accordingly. Everywhere in the top image that reads "MGVARN_LABEL", where N is an integer, will be replaced with the string "CREATE_JOB_MGVARN_LABEL" in the bottom image. To extract a different variable the user can either alter an existing variable, or duplicate one first. Add a different number to the label, (this does need to **keep** the format in order for everything to function, unless further alterations are made to multiple files), add an appropriate 'search string' and line position. Then ensure that the new label is reflected in the makefile

been done simply enter "make create-jobs", this will create a new folder inside "MadGraph/jobs/" which will contain all the files needed for the job.

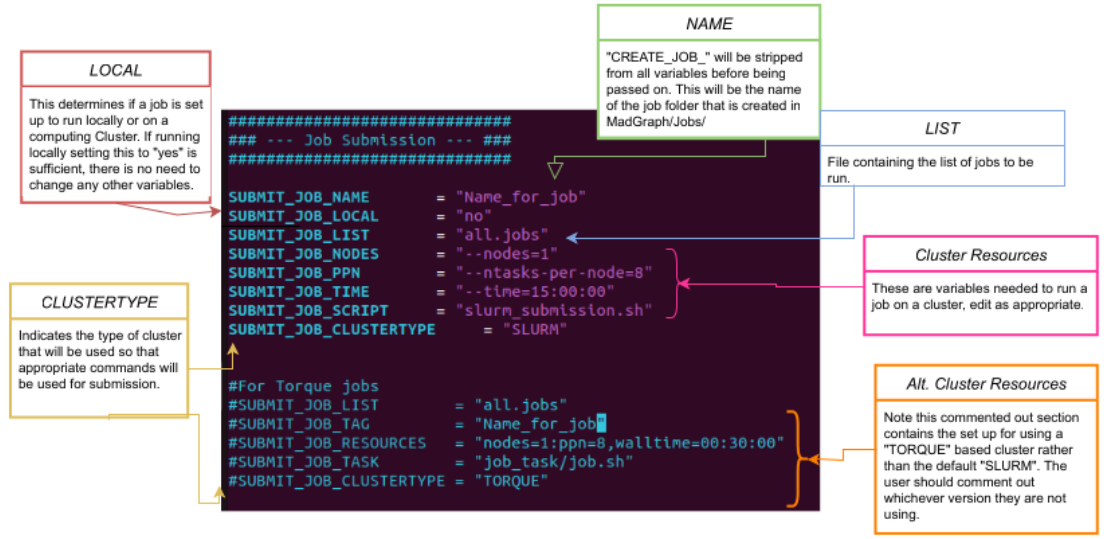
- Next the user should edit the 'submit-jobs' section of the Makefile. Please refer to Figure A.17 to see what each of the variables are for in detail. Note that Magellan is compatible with both SLURM and Torque job managers if using HPC; the user should take care to comment/uncomment the appropriate part of the section with this in mind. The first set of variables (uncommented by default) is for running on a SLURM system, while the second (commented out by default) is for a Torque system. If running locally then the user need only set the "SUBMIT_JOB_LOCAL" variable to yes and need not worry about the others.

- Once the previous step is done and the user is ready to submit they should enter "make submit-jobs". This will submit all the jobs created by the user in the previous step.

- For the final step, the code will combine the outputted cross sections from each job together into a CSV file then merge that CSV with the original input data



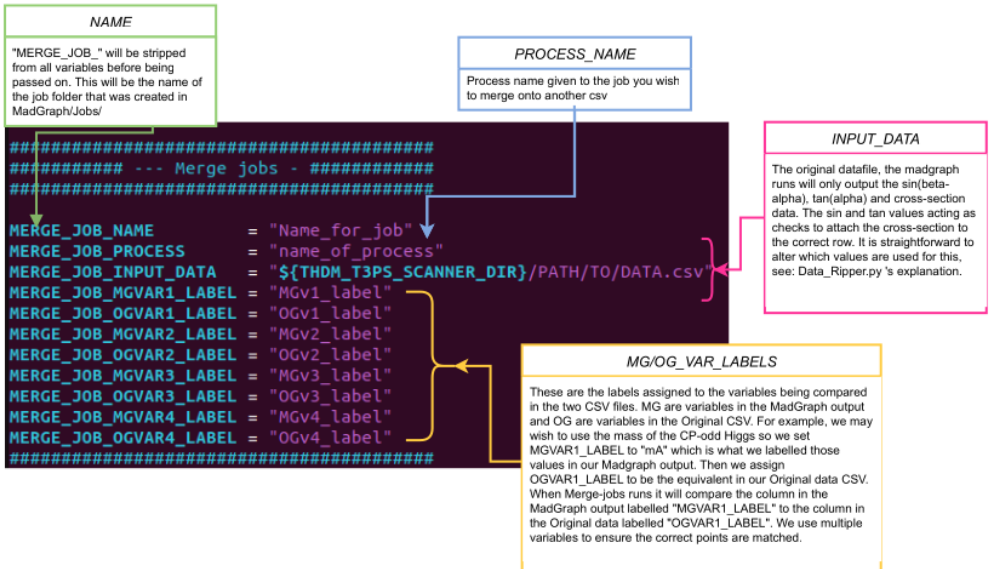
A.16 "`create-jobs.sh`" file for `MadGraph5@NLO`-type runs



A.17 "`submit-jobs.sh`" file for `MadGraph5@NLO`-type runs

to. By default the output from '`submit-jobs`' will only contain the 'comparison' variables and cross section values. The 'comparison' variables are there in order to ensure the cross section values are correctly matched up to the corresponding data point from the original file. This method was chosen because it is quicker to have the "`Data_Ripper.py`" that runs as part of the jobs already identify the

cross section and other select variables only, store these in single point CSV files and simply merge these back onto the original later. Please refer to Figure A.18 to see the variables for this section.



A.18 "`merge-jobs.sh`" file for `MadGraph5@NLO`-type runs

A.2.3.2 Creating, Submitting and Merging

Once the necessary files have been set up as described previously, the user can run the `MadGraph5@NLO` section using the Makefile as follows;

- `cd` into `job_submission/MadGraph/`
- Edit the Makefile, refer to Figure A.16, Figure A.17.
- Run "`make create-jobs`" - this will create a directory in "jobs" and copy all necessary files to it.
- Then run "`make submit-jobs` - this will either submit your jobs to your cluster or run them locally.
- Once your jobs have finished running run "`make merge-jobs` to combine all of the output together into a final output CSV file.

By default the vast majority of `MadGraph5@NLO` files are discarded due to the high volume of files created for every point that is run. By default only the "`run_01_tag_1_banner.txt`" file and the compressed "`unweighted_events.lhe.gz`" file are kept. **If you wish to keep different files or additional ones**, please refer to subsection A.2.6 for details.

A.2.4 Troubleshooting Installation

If problems are encountered when installing the first step to fixing this is to establish *where* in the process the problems are occurring. If the install is getting stuck on a specific package within `Magellan` then it is a good idea to check if it is possible to install that package manually by itself. If it is not and the package is either `2HDMC`, `HiggsBounds` or `HiggsSignals`, then we recommend referencing the relevant manual for that package. Something the authors came across on multiple occasions when installing on a cluster was the issue that while Makefiles may explicitly state which version of a compiler to use (e.g. `GCC` 11.1.0), and even when the module for this has been loaded, the cluster may use the default one instead for part of it (e.g. using the wrong version of `GFortran` compared to the specified `GCC` 11.1.0). In this instance it may be helpful to explicitly export each compiler individually before compiling (e.g. `export FC=gcc/11.1.0`).

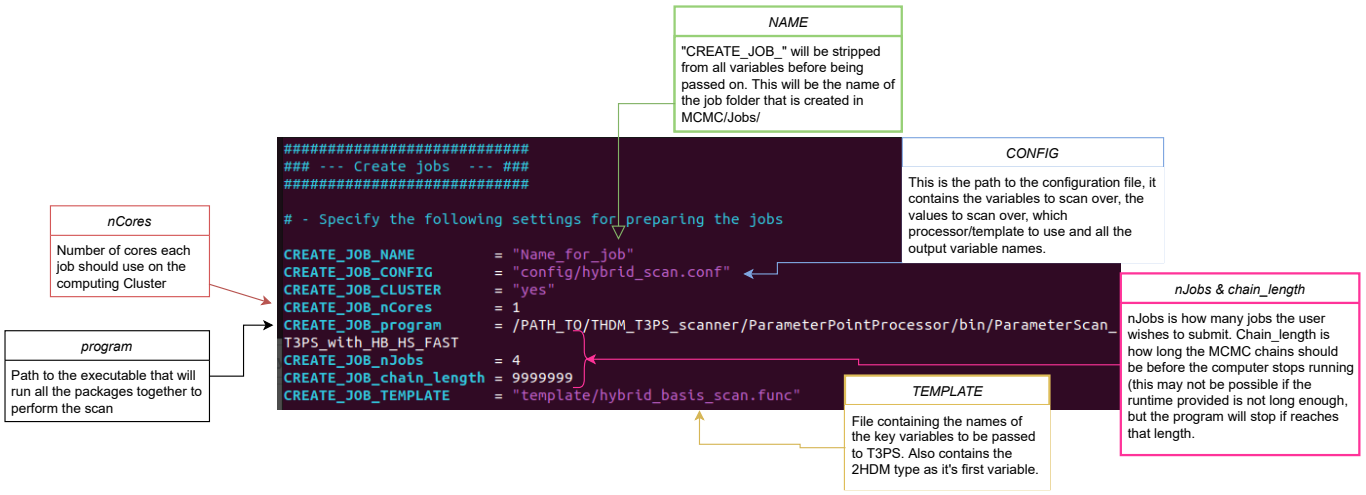
If the user does not have access to the recommended compilers then it is also possible that the versions they are using are not compatible. Anything more recent than the recommended versions `should` be compatible so these can be tried. Anything after `GCC` 11.1.0 v10.4.0 may also work.

A.2.5 MCMC Jobs

The Makefile in the MCMC directory is key to using the scanner, it is split into roughly four parts, the '`create-jobs`' shown in Figure A.19, '`submit-jobs`' shown in Figure A.20, '`merge-jobs`' shown in Figure A.21 and a section which exports variables and calls necessary utility scripts. The first three perform much as their names suggest, "`make create-jobs`" will create job directories, copy templates into them, replace any placeholder variables with those given in the Makefile and prepare everything needed to run a job. "`make submit-jobs`" will submit a job named in the Makefile alongside the cluster resources the user gives there. This can be altered for either `SLURM` or `Torque` based systems by commenting/uncommenting as appropriate. Finally "`make merge-jobs`" will combine the different chains produced into one file in either .dat, .CSV or .hdf format. Please examine the related diagrams for a clearer idea of what each variable the user can change is used for.

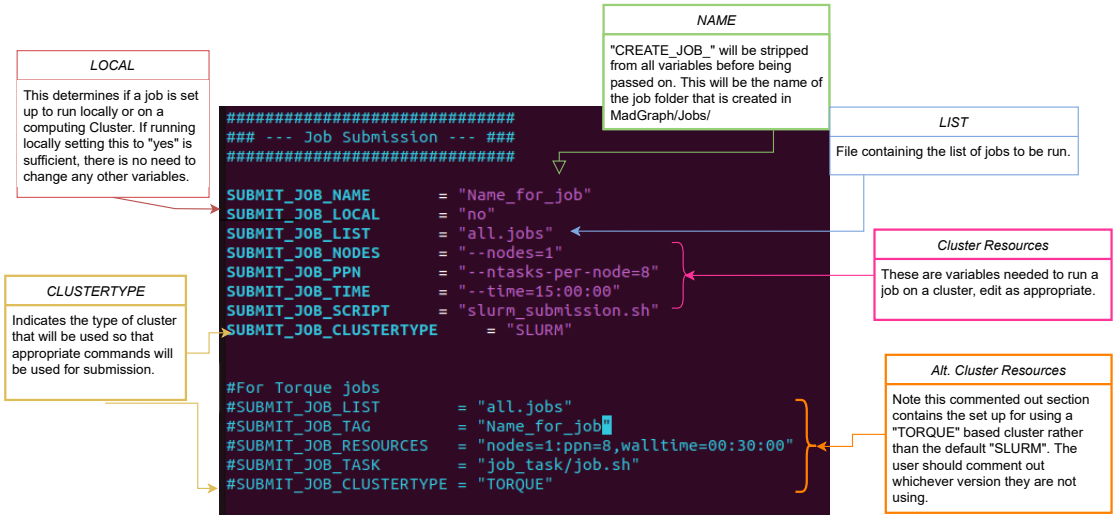
To use the Makefile directly use the following steps:

- `cd` into `job_submission/MCMC/`
- Check which config file matches your desired parametrisation in the `config` directory, open this to edit the parameter ranges you will be scanning over. Check the description for Figure A.6 for details of accepted formats.
- Open the Makefile in your preferred editor and ensure you have commented out the line '`#include utils/settings.mk`'. Otherwise old variable settings will be read in and your job may not work.



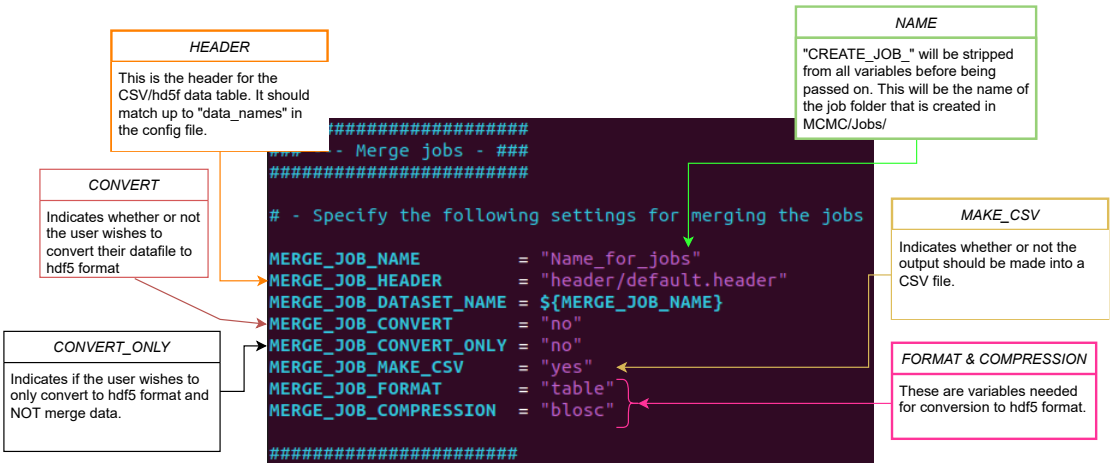
,

A.19 `create-jobs`' file for MCMC type runs



A.20 `submit-jobs.sh` file for MCMC type runs

- Set the name for your job in the "`CREATE_JOB_NAME`" parameter. Use this same name for "`SUBMIT_JOB_NAME`" (unless you intend to submit a different job before this one).
- Edit "`CREATE_JOB_CONFIG`" to equal the appropriate .conf file, i.e. the one with the parametrisation you wish to use. Be sure to edit this config file to have the correct values for the input variables.
- Check that "`CREATE_JOB_CLUSTER`" is set to "yes"



A.21 'merge-jobs.sh' file for MCMC type runs

A.2.5.1 Application of constraints

Constraints are applied at the merging stage by the python file "dat_to_csv.py" which is called by "merge-jobs.sh"; both of which are in "MCMC/utils/". By default these are set to correspond to the most recent bounds found by the authors on values for the κs, it is very simple to add further constraints here using the pre-existing ones, or to apply them at a later stage, for example when plotting. In both cases the user can apply any methods they might normally use when dealing with a pandas DataFrame, for which there are extensive tutorials online.

A.2.6 MadGraph Jobs

A.2.6.1 Steps to create MadGraph5@NLO proc_card

- Navigate to "THDM_T3PS_scanner/packages/MG5_aMC_v3_1_0/bin" and start MadGraph5@NLO with the following "python2 mg5_aMC"
- You will be prompted if you want to upgrade your MadGraph5@NLO version, **do not answer 'yes'!** The version that comes with Magellan is the latest that runs on python2, upgrading will break Magellan as it will switch MadGraph5@NLO to python3.
- Then load your chosen model with "import model YOUR_MODEL"
- Define any particle groups you need "define W = w+ w-"
- Generate your process "generate particles in > particles out"
- Then save the process output savefolder"

- Run the process, in order to edit parameter values "`launch`". You will be prompted about several options and should edit these as appropriate for your process, see the `MadGraph5@NLO` documentation for more detail.

- Next you will be prompted to edit the parameter and run cards if you wish. You should change the values of any variables you varied when scanning your parameter space. E.g. if you changed the masses of particles from point to point, these should be edited. The value does not matter at this stage, editing will simply add the needed command to the proc_card for later use.

- Now quit `MadGraph5@NLO` with "`quit()`" and navigate to "`MG5_aMC_v3_1_0/bin/savefolder/Cards`, there will be a file called "`run_card.dat`" which is the one we are interested in. Copy this proc_card to "`MadGraph_jobs/MG_utils/mg_proccards`" with a name that will distinguish it for you e.g. YOUR_MODEL_proccard.dat .

- "`CREATE_JOB_NAME`" should be set to the name of your job. This will be the name of the output directory created inside "jobs".

- "`CREATE_INPUT_DATA`" will be a CSV file containing the parameter points to be fed to `MadGraph5@NLO`. Usually this will just be an output CSV file from the `MCMC` scanner section.

- The order of variables in your "`INPUT_DATA`" header is of upmost importance and must match the '`csv_looper.sh`' file given to the Makefile to call (this is the variable '`CREATE_JOB_LOOPER`'). This is set to be a template in '`MG_utils/Looper`', this reads values into variable names according to their position along the line being read. This means that if the header for your data and the data order in the chosen '`csv_looper.sh`' file do not match up **the wrong values will be assigned to variables!**.

- "`csv_looper.sh`" is set to read files that have come straight from the MCMC scanner, using the Type-II model "2HDMtII_NLO", so the order it is reading in variables will be the same, i.e. , $m_H \to 2$, $m_{H^\pm} \to 3$, $m_A \to 4$, $\tan(\beta) \to 6$ & $\sin(\beta - \alpha) \to 7$. Create a new copy of "`csv_looper.sh`" for each model you use.Then replace the variable names in line 10 of "`csv_looper.sh`" to the ones appropriate to your data and make sure that the numbers in line 49 match those new variables. This may need to be done if using a CSV from somewhere other than the scanner as the names of variables and positions may differ. It is recommended that the user makes a copy of a template "`csv_looper.sh`" for this purpose as they may wish to have several copies corresponding to different jobs they wish to run. This reduces the chances of errors if they continuously edit one file back and forth.

- "CREATE_JOB_LOCAL" Should be set to "no" to create cluster jobs and "yes" to run on the local machine only.
- "CREATE_JOB_PROCESS_NAME" is the name that will be used to name output CSV files.
- "CREATE_JOB_PROCESS" is used to enter the process you wish to generate in MadGraph5@NLO. Where you might normally enter "generate $bg > t \sim W - h$" you would enter, "b\ g\ \> \ t~ \ W- \ h". Note the use of "\" before each space and instance of ">", this is necessary to prevent bash misinterpreting the input.
- "CREATE_JOB_nCores" this is the number of computer cores to run your job/s on.
- "CREATE_JOB_nJobs" this is the number of parallel jobs to split your CSV data between.
- "CREATE_JOB_Card_Editor", indicates which 'card_editor' file should be used to edit the values in the proc_card being submitted to MadGraph5@NLO
- "CREATE_JOB_BASECARD" is the basic template being used to create MadGraph5@NLO proc_cards for your job.
- "CREATE_JOB_RIPPER" is the script that will be used to extract output data from MadGraph5@NLO.
- "CREATE_JOB_MGVARi_LABEL" these variables should be the column names of the variables from the data CSV that the user wants to use to compare points when recombining MadGraph5@NLO output data to the original input. Essentially a minimum of 3 is needed to ensure that values are correctly matched up, otherwise cross-sections etc. may be assigned to incorrect points that happen to share one or two values for variables. The examples given in the toolbox use 4 of these 'comparison variables' as an extra precaution

A.2.6.2 Retained MadGraph5@NLO Output

Due to the storage space that would be required, most of the output files from MadGraph5@NLO are deleted before the "Looper" file moves on to the next parameter-point. By default the only files that are retained are the "run_01_tag_1_banner.txt" and the compressed "unweighted_events.lhe.gz". To change this simply edit the "Data_Ripper" file to be used. Specifically, inside the function "get_uncomp_file" add a new variable for the files you wish to keep akin to "path_to_file" and "path_to_lhe", then add a line to copy them as the two existing files are. The relevant section of code is shown in Figure A.22,

```
path_to_file = proc_out_dir + "/Events/run_01/run_01_tag_1_banner.txt"
path_to_lhe = proc_out_dir + "/Events/run_01/unweighted_events.lhe.gz"

if os.path.exists(path_to_file) and os.path.exists(path_to_lhe) :
    print("Banner-path found")
    # name of new folder for storing
    file_storage = "JOB_DIR_" + "/Processed_results/" + str(Process)
    # storing the data for later use
    if not os.path.exists(file_storage):
        os.makedirs(file_storage)

    shutil.copy(path_to_file, file_storage)# copying the file
    shutil.copy(path_to_lhe, file_storage)# moving the lhe file
    txt_stored = str(file_storage) + "/run_01_tag_1_banner.txt"
    # adding the path for the copied file to our list
    uncomped_files.append(txt_stored)
```

A.22 A section of a "Data_Ripper" script, showing how the files that are kept from the `MadGraph5@NLO` output are copied into a new location before the originals are removed. The user should add a line defining the path to the additional files they want to retain a line above or below "path_to_file" or "path_to_lhe"; then for these additional files add "shutil.copy(additionalfile, file_storage)" for each of them below/above the pre-existing shutil lines.

A.3 Under The Hood

A.3.1 MCMC Jobs

The Makefile in the MCMC directory is key to using the scanner, it is split into roughly four parts, the '`create-jobs`' shown in Figure A.19, '`submit-jobs`' shown in Figure A.20, '`merge-jobs`' shown in Figure A.21 and a section which exports variables and calls necessary utility scripts. The first three perform much as their names suggest, "`make create-jobs`" will create job directories, copy templates into them, replace any placeholder variables with those given in the Makefile and prepare everything needed to run a job. "`make submit-jobs`" will submit a job named in the Makefile alongside the cluster resources the user gives there. This can be altered for either `SLURM` or `Torque` based systems by commenting/uncommenting as appropriate. Finally "`make merge-jobs`" will combine the different chains produced into one file in either .dat, .CSV or .hdf format. Please examine the related diagrams for a clearer idea of what each variable the user can change is used for.

To use the Makefile directly use the following steps:

- `cd` into `job_submission/MCMC/`
- Check which config file matches your desired parametrisation in the `config` directory, open this to edit the parameter ranges you will be scanning over. Check the description for Figure A.6 for details of accepted formats.

- Open the Makefile in your preferred editor and ensure you have commented out the line '`#include utils/settings.mk`'. Otherwise old variable settings will be read in and your job may not work.
- Set the name for your job in the "`CREATE_JOB_NAME`" parameter. Use this same name for "`SUBMIT_JOB_NAME`" (unless you intend to submit a different job before this one).
- Edit "`CREATE_JOB_CONFIG`" to equal the appropriate .conf file, i.e. the one with the parametrisation you wish to use. Be sure to edit this config file to have the correct values for the input variables.
- Check that "`CREATE_JOB_CLUSTER`" is set to "yes"

A.3.1.1 Application of constraints

Constraints are applied at the merging stage by the python file "`dat_to_csv.py`" which is called by "`merge-jobs.sh`"; both of which are in "MCMC/utils/". By default these are set to correspond to the most recent bounds found by the authors on values for the κs, it is very simple to add further constraints here using the pre-existing ones, or to apply them at a later stage, for example when plotting. In both cases the user can apply any methods they might normally use when dealing with a `pandas` DataFrame, for which there are extensive tutorials online.

A.3.2 MadGraph Jobs

A.3.2.1 Steps to create `MadGraph5@NLO` proc_card

- Navigate to "`THDM_T3PS_scanner/packages/MG5_aMC_v3_1_0/bin`" and start `MadGraph5@NLO` with the following "`python2 mg5_aMC`"
- You will be prompted if you want to upgrade your `MadGraph5@NLO` version, **do not answer 'yes'!** The version that comes with `Magellan` is the latest that runs on `python2`, upgrading will break `Magellan` as it will switch `MadGraph5@NLO` to `python3`.
- Then load your chosen model with "`import model YOUR_MODEL`"
- Define any particle groups you need "`define W = w+ w-`"
- Generate your process "`generate particles in > particles out`"
- Then save the process `output savefolder`"
- Run the process, in order to edit parameter values "`launch`". You will be prompted about several options and should edit these as appropriate for your process, see the `MadGraph5@NLO` documentation for more detail.

- Next you will be prompted to edit the parameter and run cards if you wish. You should change the values of any variables you varied when scanning your parameter space. E.g. if you changed the masses of particles from point to point, these should be edited. The value does not matter at this stage, editing will simply add the needed command to the proc_card for later use.

- Now quit `MadGraph5@NLO` with "`quit()`" and navigate to "`MG5_aMC_v3_1_0/bin/savefolder/Cards`, there will be a file called "`run_card.dat`" which is the one we are interested in. Copy this proc_card to "`MadGraph_jobs/MG_utils/mg_proccards`" with a name that will distinguish it for you e.g. YOUR_MODEL_proccard.dat .

- "`CREATE_JOB_NAME`" should be set to the name of your job. This will be the name of the output directory created inside "jobs".

- "`CREATE_INPUT_DATA`" will be a CSV file containing the parameter points to be fed to `MadGraph5@NLO`. Usually this will just be an output CSV file from the `MCMC` scanner section.

- The order of variables in your "`INPUT_DATA`" header is of upmost importance and must match the '`csv_looper.sh`' file given to the Makefile to call (this is the variable '`CREATE_JOB_LOOPER`'). This is set to be a template in '`MG_utils/Looper`', this reads values into variable names according to their position along the line being read. This means that if the header for your data and the data order in the chosen '`csv_looper.sh`' file do not match up **the wrong values will be assigned to variables!**.

- "`csv_looper.sh`" is set to read files that have come straight from the MCMC scanner, using the Type-II model "2HDMtII_NLO", so the order it is reading in variables will be the same, i.e. , $m_H \to 2$, $m_{H^\pm} \to 3$, $m_A \to 4$, $\tan(\beta) \to 6$ & $\sin(\beta - \alpha) \to 7$. Create a new copy of "`csv_looper.sh`" for each model you use.Then replace the variable names in line 10 of "`csv_looper.sh`" to the ones appropriate to your data and make sure that the numbers in line 49 match those new variables. This may need to be done if using a CSV from somewhere other than the scanner as the names of variables and positions may differ. It is recommended that the user makes a copy of a template "`csv_looper.sh`" for this purpose as they may wish to have several copies corresponding to different jobs they wish to run. This reduces the chances of errors if they continuously edit one file back and forth.

- "`CREATE_JOB_LOCAL`" Should be set to "no" to create cluster jobs and "yes" to run on the local machine only.

- "`CREATE_JOB_PROCESS_NAME`" is the name that will be used to name output CSV files.

- "CREATE_JOB_PROCESS" is used to enter the process you wish to generate in MadGraph5@NLO. Where you might normally enter "generate $bg > t \sim W - h$" you would enter, "b\ g\ \> \ t~ \ W- \ h". Note the use of "\" before each space and instance of ">", this is necessary to prevent bash misinterpreting the input.
- "CREATE_JOB_nCores" this is the number of computer cores to run your job/s on.
- "CREATE_JOB_nJobs" this is the number of parallel jobs to split your CSV data between.
- "CREATE_JOB_Card_Editor", indicates which 'card_editor' file should be used to edit the values in the proc_card being submitted to MadGraph5@NLO
- "CREATE_JOB_BASECARD" is the basic template being used to create MadGraph5@NLO proc_cards for your job.
- "CREATE_JOB_RIPPER" is the script that will be used to extract output data from MadGraph5@NLO.
- "CREATE_JOB_MGVARi_LABEL" these variables should be the column names of the variables from the data CSV that the user wants to use to compare points when recombining MadGraph5@NLO output data to the original input. Essentially a minimum of 3 is needed to ensure that values are correctly matched up, otherwise cross-sections etc. may be assigned to incorrect points that happen to share one or two values for variables. The examples given in the toolbox use 4 of these 'comparison variables' as an extra precaution

A.3.2.2 Troubleshooting MadGraph5@NLO job merge-jobs

- If a user is having difficulty running "make merge-jobs" they should perform some simple checks.
  - As a MadGraph - type job runs it creates a 1-point CSV file for every data point run. These are split between each of the different sub-jobs which were submitted to the cluster. These CSV files are saved inside "MadGraph/jobs/your_job/job_00X/Data_Files/" (here 00X indicates the number for whichever 'sub'-job the user is interested in, 000, 006 etc). A python script, "Data_coallator.py" is inside every "job_00X" directory and this runs at the end of each job submitted to the cluster in order to concatenate the 1-point CSV files into one.
  - When "merge-jobs.sh" runs it seeks out these CSV files so they must be present for it's successful completion. If a user has submitted too many points to run in the allotted time then it will NOT have been created.
  - This is easily remedied by cd'ing into the relevant 'job_00X directory and running the "Data_coallator.py" for the 'CREATE_JOB_PROCESS_NAME'

originally entered, e.g. "bg_twh1". Looking inside "`Data_Files`" will reveal that each 1-point CSV is named with this '`CREATE_JOB_PROCESS_NAME`' variable followed by the tan value and sin value at this point, see Figure A.10 for an example.

– Inside the job_002 directory seen in the pathway for Figure A.10, if "`Data_Files/bg_twh1.CSV`" did not already exist then (e.g. if a job had been unable to finish running all points in time) a user could run the following "`python Data_coallator.py "Data_Files" "bg_twh1"`" (assuming they remembered to source '`env.sh`" in the top directory as users must **every time** `Magellan` is used) and this would create "`bg_twh1.CSV`" for them.

- If a user did find their job had not completed in time and wants to run again with just the points that where not run previously it is straightforward to obtain these.

  – After creating a combined CSV for each '`job_00X`' directory a user can check the number of lines in a combined CSV and compare this to the corresponding "split" CSV file. E.g. job_000's combined CSV corresponds to split_aa.CSV, job_004's combined CSV corresponds to "`split_ae.CSV`" etc.

  – Then the user can simply use their preferred text editor, skip the header row and then delete the number of lines that ran successfully. These shortened files can be concatenated using `pandas`'concatenate' function and this new CSV will can be used to create a new job to finish off the previous one[13].

  – There is a script within "`MadGraph/MG_utils`" that can make harvesting the leftover points slightly quicker, "`row_split.py`" contains a function; '`row_dropper`', which takes an input CSV, number of lines and output CSV name as input variables. It will drop X lines starting from the top (the header of the CSV is preserved, however) and saves the new file to the given output name.

  – A user may have several sub-jobs that did not have sufficient run time and again these can be concatenated with pandas before being used as the "`CREATE_JOB_INPUT_DATA`" variable for a new job in the Makefile.

A.3.2.3 Retained `MadGraph5@NLO` Output

Due to the storage space that would be required, most of the output files from `MadGraph5@NLO` are deleted before the "Looper" file moves on to the next

[13]The user ought to consider how many points were run in the first job when setting the time and number of jobs for this second one. Different processes will take different times for `MadGraph5@NLO` to calculate depending on their complexity.

parameter-point. By default the only files that are retained are the "run_01_tag_1_banner.txt" and the compressed "unweighted_events.lhe.gz". To change this simply edit the "Data_Ripper" file to be used. Specifically, inside the function "`get_uncomp_file`" add a new variable for the files you wish to keep akin to "path_to_file" and "path_to_lhe", then add a line to copy them as the two existing files are. The relevant section of code is shown in Figure A.22,

Bibliography

[1] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al. Measurements of the higgs boson production and decay rates and constraints on its couplings from a combined atlas and cms analysis of the lhc pp collision data at $\sqrt{s}$=7 and 8 tev. *Journal of High Energy Physics*, 2016(8), Aug 2016. ISSN 1029-8479.

[2] E. Accomando, M. Chapman, A. Maury, and S. Moretti. Below-threshold CP-odd Higgs boson search via A to Zh at the LHC. *Physics Letters B*, 818: 136342, 2021. ISSN 0370-2693.

[3] E. Accomando, C. Byers, D. Englert, J. Hays, and S. Moretti. LHC data interpretation within the 2HDM type II via a new analysis toolkit. *Phys. Rev. D*, 105:115004, Jun 2022.

[4] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. The minimal composite Higgs model. *Nuclear Physics B*, 719(1–2):165 – 187, 2005. ISSN 0550-3213.

[5] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Computer Physics Communications*, 185(8):2250–2300, Aug 2014. ISSN 0010-4655.

[6] J. Alwall, R. Frederix, S. Frixione, and et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014. ISSN 1029-8479.

[7] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014.

[8] Mayumi Aoki, Shinya Kanemura, and Osamu Seto. Neutrino Mass, Dark Matter, and Baryon Asymmetry via TeV-Scale Physics without Fine-Tuning. *Phys. Rev. Lett.*, 102:051805, Feb 2009.

[9] Peter Athron, Csaba Balazs, Torsten Bringmann, Andy Buckley, Marcin Chrząszcz, Jan Conrad, Jonathan M. Cornell, Lars A. Dal, Hugh Dickinson, and et al. GAMBIT: the global and modular beyond-the-standard-model inference tool. *The European Physical Journal C*, 77(11), Nov 2017. ISSN 1434-6052.

[10] Peter Athron, Csaba Balazs, Torsten Bringmann, Andy Buckley, Marcin Chrząszcz, Jan Conrad, Jonathan M. Cornell, Lars A. Dal, Hugh Dickinson, and et al. GAMBIT: the global and modular beyond-the-standard-model inference tool. *The European Physical Journal C*, 77(11), Nov 2017. ISSN 1434-6052.

[11] Peter Athron, Csaba Balázs, Torsten Bringmann, Andy Buckley, Marcin Chrząszcz, Jan Conrad, Jonathan M. Cornell, Lars A. Dal, Joakim Edsjö, and et al. A global fit of the MSSM with GAMBIT. *The European Physical Journal C*, 77(12), Dec 2017. ISSN 1434-6052.

[12] ATLAS and CMS. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, May 2015.

[13] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1 – 29, 2012. ATLAS Collaboration.

[14] ATLAS Collaboration. Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *JHEP*, 03:174, 2018.

[15] ATLAS Collaboration. Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. *Phys. Rev. D*, 101:012002, Jan 2020.

[16] Roberto Ruiz de Austri, Roberto Trotta, and Leszek Roszkowski. A Markov chain Monte Carlo analysis of the CMSSM. *Journal of High Energy Physics*, 2006(05):002–002, May 2006.

[17] M. Baak, M. Goebel, J. Haller, and et al. Updated status of the global electroweak fit and constraints on new physics. *The European Physical Journal C*, 72(5), May 2012.

[18] Baak, M. and Cúth, J. and Haller, J. and Hoecker, A. and et al. The global electroweak fit at NNLO and prospects for the LHC and ILC. *The European Physical Journal C*, 74(9):3046, 2014. ISSN 1434-6052.

[19] Henning Bahl, Thomas Biekötter, Sven Heinemeyer, and et al. HiggsTools: BSM scalar phenomenology with new versions of HiggsBounds and HiggsSignals. 10 2022.

[20] Phillipp Basler, Pedro M. Ferreira, Margarete Mühlleitner, and Rui Santos. High scale impact in alignment and decoupling in two-Higgs doublet models. *Phys. Rev.*, D97(9):095024, 2018.

[21] P. Bechtle, K. Desch, and P. Wienemann. Fittino, a program for determining MSSM parameters from collider observables using an iterative method. *Computer Physics Communications*, 174(1):47–70, Jan 2006. ISSN 0010-4655.

[22] Philip Bechtle, Klaus Desch, Mathias Uhlenbrock, and Peter Wienemann. Constraining SUSY models with Fittino using measurements before, with and beyond the LHC. *The European Physical Journal C*, 66(1-2):215–259, Jan 2010. ISSN 1434-6052.

[23] Philip Bechtle, Torsten Bringmann, Klaus Desch, and et al. Constrained supersymmetry after two years of LHC data: a global view with Fittino. *Journal of High Energy Physics*, 2012(6), Jun 2012. ISSN 1029-8479.

[24] Philip Bechtle, Oliver Brein, Sven Heinemeyer, and et al. HiggsBounds – 4: Improved Tests of Extended Higgs Sectors against Exclusion Bounds from LEP, the Tevatron and the LHC. *Eur. Phys. J.*, C74(3):2693, 2014.

[25] Philip Bechtle, Sven Heinemeyer, Oscar Stål, Tim Stefaniak, and Georg Weiglein. *HiggsSignals*: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC. *Eur. Phys. J.*, C74(2), 2014.

[26] Philip Bechtle, Sven Heinemeyer, Oscar Stål, Tim Stefaniak, and Georg Weiglein. HiggsSignals: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC. *The European Physical Journal C*, 74(2):1–40, 2014. ISSN 1434-6052.

[27] Philip Bechtle, Sven Heinemeyer, Tobias Klingl, Tim Stefaniak, Georg Weiglein, and Jonas Wittbrodt. HiggsSignals-2: probing new physics with precision Higgs measurements in the LHC 13 TeV era. *The European Physical Journal C*, 81(2), Feb 2021. ISSN 1434-6052.

[28] Belle Collaboration. Measurement of the inclusive $B \to X_{s+d}\gamma$ branching fraction, photon energy spectrum and HQE parameters, 2016. URL https://arxiv.org/abs/1608.02344.

[29] Florian U. Bernlochner, , Marcin Chrząszcz, Lars A. Dal, and et al. FlavBit: a GAMBIT module for computing flavour observables and likelihoods. *The European Physical Journal C*, 77(11), Nov 2017.

[30] Jérémy Bernon, John F. Gunion, Howard E. Haber, Yun Jiang, and Sabine Kraml. Scrutinizing the alignment limit in two-Higgs-doublet models. II. m_H=125 GeV. *Phys. Rev. D*, 93(3):035027, 2016.

[31] Jérémy Bernon and Béranger Dumont. Lilith: a tool for constraining new physics from Higgs measurements. *The European Physical Journal C*, 75(9), Sep 2015. ISSN 1434-6052.

[32] Enrico Bertuzzo, Tirtha Sankar Ray, Hiroshi de Sandes, and Carlos A. Savoy. On composite two Higgs doublet models. *Journal of High Energy Physics*, 2013 (5):1–23, 2013. ISSN 1029-8479.

[33] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv e-prints*, January 2017. .

[34] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo, 2018.

[35] Bokeh Development Team. *Bokeh: Python library for interactive visualization*, 2014. URL http://www.bokeh.pydata.org.

[36] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Physics Reports*, 516(1–2):1 – 102, 2012. ISSN 0370-1573. Theory and phenomenology of two-Higgs-doublet models.

[37] Alessandro Broggio, Eung Jin Chun, Massimo Passera, and et al. Limiting two-Higgs-doublet models. *Journal of High Energy Physics*, 2014(11):58, Nov 2014. ISSN 1029-8479.

[38] O. Buchmueller, R. Cavanaugh, A. De Roeck, and et al. Likelihood functions for supersymmetric observables in frequentist analyses of the CMSSM and NUHM1. *The European Physical Journal C*, 64(3):391–415, Oct 2009. ISSN 1434-6052.

[39] O. Buchmueller, R. Cavanaugh, D. Colling, and et al. Implications of initial LHC searches for supersymmetry. *The European Physical Journal C*, 71(5), May 2011. ISSN 1434-6052.

[40] O. Buchmueller, R. Cavanaugh, A. De Roeck, and et al. Supersymmetry in light of 1/fb of LHC data. *The European Physical Journal C*, 72(2), Feb 2012. ISSN 1434-6052.

[41] O. Buchmueller, R. Cavanaugh, A. De Roeck, and et al. Higgs and supersymmetry. *The European Physical Journal C*, 72(6), Jun 2012. ISSN 1434-6052.

[42] O. Buchmueller, R. Cavanaugh, A. De Roeck, M. J. Dolan, J. R. Ellis, H. Flächer, S. Heinemeyer, G. Isidori, J. Marrouche, D. Martínez Santos, and et al. The CMSSM and NUHM1 after LHC Run 1. *The European Physical Journal C*, 74(6), Jun 2014. ISSN 1434-6052.

[43] C. Byers, D. Englert, S. Jain, S. Moretti, and E Olaiya. Magellan - a new analysis toolkit. In Preparation.

[44] Ciara Byers, Shubhani Jain, Stefano Moretti, and Emmanuel Olaiya. Exploring SM-like Higgs Boson Production in Association with Single-Top at the LHC Within a 2HDM, 2023. URL https://arxiv.org/abs/2303.09225.

[45] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.

[46] Cacciari, Matteo and Salam, Gavin P. and Soyez, Gregory. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012. .

[47] Allen Caldwell, Daniel Kollár, and Kevin Kröninger. BAT – The Bayesian analysis toolkit. *Computer Physics Communications*, 180(11):2197–2209, Nov 2009.

[48] Junjie Cao, Peihua Wan, Lei Wu, and Jin Min Yang. Lepton-specific two-Higgs-doublet model: Experimental constraints and implication on Higgs phenomenology. *Phys. Rev. D*, 80:071701, Oct 2009.

[49] Alejandro Celis, Javier Fuentes-Martín, and Hugo Serôdio. Effective aligned 2HDM with a DFSZ-like invisible axion. *Physics Letters B*, 737:185 – 190, 2014. ISSN 0370-2693. Effective aligned 2HDM with a DFSZ-like invisible axion.

[50] Amit Chakraborty, Srinandan Dasmahapatra, Henry Day-Hall, Billy Ford, Shubhani Jain, Stefano Moretti, Emmanuel Olaiya, and Claire Shepherd-Themistocleous. Revisiting jet clustering algorithms for new Higgs Boson searches in hadronic final states. *Eur. Phys. J. C*, 82(4):346, 2022.

[51] Amit Chakraborty, Srinandan Dasmahapatra, Henry Day-Hall, Billy Ford, Shubhani Jain, and Stefano Moretti. Fat b-jet analyses using old and new clustering algorithms in new Higgs boson searches at the LHC. *Eur. Phys. J. C*, 83(4):347, 2023.

[52] Jung Chang, Kingman Cheung, Jae Lee, and Chih-Ting Lu. Probing the top-yukawa coupling in associated higgs production with a single top quark. *Journal of High Energy Physics*, 2014, 03 2014. .

[53] Serguei Chatrchyan and et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012.

[54] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2π Decay of the K_2^0 Meson. *Phys. Rev. Lett.*, 13:138–140, 1964.

[55] CMS. Measurement of the Higgs boson production in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81:378, 2021. .

[56] CMS2015. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$TeV . *Journal of Instrumentation*, 10(06):P06005–P06005, Jun 2015.

[57] John C. Collins, Davison E. Soper, and George Sterman. Factorization for short distance hadron-hadron scattering. *Nuclear Physics B*, 261:104–142, 1985. ISSN 0550-3213.

[58] Eric Conte and Benjamin Fuks. Confronting new physics theories to LHC data with MadAnalysis 5. *International Journal of Modern Physics A*, 33:1830027, 10 2018. .

[59] Eric Conte, Benjamin Fuks, and Guillaume Serret. MadAnalysis 5, a user-friendly framework for collider phenomenology. *Computer Physics Communications*, 184(1):222–256, 2013. ISSN 0010-4655.

[60] Andrew Curtis and Anthony Lomax. Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, 66(2):372–378, 01 2001. ISSN 0016-8033.

[61] Stefania De Curtis, Luigi Delle Rose, Stefano Moretti, and Kei Yagyu. Supersymmetry versus Compositeness: 2HDMs tell the story. *Physics Letters B*, 786:189–194, Nov 2018.

[62] Stefania De Curtis, Luigi Delle Rose, Stefano Moretti, and Kei Yagyu. A concrete composite 2-Higgs doublet model. *JHEP*, 12:051, 2018.

[63] G. D'Agostini. Bayesian reasoning in high-energy physics: Principles and applications. 7 1999. .

[64] Xinchen Dai, Miroslav Saur, Yiduo Shang, Xueting Yang, and Yanxi Zhang. CP Violation in Baryon Decays at LHCb. *Symmetry*, 15(2), 2023. ISSN 2073-8994.

[65] Jorge de Blas, Marco Ciuchini, Enrico Franco, Satoshi Mishima, Maurizio Pierini, Laura Reina, and Luca Silvestrini. The Global Electroweak and Higgs Fits in the LHC era. In *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2017)*, volume 314, page 467, 2018. .

[66] de Blas, Jorge and Ciuchini, Marco and Franco, Enrico and Mishima, Satoshi and Pierini, Maurizio and Reina, Laura and Silvestrini, Luca. Electroweak precision constraints at present and future colliders. *PoS*, ICHEP2016:690, 2017. .

[67] Céline Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. UFO - The Universal FeynRules Output. *Computer Physics Communications*, 183(6):1201–1214, Jun 2012. ISSN 0010-4655.

[68] Nilendra G. Deshpande and Ernest Ma. Pattern of symmetry breaking with two Higgs doublets. *Phys. Rev. D*, 18:2574–2576, Oct 1978.

[69] Abdelhak Djouadi. The anatomy of electroweak symmetry breaking Tome II: The Higgs bosons in the Minimal Supersymmetric Model. *Physics Reports*, 459 (1–6):1 – 241, 2008. ISSN 0370-1573.

[70] Claude Durh, Michel Herquet, and Celine Degrande. 2HDMtII_NLO. *FeynRules*, 2018. URL https://feynrules.irmp.ucl.ac.be/attachment/wiki/2HDM/2HDMtII_NLO.tar.gz.

[71] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

[72] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons, Aug 1964.

[73] David Eriksson, Johan Rathsman, and Oscar Stal. 2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual. *Comput. Phys. Commun.*, 181:189–205, 2010.

[74] David Eriksson, Johan Rathsman, and Oscar Stal. 2HDMC - two-Higgs-doublet model calculator. *"Comput. Phys. Commun."*, 181(4):833–834, 2010. ISSN 00104655.

[75] Marco Farina, Christophe Grojean, Fabio Maltoni, Ennio Salvioni, and Andrea Thamm. Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson. *JHEP*, 05:022, 2013. .

[76] Marco Farina, Maxim Perelstein, and Nicolas Rey-Le Lorier. Higgs couplings and naturalness. *Physical Review D*, 90(1), Jul 2014.

[77] P. M. Ferreira, Rui Santos, John F. Gunion, and Howard E. Haber. Probing wrong-sign Yukawa couplings at the LHC and a future linear collider. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 89(11), 2014. ISSN 15502368.

[78] Pedro M. Ferreira, Stefan Liebler, and Jonas Wittbrodt. $pp \to A \to Zh$ and the wrong-sign limit of the two-Higgs-doublet model. *Phys. Rev.*, D97(5):055008, 2018.

[79] Y. Fukuda, T. Hayakawa, E. Ichihara, and et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998.

[80] Radouane Gannouji. Introduction to Electroweak Baryogenesis. *Galaxies*, 10(6): 116, 2022. .

[81] Murray Gell-Mann. Symmetries of Baryons and Mesons. *Phys. Rev.*, 125: 1067–1084, Feb 1962.

[82] Charles J. Geyer and et al. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2011. ISBN 9781420079425.

[83] Charles J. Geyer and Elizabeth A. Thompson. Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *JASA*, 90(431):909–920, 1995.

[84] I. F. Ginzburg and I. P. Ivanov. Tree-level unitarity constraints in the most general two Higgs doublet model. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 72(11):1–6, 2005. ISSN 15507998.

[85] Sheldon L. Glashow and Steven Weinberg. Natural conservation laws for neutral currents. *Phys. Rev. D*, 15:1958–1965, Apr 1977.

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

[87] M. C. Gonzalez-Garcia and C. Peña Garay. Four-neutrino oscillations at sno. *Phys. Rev. D*, 63:073013, Mar 2001.

[88] W. Grimus, L. Lavoura, O.M. Ogreid, and P. Osland. A Precision constraint on multi-Higgs-doublet models. *J. Phys. G*, 35:075001, 2008.

[89] W. Grimus, L. Lavoura, O.M. Ogreid, and P. Osland. The Oblique parameters in multi-Higgs-doublet models. *Nucl. Phys. B*, 801:81–96, 2008.

[90] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. ISSN 2050-3911. 083C01.

[91] J.F. Gunion. *The Higgs Hunter's Guide*. CRC Press, 2018. ISBN 9780429976070.

[92] John F. Gunion, Howard E. Haber, Gordon L. Kane, and Sally Dawson. *The Higgs Hunter's Guide*. Perseus, 2000. ISBN 073820305X. ISBN: 073820305X.

[93] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[94] Howard E. Haber and Oscar Stal. New LHC benchmarks for the CP-conserving two-Higgs-doublet model. *European Physical Journal C*, 75(10):1–38, 2015. ISSN 14346052.

[95] Bruce Hajek. Cooling Schedules for Optimal Annealing. *Mathematics of Operations Research*, 13(2):311–329, 1988. ISSN 0364765X, 15265471.

[96] J. Haller, , A. Hoecker, R. Kogler, K. Mönig, T. Peiffer, and J. Stelzer. Update of the global electroweak fit and constraints on two-Higgs-doublet models. *The European Physical Journal C*, 78(8), Aug 2018.

[97] J. Haller, A. Hoecker, R. Kogler, and et al. Update of the global electroweak fit and constraints on two-Higgs-doublet models. *The European Physical Journal C*, 78(8), Aug 2018.

[98] Xiao-Fang Han, Hong-Xin Wang, and Lei Wang. Revisiting wrong sign Yukawa coupling of type II two-Higgs-doublet model in light of the recent LHC data. *Chinese Phys. C*, 44:073101, 07 2020.

[99] Robert V. Harlander, Stefan Liebler, and Hendrik Mantler. SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM. *Comput. Phys. Commun.*, 184:1605–1617, 2013.

[100] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970. ISSN 00063444.

[101] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[102] G.'t Hooft. Renormalizable Lagrangians for massive Yang-Mills fields. *Nuclear Physics B*, 35(1):167–188, 1971. ISSN 0550-3213.

[103] Shinya Kanemura, Takahiro Kubota, and Eiichi Takasugi. Lee-Quigg-Thacker bounds for Higgs boson masses in a two-doublet model. *Physics Letters B*, 313 (1-2):155–160, 1993. ISSN 03702693.

[104] Khalil, Shaaban and Moretti, Stefano. *Standard Model Phenomenology*. CRC Press, 6 2022. ISBN 978-1-138-33643-8.

[105] Jihn E. Kim. Light pseudoscalars, particle physics and cosmology. *Physics Reports*, 150(1):1–177, 1987. ISSN 0370-1573.

[106] P. Ko, Yuji Omura, and Chaehyun Yu. Dark matter and dark force in the type-I inert 2HDM with local $U(1)_H$ gauge symmetry. *Journal of High Energy Physics*, 2014(11):1–29, 2014. ISSN 1029-8479.

[107] Archil Kobakhidze, Lei Wu, and Jason Yue. Anomalous Top-Higgs Couplings and Top Polarisation in Single Top and Higgs Associated Production at the LHC. *JHEP*, 10:100, 2014. .

[108] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[109] Sabine Kraml, Tran Quang Loc, Dao Thi Nhung, and Le Duc Ninh. Constraining new physics from Higgs measurements with Lilith: update to LHC Run 2 results. *SciPost Physics*, 7(4), Oct 2019. ISSN 2542-4653.

[110] John K. Kruschke. *Doing Bayesian Data Analysis (Second Edition)*. Academic Press, Boston, second edition edition, 2015. ISBN 978-0-12-405888-0.

[111] Zoltán Kunszt, J W Stirling, A Ballestrero, and et al. Determination of the mass of the W boson, 1996. URL https://cds.cern.ch/record/296923.

[112] Remi Lafaye, Tilman Plehn, and Dirk Zerwas. SFITTER: SUSY parameter analysis at LHC and LC. 4 2004.

[113] M Romero Lamas. CP violation in beauty and charm at LHCb. *Journal of Physics: Conference Series*, 1526(1):012007, Apr 2020.

[114] T. Lancaster and S. Blundell. *Quantum Field Theory for the Gifted Amateur*. OUP Oxford, 2014. ISBN 9780199699322.

[115] E. Marinari and G. Parisi. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters*, 19(6):451, Jul 1992.

[116] Vinzenz Maurer. T3PS: tool for parallel processing in parameter scans. 2015. URL http://arxiv.org/abs/1503.01073.

[117] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[118] Mikolaj Misiak and Matthias Steinhauser. Weak radiative decays of the B meson and bounds on $m_{H^\pm}$ in the Two-Higgs-Doublet Model. *The European Physical Journal C*, 77(3):201, Mar 2017. ISSN 1434-6052.

[119] J. Mrazek, A. Pomarol, R. Rattazzi, M. Redi, J. Serra, and A. Wulzer. The other natural two Higgs doublet model. *Nuclear Physics B*, 853(1):1 – 48, 2011. ISSN 0550-3213.

[120] Siavash Neshatpour and Farvah Mahmoudi. Flavour Physics Phenomenology with SuperIso. *PoS*, CompTools2021:010, 2022.

[121] Kiran Ostrolenk and Olivier Mattelaer. Speeding up MadGraph5_aMC@NLO. *Eur Phys J. C*, 2021. .

[122] Emmanuel A. Paschos. Diagonal neutral currents. *Phys. Rev. D*, 15:1966–1972, Apr 1977.

[123] Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. OUP Oxford, 2013. ISBN 9780199671229.

[124] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.

[125] Antonio Pich and Paula Tuzón. Yukawa alignment in the two-higgs-doublet model. *Phys. Rev. D*, 80:091702, Nov 2009.

[126] Apostolos Pilaftsis and Carlos E.M Wagner. Higgs bosons in the minimal supersymmetric standard model with explicit CP violation. *Nuclear Physics B*, 553(1):3–42, 1999. ISSN 0550-3213.

[127] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Zh. Eksp. Teor. Fiz.*, 53:1717–1725, 1967.

[128] Filip Rajec, Wei Su, Martin White, and Anthony G. Williams. Exploring the 2HDM with Global Fits in GAMBIT. *EPJ Web of Conferences*, 245:06022, 2020. ISSN 2100-014X.

[129] A. Salam and J.C. Ward. Electromagnetic and weak interactions. *Physics Letters*, 13(2):168–171, 1964. ISSN 0031-9163.

[130] Prasenjit Sanyal. Limits on the charged Higgs parameters in the two Higgs doublet model using CMS $\sqrt{s}$ =13 TeV results. *The European Physical Journal C*, 79(11), Nov 2019.

[131] Oliver Schulz, Frederik Beaujean, Allen Caldwell, Cornelius Grunwald, Vasyl Hafych, Kevin Kröninger, Salvatore La Cagnina, Lars Röhrig, and Lolian Shtembari. BAT.jl: A Julia-Based Tool for Bayesian Inference. *SN Computer Science*, 2(3):210, Apr 2021. ISSN 2661-8907.

[132] Michele Selvaggi, J. Favereau, C. Delaere, Pavel Demin, A. Giammanco, Vincent Lemaitre, and Alexandre Mertens. DELPHES 3: A modular framework for fast-simulation of generic collider experiments. *Journal of Physics: Conference Series*, 523:012033, 06 2014.

[133] A. M. Sirunyan, A. Tumasyan, W. Adam, and et al. Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13$TeV. *Phys. Rev. D*, 99:092005, May 2019.

[134] Albert M Sirunyan et al. Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(5):421, 2019.

[135] Albert M Sirunyan et al. Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81(4):378, 2021. .

[136] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852–867, 2008. ISSN 0010-4655.

[137] STAN Development team. STAN Modeling Language Users Guide and Reference Manual. https://mc-stan.org, 2023. VERSION 2.31.

[138] Jean-Luc R. Stevens, Philipp Rudiger, and James A. Bednar. HoloViews: Building Complex Visualizations Easily for Reproducible Science. In Kathryn Huff and James Bergstra, editors, *Proceedings of the 14th Python in Science Conference*, pages 59–66, 2015.

[139] David M. Straub. flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond, 2018.

[140] C Strege, G Bertone, D.G Cerdeño, and et al. Updated global fits of the cMSSM including the latest LHC SUSY and Higgs searches and XENON100 data. *Journal of Cosmology and Astroparticle Physics*, 2012(03):030–030, Mar 2012. ISSN 1475-7516.

[141] C Strege, G Bertone, F Feroz, and et al. Global fits of the cMSSM and NUHM including the LHC Higgs discovery and new XENON100 constraints. *Journal of Cosmology and Astroparticle Physics*, 2013(04):013–013, Apr 2013. ISSN 1475-7516.

[142] The ATLAS Collaboration. Combined measurements of Higgs boson production and decay using up to $80fb^{-1}$ of proton-proton collision data at $\sqrt{s} = 13TeV$ collected with the ATLAS experiment, Jan 2020. ISSN 2470-0029.

[143] The CMS Collaboration. Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV, May 2019. ISSN 1434-6052.

[144] The Magellan Collaboration. Interactive dashboards. https://pprc.qmul.ac.uk/projects/magellan/2HDM/.

[145] Danny van Dyk, F. Beaujean, and et al. EOS: a software for flavor physics phenomenology. *Eur. Phys. J. C*, 82(6):569, 2022.

[146] M. Veltman. Limit on mass differences in the Weinberg model. *Nuclear Physics B*, 123(1):89–99, 1977. ISSN 0550-3213.

[147] Lei Wang, Hong-Xin Wang, and Xiao-Fang Han. Revisiting wrong sign Yukawa coupling of type II two-Higgs-doublet model in light of recent LHC data . *Chinese Physics C*, 44(7):073101, Jul 2020.

[148] Wang, Lei and Han, Xiao-Fang. A light pseudoscalar of 2HDM confronted with muon g-2 and experimental constraints. *Journal of High Energy Physics*, 2015 (5):39, May 2015. ISSN 1029-8479.

[149] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. .

[150] A Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 2004. ISBN 978-0691140346.