

LU TP 20-29
June 2020

**Using Neural Networks to Probe the Parameter Space of
a 3HDM with a $U(1) \times \mathbb{Z}_2$ Flavor Symmetry**

Ludvig Siwe

Department of Astronomy and Theoretical Physics, Lund University

Master thesis supervised by Roman Pasechnik



LUND
UNIVERSITY

Abstract

Probing the physical regions in large parameter spaces of typical Standard Model (SM) extensions can be a very difficult computational task. In this thesis project, a new framework has been developed that utilises well-known Machine Learning (ML) techniques in the form of neural networks trained by a genetic algorithm. This framework is rather generic and designed to explore new physics model parameter spaces with a large number of dimensions, implementing a given set of theoretical and experimental constraints, in a time-efficient and smart way. The ML framework has been applied for analysis of a large parameter space in a recently proposed Three Higgs Doublet Model (3HDM) with a $U(1) \times \mathbb{Z}_2$ flavor symmetry implementing theoretical constraints on tree-level unitarity and boundedness from below, as well as the experimental bounds on oblique corrections. We have developed an inversion procedure that enables us to use the scalar boson masses, mixing angles and off-alignment parameters as inputs in our ML framework. This lets us to use measured values of the SM-like Higgs boson mass (within errors) and couplings in the near-alignment regime, as well as possible experimental bounds on masses of additional scalar bosons. A similar inversion algorithm has also been implemented in the quark sector, enabling us to take the measured values of quark masses and mixing angles (within errors) as inputs randomized within the experimental uncertainties. Our ML implementation makes an important step towards an efficient and detailed exploration of large parameter spaces of new physics models highly constrained by precision experimental bounds.

Populärvetenskaplig sammanfattning

Partikelfysikens standardmodell är en av de mest exakta modellerna inom fysiken. Den beskriver interaktionerna mellan fundamentala partiklar som kvarkar och leptoner, den beskriver hur Higgspartikeln ger massa till andra partiklar och den har lyckats förena tre av universums fyra fundamentala krafter: elektromagnetism, svag växelverkan och stark kärnkraft. Tyvärr så finns det frågor som inte ens standardmodellen kan svara på. "Varför finns det fler partiklar än antipartiklar?" och "Vad är mörk materia?" är bara två exempel. För att försöka svara på de här frågorna måste fysiker undersöka nya modeller, som ofta bygger ut standardmodellen, och testa vilka numeriska värden som parametrar i dem här modellerna måste ha för att fungera med de många mätningar som redan gjorts i t.ex. partikelacceleratorer. De här nya modellerna kan dock bli väldigt komplicerade och när de blir mer komplicerade blir det svårare att numeriskt testa om vilka parametrar som fungerar bra, dagens datorer är inte kraftfulla nog. Det är inte ovanligt att man behöver flera tiotal ihopkopplade datorer som räknar i veckor för att hitta resultat.

För att försöka göra de här beräkningarna snabbare så har maskininläring använts i det här projektet. Maskininläring handlar om att utveckla metoder som tränar datorer att lösa olika problem. Ett stort område inom maskininläring är artificiella neuronnät, en metod som är baserad i att försöka replikera en biologisk hjärna med matematik. I en människohjärna finns över 80 miljarder neuroner som är ihopkopplade med ungefär 150 biljoner synapser. Detta stora nätverk av neuroner är hur människor kan bland annat minnas händelser och lösa problem. För att komma åt den här inlärningsförmågan så skapades artificiella neuronnät, med matematiska modeller av neuroner som är ihopkopplade till nätverk. Dessa nätverk lär sig genom att man ger dem ett problem som man vill att dem ska lösa, man låter dem försöka och man belönar nätverket om det gör rätt eller straffar det om det gör fel.

Här har ett program skapats som kan träna dessa artificiella neuronnät för att hitta bra värden på parametrarna i en typ av utvidgning av standardmodellen kallad tre-Higgs-dublettmodeller, eller förkortat, 3HDMs. I standardmodellen finns bara en Higgspartikel, men i en 3HDM finns ytterligare åtta Higgspartiklar som möjligtvis kan svara på obesvarade frågor och visa vägen till ny, oupptäckt fysik.

Contents

1	Introduction	4
2	A brief outline of the Standard Model	7
2.1	Higgs sector	7
2.2	Yukawa sector	9
3	Flavoured Three Higgs Doublet Model	12
3.1	$U(1) \times \mathbb{Z}_2$ flavor symmetry	12
3.2	Higgs sector	13
3.2.1	Scalar potential and minimisation	13
3.2.2	Higgs basis and alignment	14
3.2.3	Inversion	16
3.2.4	Further constraints on the Higgs sector	18
3.3	Yukawa sector	21
3.3.1	Lagrangian and mass textures	21
3.3.2	Inversion	23
3.3.3	FCNC couplings	24
4	An overview of Artificial Neural Networks	26
5	Methods and Algorithms	29
5.1	Neural Network	29
5.2	Genetic Algorithm	31
5.2.1	Training stop criteria	33
5.3	Sampling points	33
5.4	Incremental learning	34
5.4.1	Convergence criterion	35
5.5	Input and target files	35
5.6	Modular training	36
6	Numerical analysis	36
6.1	Parameter search specifications	37
6.2	Results	37
7	Conclusion	46
A	Higgs mechanism: a toy model	48
B	Inverted procedure equations list	49
C	Oblique parameters	52

1 Introduction

The Standard Model (SM) provides the best theoretical description of experimentally observed spectra of particles and their interactions at microscopic scales that exists today. It has succeeded in combining three of the four fundamental forces of nature while providing remarkable consistency with some of the most accurate measurements in physics. With the recent discovery of the Higgs particle [1, 2], the SM content was completed and it was experimentally proven that the Higgs mechanism is responsible for spontaneous breaking of the electroweak (EW) $SU(2)_L \times U(1)_Y$ symmetry of the SM Lagrangian, effectively generating the physical masses of the fermions and vector bosons.

Even with such major success, the SM still lacks a solid theoretical explanation for several experimentally observed facts (for a review of SM strengths and weaknesses, see e.g. Ref. [3]). For example, there is no particle in the SM framework that could play a role of dark matter evident from cosmological and astrophysical observations. The SM can also not adequately explain the origin of the observed baryon asymmetry in the universe i.e. why there is more matter than antimatter. While the EW baryogenesis mechanism enables violation of the baryon number in non-perturbative (sphaleron) transitions [4], the SM neither contains the required amount of \mathcal{CP} violation nor provides the necessary mechanism for a strong departure from thermal equilibrium (e.g. via strong first-order phase transitions), both required for efficient generation of the baryon asymmetry in the universe. Besides, certain features in the SM fermion spectra have no explanation within the SM framework, in particular, the observed mass and mixing hierarchies. This so-called *flavor problem*, is also often considered a drawback. For example, the top quark appears to be heavier than the up quark by roughly five orders of magnitude and heavier than the upper bound on neutrino mass by over eleven orders of magnitude. An explanation for such pronounced hierarchies is one of the desired properties of a consistent Beyond the SM (BSM) framework. These are some of the typical reasons to explore possible extensions of the SM where one or several such issues are addressed. On the other hand, the precision and the coverage of typical measurements designed to search for new physics implications at particle colliders, most notably at the Large Hadron Collider (LHC), grows tremendously over time, setting more and more stringent bounds on viable parameter spaces in popular BSM scenarios. While the available space for new physics gets reduced, it makes it more and more technically challenging to reveal remaining (possibly, fine-tuned) regions in multi-dimensional parameter spaces of such models that are still consistent with the wealth of precision experimental constraints.

Despite of relatively tight bounds on Higgs boson couplings to the SM gauge bosons and heavy fermions [5], the Higgs sector still remains among the least constrained ones in the SM. While the minimality of the Higgs sector postulated by the SM is not in immediate contradiction with the measurements at the current level of experimental precision, it is not manifestly required by the existing data. The known viable extensions of the SM typically incorporate extended Higgs sectors obtained by adding extra Higgs weak ($SU(2)_L$) doublets and/or complex scalar weak-singlets to the SM. Such models can emerge in the low-energy limit of high-scale Grand unified field theories, see e.g. Ref. [6].

The SM extensions featuring non-minimal Higgs sectors with extra Higgs doublets in analogy to fermion generations in the SM provide a fruitful playground for constructing successful BSM scenarios (for a detailed review, see e.g. Refs. [7, 8, 9]). For example, even simple extensions of the SM Higgs sector can dramatically enhance the strength of the EW phase transition in the early universe and provide extra sources for \mathcal{CP} violation, both relevant

for an efficient baryogenesis. Moreover, fermion mass and mixing patterns relate to specific structures of the Yukawa interactions, and hence are directly linked to the properties of the Higgs sector. A popular class of models with additional family symmetries acting in the space of fermion and Higgs generations [10] may help in explaining the observed fermion mass and mixing patterns, including those in the spectrum of light neutrinos. In addition, they provide rich collider and flavor physics phenomenology considered to be an important source of constraints on parameter spaces in such models.

One of the most simple and well explored scenarios of new physics, the so-called Two Higgs Doublet Model (or 2HDM), contains one extra Higgs doublet. This model was initially proposed to motivate the observed \mathcal{CP} violation [11, 12], but has since become one of the most popular and well-studied new physics frameworks [8]. The scalar potential of a generic 2HDM (i.e. not containing any additional symmetries) contains a lot of free parameters. Also, it implies the existence of potentially large tree-level Flavor Changing Neutral Currents (FCNCs) that, however, are required to be suppressed to comply with measurements. A more predictive BSM framework could be constructed by imposing an additional symmetry in the extended Higgs and/or fermion sectors yielding many new realisations with a constrained parametric freedom [10]. In this way, one could explain the observed hierarchies in the fermion sectors of the SM and avoid large FCNCs [13].

A necessary and sufficient condition for absence of the tree-level FCNCs is that all the Yukawa matrices are simultaneously diagonalisable in the mass basis. One simple realisation is to enforce that the charged right-handed fermions with the same charge couple to only one of the Higgs doublets as was proposed in Refs. [14, 15]. This so-called Glashow-Weinberg model is built by imposing a discrete symmetry such as \mathbb{Z}_2 in a 2HDM that leads to natural flavour conservation (NFC) [16] and helps to avoid such problems as basis dependence [17]. An alternative method to suppress the FCNCs is to require an alignment of the Yukawa couplings between the two Higgs doublets in the flavor space as is realised in Ref. [18].

By adding two extra Higgs doublets to the SM Higgs sector, one arrives at a less explored class of Three Higgs Doublet Models (3HDMs) which, together with an imposed flavor symmetry, offer further opportunities for addressing the flavor problem (for a few particular realisations, see e.g. Refs. [19, 20, 21, 22]). Like in the case of 2HDMs, an extra flavor symmetry may help in turning a rather complicated generic 3HDM into a predictive New Physics framework and also in suppressing the size of FCNC contributions coming from additional scalar states. Instead of entirely forbidding the tree-level FCNCs, one could secure an approximate flavor conservation by suppressing them strongly enough to fit the measurements. In the case of 2HDM, such a scenario was for the first time implemented by Branco, Grimus and Lavoura (BGL) [23] where the tree-level FCNCs in the quark sector are suppressed by combinations of small elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [24, 25]. In the original BGL-2HDM scenario [23] (see also Ref. [26]), this is realised due to the presence of a family symmetry in the Lagrangian being broken by the neutral Higgs field vacuum expectation values (VEVs).

A BGL-like realisation in the case of 3HDMs has been developed and its phenomenological analysis has been performed for the first time in Ref. [27]. This is achieved by using a $U(1) \times \mathbb{Z}_2$ flavor symmetry in the Higgs and quark sectors providing a very specific structure of the quark Yukawa matrices¹. In what follows, we refer to this model as the *BGL-3HDM scenario*. In

¹In the lepton sector, lepton flavor violation has not been considered for simplicity, such that the lepton Yukawa matrices are assumed to be diagonal. We follow the same simplifying assumption in what follows.

this project, we continue this line of research by considering the most generic soft breaking parameters of the $U(1) \times \mathbb{Z}_2$ flavor symmetry in the Higgs potential and by going beyond the Higgs alignment limit imposed previously in Ref. [27]. To get a better control of the searches for viable regions in the parameter space, an inverted procedure has been developed, both in the scalar and quark sectors, where the Lagrangian parameters are calculated in terms of the physical masses and mixing angles such that the parameter scans are performed in the space of physical inputs (such as masses and mixing angles varied within the experimentally allowed ranges).

In such complex models as the considered BGL-3HDM, the time-efficient parameter fits to the theoretical (e.g. tree-level unitarity) and precision experimental (such as EW precision observables) constraints become a challenging problem. It is common to use a random scan to search for the physical allowed regions in parameter spaces, but even with only 20 parameters and two points on each axis, that turns out to $2^{20} \approx 10^6$ points, which however would not be nearly enough to explore the space well. When the number of parameters increases, the computational time necessary to find a good fit through a random scan typically become extremely large for a set of imposed stringent constraints. Even now, it is not uncommon to run such parameter scans on a large computer cluster for many days if not weeks to find any allowed ranges in the parameter space, and only very few points that pass the most relevant constraints may eventually show up (as e.g. in Ref. [19]). No doubts, measurements will become more and more precise in the future, making the parameter fits even more challenging. With this in mind, one would want to adopt a smarter algorithm for this purpose.

In this project, we build a new computational framework based upon a Machine Learning (ML) technique applied for a more sophisticated analysis of the BGL-3HDM parameter space. This technique is realised in the form of artificial neural networks implemented with a genetic algorithm that trains the networks on how to best transform the full input parameter space into a subspace that passes a considered subset of physical constraints. Neural networks have been proven very successful in many other fields such as automation and economy, but also for various purposes in particle physics (see e.g. Refs. [28, 29, 30]). The constraints implemented in this framework are the tree-level unitarity bounds, the EW precision constraints formulated in terms of the Peskin-Takeuchi parameters [31, 32] S, T and U as well as boundness from below constraints. Provided the BGL-like nature of the considered 3HDM, we have verified that many physical points generated by the ML implementation yield rather suppressed FCNC observables. The latter have been computed using the Python package `flavio` [33] through an interface with `SARAH` [34, 35, 36] and `SPheno` [37, 38] that are utilized to generate a set of Wilson coefficients for each potentially valid parameter point. Constraints on allowed Higgs masses and couplings from collider data have not been considered in this project, but need to be considered in further studies for a more proper phenomenological analysis.

The structure of this thesis is as follows. Section 2 defines the SM framework. Section 3 discusses the BGL-3HDM scenario with $U(1) \times \mathbb{Z}_2$ flavor symmetry used in this project. Section 4 briefly introduces neural networks and how they are trained. Section 5 discusses the specifics about the networks used in the project and also introduces the training procedure and the algorithms used in our implementation. Then, Section 6 presents the relevant input parameter ranges and the numerical results of the training and the corresponding ML parameter scans. Finally, a brief summary and an outlook for future studies are given in Section 7.

2 A brief outline of the Standard Model

In order to introduce notations and basic concepts behind the SM extensions let us first start with defining the SM framework in those details that are particularly relevant for further BGL-3HDM model-building and exploration.

2.1 Higgs sector

The scalar (or Higgs) sector of the SM Lagrangian invariant under $SU(2)_L \times U(1)_Y$ gauge transformation (that will be the main subject for further extensions) reads

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad \phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (2.1)$$

where ϕ is the complex $SU(2)_L$ doublet of Higgs fields (or the Higgs doublet), μ^2 and λ are the Higgs mass term and quartic interaction coupling – the only parameters of the SM Higgs potential $V(\phi)$. Above, the Higgs kinetic terms are conventionally written in terms of D_μ being the standard EW $SU(2)_L \times U(1)_Y$ covariant derivative defined as

$$D_\mu \phi = \left(\partial_\mu - ig' Y B_\mu - ig \frac{\vec{\tau}}{2} \cdot \vec{W}_\mu \right) \phi, \quad (2.2)$$

where g' and g are the $U(1)_Y$ and $SU(2)_L$ gauge couplings, respectively, $Y = 1/2$ is the hypercharge of the Higgs doublet, $\vec{\tau}$ and $\vec{W}_\mu \equiv \{W_\mu^1, W_\mu^2, W_\mu^3\}$ are the isotopic vectors of Pauli matrices and $SU(2)_L$ gauge bosons, respectively, B_μ is the $U(1)_Y$ gauge boson, and $\partial_\mu \equiv \partial/\partial x^\mu$ is the usual spacetime derivative.

Just as in the case of Abelian Higgs mechanism discussed in Appendix A, the condition $\mu^2 < 0$ makes the EW-symmetric vacuum unstable leading to a non-zero VEV of the Higgs doublet ϕ spontaneously breaking the EW symmetry as

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{e.m.}} \quad (2.3)$$

down to the $U(1)_{\text{e.m.}}$ symmetry of electromagnetism, with the conserved electric charge

$$Q = T^3 + Y \quad (2.4)$$

for a given state having the third component of the weak-isospin generator, T^3 , and hypercharge, Y . Indeed, upon an appropriate gauge transformation, the Higgs VEV v can be conveniently placed in the electrically neutral ϕ^0 component of the Higgs doublet, namely,

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1(x) + i\varphi_2(x) \\ v + h(x) + i\varphi_3(x) \end{pmatrix}, \quad \langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v = \sqrt{\frac{-\mu^2}{\lambda}}, \quad (2.5)$$

such that the Higgs vacuum manifestly preserves the electric charge conservation. In the SM, the three Goldstone states $\varphi_{1,2,3}$ emerge corresponding to three broken generators of the EW symmetry. When a gauge symmetry breaking is concerned like in the considered case, the Goldstone modes are always eaten up by the gauge bosons corresponding to the broken gauge symmetry generators and become their longitudinal polarisations and the bosons themselves acquire masses. Indeed, one can show that, through the kinetic terms in the Higgs

Lagrangian, φ_3 gets absorbed by Z^0 , while two complex conjugated combinations of $\varphi_{1,2}$ become longitudinal components of W^\pm .

Let us show that Z^0 and W^\pm are indeed mass eigenstates of the SM gauge sector, i.e. emerge as massive states in the physical spectrum upon the spontaneous symmetry breaking (SSB). The mass terms for the Z^0 and W^\pm vector bosons can be readily obtained by a direct substitution of the vacuum value of the Higgs doublet (2.5) into its kinetic term (2.1) (which is equivalent to taking the vacuum average of the kinetic term) yielding

$$(D_\mu \phi)^\dagger (D^\mu \phi) \xrightarrow{\text{SSB}} \left| \left(\partial_\mu - \frac{i}{2} g' B_\mu - \frac{i}{2} g \vec{\tau} \cdot \vec{W}_\mu \right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 \quad (2.6)$$

$$= \frac{v^2}{8} \left| \begin{pmatrix} g W_\mu^1 + i g W_\mu^2 \\ -g W_\mu^3 + g' B_\mu \end{pmatrix} \right|^2, \quad (2.7)$$

where the last step employs the fact that $v = \text{const.}$ Thus, we arrive at the gauge boson mass form that is readily diagonalised by switching from the gauge basis, $\{W_\mu^1, W_\mu^2, W_\mu^3, B_\mu\}$, to the mass basis, $\{W_\mu^+, W_\mu^-, Z_\mu^0, A_\mu\}$, as follows

$$W_\mu^\pm = \frac{1}{\sqrt{2}} [W_\mu^1 \mp i W_\mu^2], \quad Z_\mu^0 = \frac{1}{\sqrt{g'^2 + g^2}} [g W_\mu^3 - g' B_\mu], \quad A_\mu = \frac{1}{\sqrt{g'^2 + g^2}} [g' W_\mu^3 + g B_\mu],$$

with masses

$$m_W = \frac{gv}{2} = \frac{ev}{2 \sin \theta_W}, \quad m_Z = \frac{v}{2} \sqrt{g'^2 + g^2} = \frac{m_W}{\cos \theta_W}, \quad m_\gamma = 0, \quad (2.8)$$

respectively. Here, the weak mixing (or Weinberg) angle is found as $\tan \theta_W \equiv g'/g$, and $e = g \sin \theta_W$ is the absolute value of electron charge. The measured values of the vector and Higgs boson masses are [39]

$$\begin{aligned} m_W &= 80.379 \pm 0.012 \text{ GeV}, \\ m_Z &= 91.1876 \pm 0.0021 \text{ GeV}, \\ m_h &= \sqrt{2\lambda}v = 125.10 \pm 0.14 \text{ GeV} \end{aligned} \quad (2.9)$$

The measurement of $m_{Z,W}$ provides a rather precise value of the Weinberg angle (determined at the $Q^2 = m_Z^2$ scale and in a particular renormalization scheme called the $\overline{\text{MS}}$ scheme),

$$\sin^2 \theta_W(m_Z^2) = 0.23120 \pm 0.00015. \quad (2.10)$$

Together with the value of the fine structure constant measured extremely precisely at small momentum transfers [39],

$$\alpha = \frac{e^2}{4\pi}, \quad \alpha^{-1}(0) \simeq 137.036, \quad (2.11)$$

one can access the value of the Higgs VEV,

$$v = \frac{2m_W}{g} \simeq 246 \text{ GeV}, \quad (2.12)$$

that is widely used in practical calculations.

2.2 Yukawa sector

The Yukawa sector of the SM Lagrangian describes the interactions between the Higgs and chiral fermion fields,

$$-\mathcal{L}_Y = \bar{q}_L^k \Gamma^{kj} \phi d_R^j + \bar{q}_L^k \Delta^{kj} \tilde{\phi} u_R^j + \bar{l}_L^k \Pi^{kj} \phi e_R^j + \text{h.c.}, \quad \tilde{\phi} = i\tau_2 \phi^*, \quad (2.13)$$

where τ_2 is the second Pauli matrix, and $q_L^j = (u_L^j, d_L^j)^T$ and $l_L^j = (\nu_L^j, e_L^j)^T$ represent the left-handed quark and lepton $\text{SU}(2)_L$ doublets of j -th generation ($j = 1, 2, 3$) while u_R^j , d_R^j and e_R^j correspond to the j -th generation of right-handed up-type quark, down-type quark and lepton $\text{SU}(2)_L$ singlets, respectively. The matrices Γ , Δ and Π are the 3×3 , generally complex, Yukawa interaction matrices (or textures) in the space of fermion generations (in the down-quark, up-quark and lepton sectors, respectively) that encode all possible coupling constants between the Higgs boson and a given pair of chiral fermions from generations k and j . In Eq. (2.13), the sums over j and k indices have been left implicit.

After the EW symmetry gets spontaneously broken, the Yukawa Lagrangian (2.13) provides a consistent description of fermion masses and mixings, as well as the physical Higgs boson h interactions with Dirac fermions. Taking its vacuum average, the charge fermion mass terms take the following form

$$\begin{aligned} -\mathcal{L}_Y &\xrightarrow{\text{SSB}} (\bar{u}_L \quad \bar{d}_L)^k \Gamma^{kj} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} d_R^j + (\bar{u}_L \quad \bar{d}_L)^k \Delta^{kj} \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ 0 \end{pmatrix} u_R^j \\ &+ (\bar{\nu}_L \quad \bar{e}_L)^k \Pi^{kj} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} e_R^j + \text{h.c.} \\ &= \bar{d}_L^k M_d^{kj} d_R^j + \bar{u}_L^k M_u^{kj} u_R^j + \bar{e}_L^k M_e^{kj} e_R^j. \end{aligned} \quad (2.14)$$

where

$$M_d = \frac{v}{\sqrt{2}} \Gamma, \quad M_u = \frac{v}{\sqrt{2}} \Delta, \quad M_e = \frac{v}{\sqrt{2}} \Pi, \quad (2.15)$$

are the down, up and lepton mass matrices in the generations space, respectively, each containing nine complex entries in general. Transforming to the mass basis, each mass matrix becomes a real diagonal matrix by multiplying it on the left and right by appropriate unitary transformation matrices, $U^{-1} = U^\dagger$. For example, for the quark sector such a diagonalisation reads

$$M_u^{\text{diag}} \equiv U_{R,u}^{-1} M_u U_{L,u} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}, \quad M_d^{\text{diag}} \equiv U_{R,d}^{-1} M_d U_{L,d} = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} \quad (2.16)$$

in terms of the real physical up-type u, c, t and down-type d, s, b quark masses. Because of the linear relations in Eq. (2.15), diagonalisation of mass forms $M_{u,d}$ simultaneously diagonalises the corresponding Yukawa matrices, Δ and Γ , such that the Higgs couplings to $q\bar{q}$ appear to be automatically real and also flavour-diagonal in the quark mass basis. This means that there is no quark (and also lepton) flavor violation in the SM at tree level, so any possible FCNC reactions in the SM can only happen radiatively at a loop level (typically, at one loop), and hence are suppressed. A real eigenvalue of the Yukawa matrices for each q, e fermion

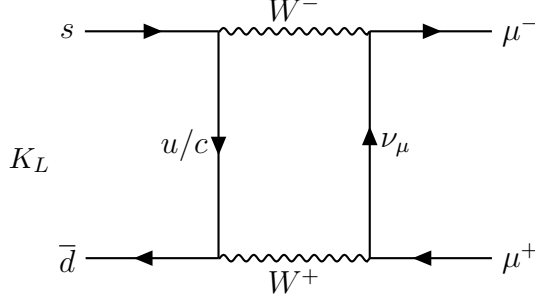


Figure 1: Feynman loop diagram for $K_L \rightarrow \mu^+ \mu^-$.

generation can thus be readily found in terms of the Higgs VEV v and the corresponding fermion mass, $m_{q,e}$, for example, $y_{q,e} = \sqrt{2}m_{q,e}/v$.

On top of loop suppression of FCNCs in the SM, an extra source of their smallness comes due to the Glashow-Iliopoulos-Maiani mechanism [40]. This mechanism shows that the FCNC-inducing loop diagrams involving the quarks of different flavors cancel in the limit of exact flavor symmetry. When flavor symmetry breaks down and quarks get different masses, the surviving contribution depends on the differences between the quark masses. For example, the total of contributions to the $K_L \rightarrow \mu^+ \mu^-$ decay amplitude shown in Fig. 1 from diagrams mediated by up and charm quarks is of order $g^4(m_c^2 - m_u^2)/m_W^2$, which contains an extra suppression factor compared to the contribution of the order g^4 for only one of those diagrams. For this reason, typical FCNC reactions are strongly suppressed in the SM in consistency with the measurements while new physics models often introduce new contributions to such FCNCs that must then be relatively small and highly constrained to comply with experiment. For a detailed review on typical FCNC observables, methodology and theoretical results, see e.g. Refs. [41, 42, 43].

In the framework of the SM, the W^\pm boson mediates the FCNC loop diagrams since the charged currents are the only ones that are capable of changing flavor (the neutral currents are automatically flavor diagonal). This happens due to the CKM matrix which is the only source of flavor and \mathcal{CP} violation in the quark sector of the SM².

The unitary CKM matrix V_{CKM} appears in the charged currents as follows

$$J_W^\mu = \frac{1}{\sqrt{2}} (\bar{u}_L \quad \bar{c}_L \quad \bar{t}_L) \gamma^\mu V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}, \quad V_{\text{CKM}} \equiv U_{L,u}^\dagger U_{L,d}. \quad (2.17)$$

In the so-called standard parameterisation, it is expressed in terms of three angles $\theta_{12}, \theta_{13}, \theta_{23}$ and one \mathcal{CP} violating phase δ_{13} as

$$V_{\text{CKM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{13}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.18)$$

where s_{ij} and c_{ij} are the shorthand notations for $\sin\theta_{ij}$ and $\cos\theta_{ij}$, respectively, and the hierarchy $s_{13} \ll s_{23} \ll s_{12} \ll 1$ has been experimentally observed.

²In the neutrino sector, the complex phases in the PMNS neutrino mixing are considered to be yet another source of \mathcal{CP} violation in the SM that is not discussed in this thesis.

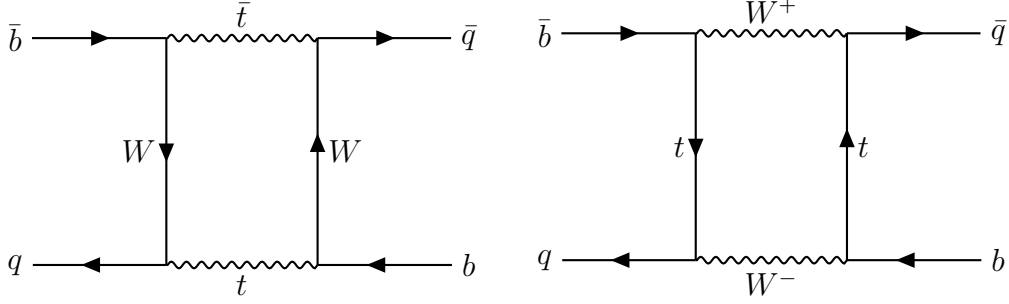


Figure 2: Dominant loop diagrams for $B_q^0 - \bar{B}_q^0$ where $(q = d, s)$.

In this thesis, we work with a different parametrization introduced by Wolfenstein [44],

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \quad (2.19)$$

where λ , A , ρ and η can be used to calculate the angles in the standard parametrization (2.18) through

$$s_{12} = \lambda, \quad s_{23} = A\lambda^2 \quad \text{and} \quad s_{13}e^{-i\delta_{13}} = A\lambda^3(\rho - i\eta). \quad (2.20)$$

The measured values for the Wolfenstein parameters can be found in the Particle Data Group review [39] and read

$$\begin{aligned} \lambda &= 0.22506 \pm 0.00050, \\ A &= 0.81 \pm 0.026, \\ \bar{\rho} &= 0.124_{-0.018}^{+0.019} \quad \text{and} \\ \bar{\eta} &= 0.356 \pm 0.011, \end{aligned} \quad (2.21)$$

where $\bar{\rho} = \rho(1 - \lambda^2/2 + \dots)$ and $\bar{\eta} = \eta(1 - \lambda^2/2 + \dots)$ such that an approximate form of V_{CKM} in Eq. (2.19) is retained up to an order of $\mathcal{O}(\lambda^4)$ reproducing all results in literature with a sufficiently good precision.

The FCNC processes that will be considered in this project are the branching ratios for rare decays like $K_L \rightarrow \mu^+\mu^-$ mentioned above as well as the $B - \bar{B}$ oscillation frequencies ΔM_s and ΔM_d . The latter are computed as

$$\Delta M_q = 2|M_{12}|, \quad (2.22)$$

where, since the oscillations are dominated by the box loop diagrams shown in Fig. 2, M_{12} can be approximated as [45]

$$M_{12} \simeq \frac{G_F^2 B_q f_B^2 m_B^q m_W^2}{12\pi^2} [(V_{\text{CKM}})_{tq}^* (V_{\text{CKM}})_{tb}]^2 S_0(m_t^2/m_W^2) \eta^B, \quad (2.23)$$

where G_F^2 is the Fermi constant, m_B^q is the mass of the B_q^0 meson (for $q = (d, s)$), B_q is the so-called bag parameter, f_B^2 is the B -meson weak decay constant, $S_0(m_t^2/m_W^2)$ is a known

function [46] that can be approximated by $0.784(m_t^2/m_W^2)^{0.76}$ and η^B is the QCD correction factor of order one. The $B - \bar{B}$ oscillation frequencies are experimentally constrained to be [39]

$$\begin{aligned}\Delta M_d &= (3.334 \pm 0.013) \cdot 10^{-13}, \\ \Delta M_s &= (1.1688 \pm 0.0014) \cdot 10^{-11},\end{aligned}\tag{2.24}$$

providing an important source of phenomenological constraints on new physics models.

3 Flavoured Three Higgs Doublet Model

Let us now consider an extended version of the SM, with an enlarged Higgs sector that contains three generations of scalar-doublets, ϕ_i ($i = 1, 2, 3$). On top of that, in order to comply with an approximate flavor conservation apparent in measurements, following the first footsteps of Ref. [27], we impose an additional horizontal $U(1) \times \mathbb{Z}_2$ global flavor symmetry acting in the space of fermion and Higgs generations. This symmetry is considered to be exact in the quark Yukawa sector and softly broken in the scalar potential. As will be discussed in more detail below, this turns the model into a phenomenologically viable framework sharing similar features to those of the well-known BGL scenario of 2HDM [23]. We refer readers to Refs. [19, 20, 21, 22] for detailed discussions of other possible 3HDM realisations and key results.

3.1 $U(1) \times \mathbb{Z}_2$ flavor symmetry

Based upon earlier work [27], we know that the global $U(1) \times \mathbb{Z}_2$ flavor symmetry of the Higgs and quark Yukawa sectors provides a very specific structure on Yukawa interactions and the scalar potential introduced below. After spontaneous EW symmetry breaking, it results in a minimal flavour violation scenario in the quark sector á la BGL [23] where the smallness of the tree-level FCNC contributions is directly correlated with the smallness of off-diagonal elements in the quark CKM mixing matrix. In order to achieve this, one adopts the following transformations under the $U(1) \times \mathbb{Z}_2$ flavor symmetry for each of the three Higgs doublets,

$$\begin{array}{ll} U(1) : & \mathbb{Z}_2 : \\ \phi_1 \rightarrow e^{i\alpha} \phi_1 & \phi_1 \rightarrow -\phi_1 \\ \phi_2 \rightarrow \phi_2 & \phi_2 \rightarrow \phi_2 \\ \phi_3 \rightarrow e^{i\alpha} \phi_3 & \phi_3 \rightarrow \phi_3. \end{array}\tag{3.1}$$

and, simultaneously, for the third quark generation

$$\begin{array}{ll} U(1) : & \mathbb{Z}_2 : \\ q_{L,3} \rightarrow e^{i\alpha} q_{L,3} & q_{L,3} \rightarrow -q_{L,3} \\ u_{R,3} \rightarrow e^{2i\alpha} u_{R,3} & u_{R,3} \rightarrow -u_{R,3} \\ d_{R,3} \rightarrow d_{R,3} & d_{R,3} \rightarrow -d_{R,3} \end{array}\tag{3.2}$$

while the other quark generations remain unaffected by this symmetry.

We should also mention that the leptonic fields are assumed to couple only to ϕ_1 in the Yukawa sector, for simplicity, which implies no lepton flavor violation in the current version. This can be achieved by assigning the following transformations to the leptonic fields

$$\begin{aligned} U(1) : \quad & l_{L,i} \rightarrow e^{i\alpha} l_{L,i} \\ & e_{R,i} \rightarrow e_{R,i} \end{aligned} \quad \mathbb{Z}_2 : \quad \begin{aligned} & l_{L,i} \rightarrow -l_{L,i} \\ & e_{R,i} \rightarrow e_{R,i} . \end{aligned} \quad (3.3)$$

$$(3.4)$$

for $i = 1, 2, 3$ labeling the lepton generations. In addition, it is also worth mentioning that we do not introduce right handed neutrinos in the current analysis, such that neutrinos are assumed to be massless in this model.

3.2 Higgs sector

3.2.1 Scalar potential and minimisation

The part of the 3HDM Higgs potential that is governed by the EW and flavor symmetries can be written as

$$\begin{aligned} V_0(\phi_1, \phi_2, \phi_3) = & \mu_1^2 (\phi_1^\dagger \phi_1) + \mu_2^2 (\phi_2^\dagger \phi_2) + \mu_3^2 (\phi_3^\dagger \phi_3) + \lambda_1 (\phi_1^\dagger \phi_1)^2 \\ & + \lambda_2 (\phi_2^\dagger \phi_2)^2 + \lambda_3 (\phi_3^\dagger \phi_3)^2 + \lambda_4 (\phi_1^\dagger \phi_1) (\phi_2^\dagger \phi_2) + \lambda_5 (\phi_1^\dagger \phi_1) (\phi_3^\dagger \phi_3) \\ & + \lambda_6 (\phi_2^\dagger \phi_2) (\phi_3^\dagger \phi_3) + \lambda_7 (\phi_1^\dagger \phi_2) (\phi_2^\dagger \phi_1) + \lambda_8 (\phi_1^\dagger \phi_3) (\phi_3^\dagger \phi_1) \\ & + \lambda_9 (\phi_2^\dagger \phi_3) (\phi_3^\dagger \phi_2) + \lambda_{10} \left\{ (\phi_1^\dagger \phi_3)^2 + \text{h.c.} \right\} , \end{aligned} \quad (3.5)$$

while the most general potential that softly breaks the $U(1) \times \mathbb{Z}_2$ flavor symmetry is given by

$$V_{\text{soft}}(\phi_1, \phi_2, \phi_3) = \mu_{12}^2 \phi_1^\dagger \phi_2 + \mu_{13}^2 \phi_1^\dagger \phi_3 + \mu_{23}^2 \phi_2^\dagger \phi_3 + \text{h.c.} . \quad (3.6)$$

V_{soft} therefore prevents the occurrence of a massless Goldstone scalar in the physical spectrum. Note, only the term that is proportional to μ_{23}^2 , which respects the \mathbb{Z}_2 part of the flavor symmetry, has been included in the first analysis of Ref. [27]. In this work, we incorporate all three terms in Eq. (3.6) thus also enabling a soft breaking of the \mathbb{Z}_2 symmetry. The total Higgs potential of the considered 3HDM is then given by the sum of the flavor-symmetric and flavor-breaking parts, $V = V_0 + V_{\text{soft}}$. In this work, we assume that all the parameters in the potential are real, i.e. the \mathcal{CP} symmetry is not broken explicitly.

Similarly to the SM, each doublet acquires a VEV upon spontaneous EW symmetry breaking and gets expanded around an extremum of the potential as follows

$$\phi_k = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} H_k^+ \\ v_k + h_k + i\eta_k \end{pmatrix} , \quad \langle \phi_k \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_k \end{pmatrix} , \quad (3.7)$$

in terms of three real Higgs VEVs, v_k , $k = 1, 2, 3$, charged Higgs fields, H_k^+ , neutral \mathcal{CP} -odd, η_k , and \mathcal{CP} -even, h_k , Higgs fields. Out of these 12 scalar degrees of freedom, 3 will become Goldstone modes showing up as massless states in the mass basis. They will be absorbed by the massive gauge bosons and hence will be discarded from the scalar mass spectrum leaving us

with 9 physical states (including one pseudo-Goldstone state corresponding to softly-broken flavor U(1) symmetry).

Then the standard procedure to identify physical regions in the parameter space is to find all the (global and local) minima of the scalar potential $V(\phi_k)$ by first solving a coupled set of three tadpole equations (extremum conditions)

$$\left. \frac{\partial V(\phi_k)}{\partial \phi_k} \right|_{\phi_k = \langle \phi_k \rangle} = 0, \quad (3.8)$$

and then requiring that the masses of all physical scalars are strictly positive for each such solution (minimisation conditions). For this purpose, it is convenient to trade the bilinear coefficients $\mu_{1,2,3}^2$ in favor of the three Higgs VEVs yielding

$$\mu_1^2 = -\frac{2\lambda_1 v_1^3 + v_1 v_3^2(2\lambda_{10} + \lambda_5 + \lambda_8) + v_1 v_2^2(\lambda_4 + \lambda_7) + 2\mu_{12}^2 v_2 + 2\mu_{13}^2 v_3}{2v_1}, \quad (3.9)$$

$$\mu_2^2 = -\frac{2\lambda_2 v_2^3 + v_1^2 v_2(\lambda_4 + \lambda_7) + v_2 v_3^2(\lambda_6 + \lambda_9) + 2\mu_{12}^2 v_1 + 2\mu_{23}^2 v_3}{2v_2}, \quad (3.10)$$

$$\mu_3^2 = -\frac{v_1^2 v_3(2\lambda_{10} + \lambda_5 + \lambda_8) + 2\lambda_3 v_3^3 + v_2^2 v_3(\lambda_6 + \lambda_9) + 2\mu_{13}^2 v_1 + 2\mu_{23}^2 v_2}{2v_3}. \quad (3.11)$$

3.2.2 Higgs basis and alignment

Note, the gauge boson mass terms

$$\frac{1}{4}[v_1^2 + v_2^2 + v_3^2]g^2 W_\mu^+ W^{-\mu} + \frac{1}{8}[v_1^2 + v_2^2 + v_3^2](g'^2 + g^2)Z_\mu Z^\mu \quad (3.12)$$

retain the same form as in the SM as long as

$$\sqrt{v_1^2 + v_2^2 + v_3^2} = v \simeq 246 \text{ GeV}. \quad (3.13)$$

This represents an important constraint on the allowed values of the Higgs VEVs that enables us to reparametrize the doublet VEVs in terms of two independent parameters $\beta_{1,2}$ as follows

$$\left. \begin{aligned} v &= \sqrt{v_1^2 + v_2^2 + v_3^2} \\ \tan(\beta_1) &= v_1/v_3 \\ \tan(\beta_2) &= v_2/v_{13} \end{aligned} \right\} \leftrightarrow \left\{ \begin{aligned} v_1 &= v \sin(\beta_1) \cos(\beta_2) \\ v_2 &= v \sin(\beta_2) \\ v_3 &= v \cos(\beta_1) \cos(\beta_2) \end{aligned} \right., \quad (3.14)$$

where $v_{13} = \sqrt{v_1^2 + v_3^2}$. Thus, the $\beta_{1,2}$ parameters can be considered as rotation angles to a new basis in the space of Higgs VEVs, $v_1 \rightarrow v$ and $v_2, v_3 \rightarrow 0$, in which only one Higgs doublet acquires a non-zero VEV, v . This new basis is called the Higgs basis which can be conveniently utilised as an intermediate basis between the gauge (or Lagrangian) and the mass (or physical) bases. An orthogonal matrix that performs such a “rotation” between the gauge and the Higgs bases reads

$$\mathcal{O}_\beta = \begin{pmatrix} \sin(\beta_1) \cos(\beta_2) & \sin(\beta_2) & \cos(\beta_1) \cos(\beta_2) \\ \cos(\beta_1) & 0 & -\sin(\beta_1) \\ \sin(\beta_1) \sin(\beta_2) & -\cos(\beta_2) & \cos(\beta_1) \sin(\beta_2) \end{pmatrix}, \quad (3.15)$$

which is one of the possible representations of such a basis transformation that we employ in our work. In order to realise the inversion procedure in the Higgs sector, i.e. to get the Lagrangian parameters in terms of physical masses and rotation angles, in what follows we need to determine all relevant relations between the physical (measured) masses and the Lagrangian parameters in terms of the rotation angles between the gauge and the Higgs bases and between the Higgs and the mass bases.

Besides the gauge boson mass terms, the Higgs kinetic Lagrangian contains the following interaction terms of the \mathcal{CP} -even Higgs states and the W^\pm bosons:

$$\frac{g^2 v}{2} W_\mu^+ W^{\mu-} \left(\frac{1}{v} \sum_{k=1}^3 v_k h_k \right), \quad (3.16)$$

Here, the following superposition of neutral scalar states emerges

$$H_0 = \frac{1}{v} \sum_{k=1}^3 v_k h_k, \quad (3.17)$$

which represents a scalar state that features the same tree-level coupling to the gauge bosons as those of the SM Higgs boson. Similarly, one can explicitly demonstrate that this would-be SM Higgs boson H_0 has the SM-like Yukawa couplings too, see Ref. [47].

In general, however, H_0 is not a physical mass eigenstate. Only in the so-called *Higgs alignment limit* corresponding to a particular configuration in the parameter space of the 3HDM, the Higgs-basis state H_0 becomes a mass eigenstate and thus gets completely aligned with the lightest \mathcal{CP} -even physical Higgs boson (h) in the SM. In the earlier work of Ref. [27], such an alignment limit has been imposed on the parameter space such that the lightest \mathcal{CP} -even state of the 3HDM had exactly the same couplings as those of the SM Higgs bosons.

In this project, we go beyond the Higgs alignment limit and enable the H_0 state to become somewhat misaligned with respect to the SM Higgs direction in a controllable way. Indeed, due to the current stringent experimental constraints on the Higgs boson couplings to the SM fermions and gauge bosons, such a misalignment cannot be too large and constrains these Higgs couplings to be within 10% or so of those predicted by the SM [5].

Instead of the direct diagonalisation of the gauge-basis mass forms extracted from the bilinear terms of the potential V expanded around the vacuum state by using Eq. (3.7), the use of the intermediate Higgs basis significantly simplifies the analysis of the scalar, pseudoscalar and the charged scalar sectors. Indeed, in the Higgs basis the mass forms acquire a block-diagonal form where the Goldstone modes manifestly decouple from the physical spectrum. With the choice of the first Higgs doublet getting the VEV v only, the $(1, 1)$ -components of the Higgs-basis pseudoscalar and charged scalar mass matrices correspond to neutral and charged Goldstone modes, respectively. Then, the pseudoscalar and charged scalar mass forms in the Higgs basis reduce to 2×2 matrices whose diagonalisation becomes a trivial step.

More explicitly, the gauge-basis mass matrix M^G gets rotated to the Higgs-basis mass matrix M^H by means of the orthogonal transformation in Eq. (3.15) as

$$M^H = \mathcal{O}_\beta M^G \mathcal{O}_\beta^T, \quad (3.18)$$

where M^G and M^H have the forms

$$M^G = \begin{pmatrix} M_{11}^G & M_{12}^G & M_{13}^G \\ M_{21}^G & M_{22}^G & M_{23}^G \\ M_{31}^G & M_{32}^G & M_{33}^G \end{pmatrix} \quad \text{and} \quad M^H = \begin{pmatrix} 0 & 0 & 0 \\ 0 & M_{22}^H & M_{23}^H \\ 0 & M_{32}^H & M_{33}^H \end{pmatrix}, \quad (3.19)$$

for both the \mathcal{CP} -odd η and the charged H^\pm Higgs fields. The residual diagonalisation can be performed in a trivial way by means of the block-diagonal rotation matrices

$$\mathcal{O}_{\gamma_i} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\gamma_i) & -\sin(\gamma_i) \\ 0 & \sin(\gamma_i) & \cos(\gamma_i) \end{pmatrix}, \quad i = 1, 2, \quad (3.20)$$

given in terms of two angles γ_1 and γ_2 for the charged and \mathcal{CP} -odd scalar sectors, respectively. Then, the transformations between the gauge and mass bases (through the intermediate Higgs basis) can be represented as

$$M_{H^\pm}^{\text{diag}} \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & m_{H_1^\pm}^2 & 0 \\ 0 & 0 & m_{H_2^\pm}^2 \end{pmatrix} = \mathcal{O}_{\gamma_1} \mathcal{O}_\beta M_{H^\pm}^G \mathcal{O}_\beta^T \mathcal{O}_{\gamma_1}^T, \quad (3.21)$$

$$M_\eta^{\text{diag}} \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & m_{A_1}^2 & 0 \\ 0 & 0 & m_{A_2}^2 \end{pmatrix} = \mathcal{O}_{\gamma_2} \mathcal{O}_\beta M_\eta^G \mathcal{O}_\beta^T \mathcal{O}_{\gamma_2}^T. \quad (3.22)$$

At the same time, the \mathcal{CP} -even Higgs mass matrix in the Higgs basis reveals a specific structure when the $(1, 1)$ -component corresponds to the H_0 state defined in Eq. (3.17). Thus, if the Higgs alignment limit is concerned, the $(1, 1)$ element for the mass matrix becomes equal to the Higgs mass squared m_h^2 in the SM and decouples from other \mathcal{CP} -even scalars. In this case, the matrix reduces to a block-diagonal form, with a 2×2 residual mixing between the heavier scalar states only. We, however, do not impose this limit considering a more generic 3×3 \mathcal{CP} -even mass form in the Higgs basis whose diagonalisation is performed by three subsequent 2×2 rotations

$$\mathcal{O}_\alpha = \begin{pmatrix} \cos(\alpha_1) & -\sin(\alpha_1) & 0 \\ \sin(\alpha_1) & \cos(\alpha_1) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\alpha_2) & 0 & \sin(\alpha_2) \\ 0 & 1 & 0 \\ -\sin(\alpha_2) & 0 & \cos(\alpha_2) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_3) & -\sin(\alpha_3) \\ 0 & \sin(\alpha_3) & \cos(\alpha_3) \end{pmatrix}, \quad (3.23)$$

such that the diagonalisation procedure reads

$$M_h^{\text{diag}} \equiv \begin{pmatrix} m_h^2 & 0 & 0 \\ 0 & m_{H_1}^2 & 0 \\ 0 & 0 & m_{H_2}^2 \end{pmatrix} = \mathcal{O}_\alpha \mathcal{O}_\beta M_h^G \mathcal{O}_\beta^T \mathcal{O}_\alpha^T. \quad (3.24)$$

Here, m_h is the measured value of the SM Higgs boson given in Eq. (2.9). As was mentioned above, in the Higgs alignment limit one enforces the $(1, 1)$ elements of M_h^{diag} and M_h^H to be the same, while enforcing the absence of any mixing between this $H_0 \equiv h$ state with the heavier non-SM scalars. One way such an alignment can be imposed is by setting $\alpha_1 = 0$ and $\alpha_2 = 0$ (c.f. Ref. [27]). In the numerical analysis of the 3HDM parameter space, this limit can be easily relaxed to an approximate Higgs alignment $H_0 \approx h$ enabling a small mixing between h and $H_{1,2}$ by setting α_1 and α_2 to small values varying within certain ranges allowed by the Higgs couplings measurements.

3.2.3 Inversion

Next, we would like to define the physical input parameters in order to ensure that the scalar mass spectrum is positively definite, i.e. no tachyonic solutions are considered from the

beginning and the vacuum is manifestly stable, at least, at tree level. For this purpose, we adopt an inversion procedure in the scalar sector enabling us to express many of the Higgs potential parameters in terms of the physical masses, Higgs VEVs and mixing angles. This is a particularly convenient representation of the parameter space as it enables to account for possible experimental bounds that are often imposed on the mass values of additional (typically, heavy) scalar boson states, as well as on their interactions with SM gauge bosons and fermions that can be readily represented in terms of the scalar mixing angles. If for example heavier Higgs masses need to be above a certain value in the search, this constraint would just mean adjusting the lower bound for the corresponding input parameter.

Of course, such a connection can be established analytically at tree level only (see below), while the measured bounds on the physical parameters effectively account for higher order radiative corrections. Nevertheless, such an inversion procedure provides an important guidance assuming that radiative corrections are not dominant but rather subleading in the masses and physical couplings for relevant parts of the model parameter space. This way, we assume that only a minor subset of such parameter points would be discarded as soon as one-loop corrections are incorporated to the scalar boson masses and couplings.

Likewise, in the quark sector discussed below, it is also instructive to determine the physical input parameter space in terms of measured physical masses of quarks and their mixing parameters entering the procedure via measured values of the CKM matrix. Such an inversion would then enable computation of the relevant entries in the Yukawa matrices, while a spread in input values of the quark masses and mixing angles would roughly correspond to their experimental error bars. This means that constraints on the physical parameters would be easy to fulfill. Thus, the numerical scan would be performed within the physical ranges of Yukawa couplings and Higgs VEVs that provide an approximately correct SM quark spectrum.

A more standard procedure of fitting the scalar self-interaction and Yukawa couplings in the Lagrangian to the spectra and exclusion bounds quickly deteriorates with the growth of dimensionality of the parameter space, and could exhibit large problems when one accounts for very small experimental error bars on (e.g. quark) masses and mixing angles. The inverted procedure implemented in this project does not reveal the same problem and can be applied also for a generic scalar potential and Yukawa textures.

On the other hand, the inversion described above may represent an apparent difficulty with theoretical constraints such as perturbativity or unitarity bounds and vacuum stability formulated at the level of Higgs couplings. These, however, are less problematic than fitting the physical masses in very narrow allowed ranges or error bars, when precision observables (like quark masses or oblique corrections) are concerned. In order to also satisfy those constraints that remain to be satisfied by the parameter space after the inversion, we further develop an ML algorithm that trains a neural network on such constraints, further steering towards the physically interesting regions.

There is also a third possibility of finding a so-called hybrid basis, where one uses a suitable mix of Lagrangian and physical parameters to calculate the remaining ones, to hopefully be able to use the parameters with strong constraints as inputs, thereby simplifying the constraints overall. One example of this has been done for a softly-broken \mathcal{CP} conserving 2HDM in Ref. [48].

In order to relate the Lagrangian parameters with physical masses and rotation angles, it is instructive to rewrite Eqs. (3.22) and (3.24) such that one first rotates the gauge basis into the Higgs basis, and separately rotates the physical basis to the Higgs basis, and then

equates the resulting matrices as

$$\mathcal{O}_{\gamma_1}^T M_{H^\pm}^{\text{diag}} \mathcal{O}_{\gamma_1} = \mathcal{O}_\beta M_{H^\pm}^G \mathcal{O}_\beta^T, \quad (3.25)$$

$$\mathcal{O}_{\gamma_2}^T M_\eta^{\text{diag}} \mathcal{O}_{\gamma_2} = \mathcal{O}_\beta M_\eta^G \mathcal{O}_\beta^T, \quad (3.26)$$

$$\mathcal{O}_\alpha^T M_h^{\text{diag}} \mathcal{O}_\alpha = \mathcal{O}_\beta M_h^G \mathcal{O}_\beta^T. \quad (3.27)$$

This provides a linear system of coupled algebraic equations that has been analytically solved for λ_{1-10} and μ_{ij}^2 in terms of the scalar masses, $v_{1,2,3}$ and the angles $\alpha_{1,2,3}$ and $\gamma_{1,2}$. This is the essence of the inversion procedure implemented in our analysis of the Higgs sector. The full set of Lagrangian parameters resolved in terms of the input parameters is given in Appendix B. The number of equations in (3.27) is one less than the number of quartic couplings and soft-breaking mass terms, so we have chosen λ_{10} to be an input parameter for the inversion procedure as well. Since only the lightest SM-like state h has been discovered so far, with no evidence of additional Higgs “partners” yet, we can set the numerical values of the masses of other scalar states to be above certain lower bounds that can be potentially suggested by the experimental measurements at the LHC.

3.2.4 Further constraints on the Higgs sector

While the tree-level scalar mass and mixing constraints can be directly implemented by choosing appropriate ranges of physical input parameters, there are several additional sources of theoretical and phenomenological constraints on the scalar sector that should be satisfied. These concern, for example, the unitarity bounds and boundedness from below conditions on the parameters of the scalar potential, as well as the EW precision constraints encoded in the form of bounds on the Peskin-Takeuchi parameters, also known as the oblique corrections, S , T and U [31].

Other constraints, such as vacuum stability, particle masses at one loop and experimental bounds on SM Higgs couplings and non-SM scalar masses and decays, should be implemented on top of these in further more accurate phenomenological studies of the model.

Boundedness from below For the Higgs potential V introduced above to be stable, it should be bounded from below (BFB) in all directions. In our numerical analysis, we have used the following BFB conditions sufficient to ensure such a stability [49]

$$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0, \quad (3.28)$$

$$\sqrt{4\lambda_2\lambda_3} + \lambda_6 + \text{Min}[0, \lambda_9] > 0, \quad (3.29)$$

$$\sqrt{4\lambda_1\lambda_2} + \lambda_4 + \text{Min}[0, \lambda_7] > 0, \quad (3.30)$$

$$\sqrt{4\lambda_1\lambda_3} + \lambda_5 + \text{Min}[0, \lambda_8 + 2\lambda_{10}, \lambda_8 - 2\lambda_{10}] > 0. \quad (3.31)$$

Tree-level unitarity Description following Refs. [49, 50]. Under the basic assumption that all Higgs couplings are independent, the scattering amplitudes for $2 \rightarrow 2$ scattering of gauge and Higgs bosons can indefinitely grow with the center-of-mass energy. This is inconsistent with unitarity. This violation of unitarity shows that the assumption that the Higgs couplings are independent is incorrect. One can derive the key relations among the Higgs couplings such that the strong high-energy behavior of the scattering cancels out for all scattering amplitudes.

With these relations, the scattering amplitude will either approach a constant value or zero in the high-energy limit. Since the interest lies in the high-energy limit, one only needs to evaluate the scattering S-matrix for the two body scalar bosons, which only consists of the quartic part of the scalar potential V . The requirement of unitarity of the S-matrix can be translated into relations for the partial wave amplitude a_J , which in the high-energy limit becomes

$$\Im(a_J) = |a_J|^2. \quad (3.32)$$

This requirement can be reformulated to a_J lying on a circle centered on $(0, 1/2)$ with radius $1/2$ in the complex plane. From this, one can put a requirement on the tree-level amplitude of a_J as

$$|\Re(a_J)| < 1/2. \quad (3.33)$$

This can be translated to

$$|\Re(\mathcal{M})| < 8\pi, \quad (3.34)$$

where \mathcal{M} is the amplitude for the process. Now, conditions for unitarity can be found with these conditions by considering all possible $S_1 S_2 \rightarrow S_3 S_4$ processes, where S_i 's represent the different (pseudo)scalar bosons. In the high-energy limit, each element a_J is given by the (pseudo)scalar four point interactions, which means that only the s-wave amplitude contributes to the scattering process.

The unitarity conditions on the quartic couplings in the considered 3HDM can be adopted from a more generic 3HDM analysis in Ref. [49]. They read

$$|y_j| < 8\pi, \quad (j = 1, \dots, 4), \quad (3.35)$$

$$|y_j^\pm| < 8\pi, \quad (j = 5, \dots, 9), \quad (3.36)$$

$$|x_i| < 8\pi, \quad (i = 1, \dots, 9), \quad (3.37)$$

where y_j are defined as

$$y_1 = \lambda_4, \quad y_2 = \lambda_6, \quad (3.38)$$

$$y_3 = \lambda_4 + 2\lambda_7, \quad y_4 = \lambda_6 + 2\lambda_9, \quad (3.39)$$

while y_j^\pm are given by

$$y_5^\pm = \lambda_5 + 2\lambda_8 \pm 6\lambda_{10}, \quad y_6^\pm = \lambda_5 \pm \lambda_{10}, \quad (3.40)$$

$$y_7^\pm = \lambda_4 \pm \lambda_7, \quad y_8^\pm = \lambda_5 \pm \lambda_8, \quad y_9^\pm = \lambda_6 \pm \lambda_9, \quad (3.41)$$

and x_i are the eigenvalues of X_1 , X_2 and X_3 matrices built as follows

$$X_1 = \begin{pmatrix} 6\lambda_1 & 2\lambda_4 + \lambda_7 & 2\lambda_5 + \lambda_8 \\ 2\lambda_4 + \lambda_7 & 6\lambda_2 & 2\lambda_6 + \lambda_9 \\ 2\lambda_5 + \lambda_8 & \lambda_6 + \lambda_9 & 6\lambda_3 \end{pmatrix}, \quad (3.42)$$

$$X_2 = \begin{pmatrix} 2\lambda_1 & \lambda_7 & \lambda_8 \\ \lambda_7 & 2\lambda_2 & \lambda_9 \\ \lambda_8 & \lambda_9 & 2\lambda_3 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 2\lambda_1 & 0 & 2\lambda_{10} \\ 0 & 2\lambda_2 & 0 \\ 2\lambda_{10} & 0 & 2\lambda_3 \end{pmatrix}. \quad (3.43)$$

Electroweak precision tests The oblique corrections effectively parametrize new physics contributions to the EW observables such as vector boson masses and decay rates, etc. Given a number of high precision constraints on the EW observables from LEP collider, and more recently on the Higgs boson masses and couplings from the LHC, the room for new physics particles interacting with the SM vector bosons has become very constrained.

If the relevant measurements are performed at various momentum scales, from small $Q^2 \approx 0$ up to a vicinity of Z^0 - and W^\pm -mass scales, $Q^2 = m_Z^2$ and $Q^2 = m_W^2$, the new physics contributions can be generically parametrized by only the following six quantities [32],

$$S = \frac{4s_W^2 c_W^2}{\alpha} \bar{S}, \quad T = \frac{1}{\alpha} \bar{T}, \quad U = \frac{4s_W^2}{\alpha} \bar{U}, \quad (3.44)$$

$$X = \frac{s_W c_W}{\alpha} \bar{X}, \quad V = \frac{1}{\alpha} \bar{V}, \quad W = \frac{1}{\alpha} \bar{W}, \quad (3.45)$$

where α is the fine structure constant found in Eq. (2.11), s_W and c_W are the sine and cosine of the weak mixing angle θ_W whose experimental value is given in Eq. (2.10). In these definitions, a subtraction of the SM contribution is always present. For example, \bar{T} can be calculated as [32]

$$\begin{aligned} \bar{T} = & \frac{g^2}{64\pi^2 m_W^2} \left[\sum_{a=2}^n \sum_{b=2}^m |(\mathcal{U}^\dagger \mathcal{V})_{ab}|^2 F(m_a^2, M_b^2) \right. \\ & - \sum_{b=2}^{m-1} \sum_{b'=b+1}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{bb'}]^2 F(M_b^2, M_{b'}^2) \\ & - 2 \sum_{a=2}^{n-1} \sum_{a'=a+1}^n |(\mathcal{U}^\dagger \mathcal{U})_{aa'}|^2 F(m_a^2, m_{a'}^2) \\ & + 3 \sum_{b=2}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{1b}]^2 [F(m_Z^2, M_b^2) - F(m_W^2, M_b^2)] \\ & \left. - 3 [F(m_Z^2, m_h^2) - F(m_W^2, m_h^2)] \right], \end{aligned} \quad (3.46)$$

where m_a , $a = 1, 2, 3$, are the charged scalar masses, M_b , $b = 1, \dots, 6$, are the neutral scalar masses, with $b = 1, 2, 3$ entries corresponding to the \mathcal{CP} -odd scalar masses, and $b = 4, 5, 6$ are the \mathcal{CP} -even scalar masses, such that $M_4 = m_h$ is the SM-like Higgs boson mass, \mathcal{U} is a 3×3 matrix defined in our notation as

$$\mathcal{U} = \mathcal{O}_\beta^T \mathcal{O}_{\gamma_2}^T, \quad (3.47)$$

\mathcal{V} is a 3×6 matrix given by

$$\mathcal{V} = \begin{pmatrix} i & 0 & 0 & 1 & 0 & 0 \\ 0 & i & 0 & 0 & 1 & 0 \\ 0 & 0 & i & 0 & 0 & 1 \end{pmatrix} {}^6\mathcal{O}_{\beta\beta}^T {}^6\mathcal{O}_{\gamma_1\alpha}^T, \quad (3.48)$$

where ${}^6\mathcal{O}_\beta$ and ${}^6\mathcal{O}_\alpha$ are 6×6 matrices defined as

$${}^6\mathcal{O}_{\beta\beta} = \begin{pmatrix} \mathcal{O}_\beta & 0_{3 \times 3} \\ 0_{3 \times 3} & \mathcal{O}_\beta \end{pmatrix}, \quad {}^6\mathcal{O}_{\gamma_1\alpha} = \begin{pmatrix} \mathcal{O}_{\gamma_1} & 0_{3 \times 3} \\ 0_{3 \times 3} & \mathcal{O}_\alpha \end{pmatrix}, \quad (3.49)$$

and the function F is defined as

$$F(I, J) = \begin{cases} \frac{I+J}{2} - \frac{IJ}{I-J} \ln \frac{I}{J} & \text{if } I \neq J, \\ 0 & \text{if } I = J. \end{cases} \quad (3.51)$$

Definitions for the other oblique corrections and relevant functions can be found in Appendix C and in Ref. [32].

In what follows, only the set of first three oblique corrections S , T and U are employed. These are the most relevant for probing the effect of heavy new physics particles (heavy scalars, in our case) with mass scales above the EW scale. These are also known in the literature as the Peskin-Tacheuchi parameters [31]. In our numerical analysis, we constrain these parameters to be within the following experimentally allowed ranges [39]

$$-0.08 < S < 0.12, \quad (3.52)$$

$$-0.05 < T < 0.19, \quad (3.53)$$

$$-0.09 < U < 0.09. \quad (3.54)$$

The previously mentioned X , V and W parameters are relevant when new scalar masses are of similar size to m_W and m_Z [51], but are not used in this project.

Here, the S , T and U bounds are treated independently, but these parameters are strongly correlated, with a 92% correlation between S and T and a -80% (-93%) anti-correlated between U and $S(T)$ [39]. A more detailed analysis of the correlations between S , T and U can be found in Ref. [52].

3.3 Yukawa sector

3.3.1 Lagrangian and mass textures

The Lagrangian of Yukawa interactions \mathcal{L}_Y for the considered 3HDM can be written as follows

$$-\mathcal{L}_Y = \sum_{\rho=1}^3 \bar{q}_L^k \Gamma_{\rho}^{kj} \phi_{\rho} d_R^j + \sum_{\rho=1}^3 \bar{q}_L^k \Delta_{\rho}^{kj} \tilde{\phi}_{\rho} u_R^j + \bar{e}_L^k \Pi_1^{kj} \phi_1 e_R^j + \text{h.c.}, \quad (3.55)$$

where the sums run over the Higgs doublet generation index, ρ . Note that there is no such sum for the leptons since they are assumed to only couple to the first Higgs doublet ϕ_1 , for simplicity. This means that the lepton interactions in the Higgs alignment limit are the same as in the SM: $\Pi_1^{kj} \equiv \Pi^{kj}$. If there were any changes to the lepton part of the Yukawa sector, it could be explored in a similar way as the quark sector. A potentially relevant direction for future research is to introduce a possibility for a (very small) lepton flavor violation on the same footing as is done in the quark sector, having a potentially interesting prospects for new physics (see e.g. Ref. [53]). This can also be done in connection to the neutrino mass generation mechanisms, a topic for extensive research in the literature.

From the requirement that the Yukawa Lagrangian is invariant under the exact flavor $U(1) \times \mathbb{Z}_2$ symmetry, with transformations defined in Eq. (3.2), only very few elements of the down-quark and up-quark Yukawa matrices Γ_{ρ} and Δ_{ρ} are allowed to be non-zero. Indeed, with Γ_1 as an example, the products of phases and signs of the fields emerging due to a

combined set of transformations (3.2) in the corresponding term in Eq. (3.55) can be written as

$$\begin{pmatrix} -e^{i\alpha} & -e^{i\alpha} & e^{i\alpha} \\ -e^{i\alpha} & -e^{i\alpha} & e^{i\alpha} \\ 1 & 1 & -1 \end{pmatrix} \quad (3.56)$$

which then translates into the allowed structure of the Yukawa texture Γ_1

$$\Gamma_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \times & \times & 0 \end{pmatrix}, \quad (3.57)$$

where \times is an arbitrary complex number. For the other five matrices, the same procedure yields [27]

$$\Delta_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Gamma_2, \Delta_2 = \begin{pmatrix} \times & \times & 0 \\ \times & \times & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Gamma_3, \Delta_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \times \end{pmatrix}. \quad (3.58)$$

Expanding the quark Yukawa Lagrangian around the physical vacuum, one derives the explicit form of quark mass matrices and physical Yukawa interactions as follows

$$\begin{aligned} -\mathcal{L}_Y^Q &= \sum_{\rho=1}^3 \bar{q}_L^k \Gamma_\rho^{kj} \phi_\rho d_R^j + \sum_{\rho=1}^3 \bar{q}_L^k \Delta_\rho^{kj} \tilde{\phi}_\rho u_R^j + \text{h.c.} \\ &\stackrel{\text{SSB}}{=} \sum_{\rho=1}^3 (\bar{u}_L \quad \bar{d}_L)^k \Gamma_\rho^{kj} \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} H_\rho^+ \\ v_\rho + h_\rho + i\eta_\rho \end{pmatrix} d_R^j \\ &\quad + \sum_{\rho=1}^3 (\bar{u}_L \quad \bar{d}_L)^k \Delta_\rho^{kj} \frac{1}{\sqrt{2}} \begin{pmatrix} v_\rho + h_\rho - i\eta_\rho \\ -\sqrt{2} H_\rho^- \end{pmatrix} u_R^j + \text{h.c.} \\ &= \sum_{\rho=1}^3 \left[\frac{1}{\sqrt{2}} \bar{d}_L^k \Gamma_\rho^{kj} (v_\rho + h_\rho + i\eta_\rho) d_R^j + \bar{u}_L^k \Gamma_\rho^{kj} H_\rho^+ d_R^j \right] \\ &\quad + \sum_{\rho=1}^3 \left[\frac{1}{\sqrt{2}} \bar{u}_L^k \Delta_\rho^{kj} (v_\rho + h_\rho - i\eta_\rho) u_R^j - \bar{d}_L^k \Delta_\rho^{kj} H_\rho^- u_R^j \right] + \text{h.c.} \\ &= \bar{d}_L M_d d_R + \bar{u}_L M_u u_R \end{aligned} \quad (3.59a)$$

$$+ \frac{1}{\sqrt{2}} \bar{d}_L \left[\sum_{\rho=1}^3 \Gamma_\rho (h_\rho + i\eta_\rho) \right] d_R + \frac{1}{\sqrt{2}} \bar{u}_L \left[\sum_{\rho=1}^3 \Delta_\rho (h_\rho - i\eta_\rho) \right] u_R \quad (3.59b)$$

$$+ \bar{u}_L \left[\sum_{\rho=1}^3 \Gamma_\rho H_\rho^+ \right] d_R - \bar{d}_L \left[\sum_{\rho=1}^3 \Delta_\rho H_\rho^- \right] u_R + \text{h.c.}, \quad (3.59c)$$

where in the last three lines the quark generation indices are implicit. Here, Eq. (3.59a) represents the down- and up-mass terms, with the following mass textures

$$M_d = \frac{1}{\sqrt{2}} \left[\sum_{\rho=1}^3 v_\rho \Gamma_\rho \right] = \begin{pmatrix} \checkmark & \checkmark & 0 \\ \checkmark & \checkmark & 0 \\ \checkmark & \checkmark & \checkmark \end{pmatrix}, \quad M_u = \frac{1}{\sqrt{2}} \left[\sum_{\rho=1}^3 v_\rho \Delta_\rho \right] = \begin{pmatrix} \checkmark & \checkmark & 0 \\ \checkmark & \checkmark & 0 \\ 0 & 0 & \checkmark \end{pmatrix}, \quad (3.60)$$

whereas Eq. (3.59b) shows the neutral scalar Yukawa interactions and Eq. (3.59c) provides the charged scalar Yukawa interactions.

Because of the Yukawa textures in this model, Eqs. (3.57) and (3.58), there is no overlap between matrix elements inside the up- or down sectors, which makes it easy to relate elements in the mass matrices to elements in the Yukawa matrices. As an example, Γ_1 from Eq. (3.57) is the only Yukawa matrix that contributes to the $(3, 1)$ and $(3, 2)$ elements of M_d , giving the simple relations:

$$(M_d)_{31} = \frac{1}{\sqrt{2}}v_1(\Gamma_1)_{31}, \quad (M_d)_{32} = \frac{1}{\sqrt{2}}v_1(\Gamma_1)_{32}. \quad (3.61)$$

Similar relations hold true for all other elements of M_d and M_u .

3.3.2 Inversion

As usual, the unitary transformation between the gauge-basis states u, d and the mass-basis states \hat{u}, \hat{d} is different for left- and right-handed quark fields, namely,

$$\begin{aligned} u_L^i &= U_{L,u}^{ij} \hat{u}_L^j, & u_R^i &= U_{R,u}^{ij} \hat{u}_R^j, \\ d_L^i &= U_{L,d}^{ij} \hat{d}_L^j, & d_R^i &= U_{R,d}^{ij} \hat{d}_R^j, \end{aligned}$$

Each of these four generic unitary transformation matrices can be parametrized in the following way [54]

$$U = \Theta Z_{xy} Z_{xz} Z_{yz}, \quad (3.62)$$

where

$$\Theta = \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix}, \quad (3.63)$$

$$Z_{xy} = \begin{pmatrix} \cos(\zeta_1) & \sin(\zeta_1)e^{i\varphi_1} & 0 \\ -\sin(\zeta_1)e^{-i\varphi_1} & \cos(\zeta_1) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (3.64)$$

$$Z_{xz} = \begin{pmatrix} \cos(\zeta_2) & 0 & -\sin(\zeta_2)e^{-i\varphi_2} \\ 0 & 1 & 0 \\ \sin(\zeta_2)e^{i\varphi_2} & 0 & \cos(\zeta_2) \end{pmatrix}, \quad (3.65)$$

$$Z_{yz} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\zeta_3) & \sin(\zeta_3)e^{i\varphi_3} \\ 0 & -\sin(\zeta_3)e^{-i\varphi_3} & \cos(\zeta_3) \end{pmatrix}. \quad (3.66)$$

Through such parametrisations, the measured elements of the quark CKM mixing matrix $V_{\text{CKM}} = U_{L,u}^\dagger U_{L,d}$ are directly related to the quark mixing angles and phases in the left-handed sector of the theory. On the other hand, the unitary transformation matrices are used to diagonalise the mass forms in Eq. (3.59a)

$$\begin{aligned} \bar{d}_L M_d d_R + \bar{u}_L M_u u_R &= \bar{\hat{d}}_L (U_{L,d}^\dagger M_d U_{R,d}) \hat{d}_R + \bar{\hat{u}}_L (U_{L,u}^\dagger M_u U_{R,u}) \hat{u}_R \\ &= \bar{\hat{d}}_L M_d^{\text{diag}} \hat{d}_R + \bar{\hat{u}}_L M_u^{\text{diag}} \hat{u}_R, \end{aligned} \quad (3.67)$$

where the physical quark mass forms $M_{u,d}^{\text{diag}}$ are given in Eq. (2.16). Using $U_{L,d}^\dagger = V_{\text{CKM}}^\dagger U_{L,u}^\dagger$, we obtain

$$M_d^{\text{diag}} = V_{\text{CKM}}^\dagger U_{L,u}^\dagger M_d U_{R,d}, \quad (3.68)$$

which together with $M_u^{\text{diag}} = U_{L,u}^\dagger M_u U_{R,u}$ provide us with a set of coupled linear equations for the inversion procedure in the Yukawa sector. This set of equations can be resolved with respect to non-zero elements of the Yukawa textures in Eqs. (3.57) and (3.58) in terms of the physical quark masses, CKM elements and the parameters of the unitary “rotation” matrices $U_{L,u}$, $U_{R,u}$ and $U_{R,d}$, by relating $M_{u,d}$ to the Yukawa matrices as in the example given in Eq. (3.61). Such an inversion procedure will be employed in our numerical analysis in what follows allowing the measured quark masses and mixing parameters to vary within narrow intervals suggested by their experimental error bars. Note, the parameters of the right-handed unitary rotations, $U_{R,u}$ and $U_{R,d}$, are totally free since none of the measured observables are sensitive to those. At the same time, the left-handed transformation matrix in the up-sector $U_{L,u}$ is expected to be tightly constrained by the FCNC measurements.

3.3.3 FCNC couplings

To illustrate the BGL-like nature of the FCNC couplings, let us define the Higgs basis for the \mathcal{CP} -even scalars in terms of the gauge-basis eigenstates

$$\begin{pmatrix} H_0 \\ H'_1 \\ H'_2 \end{pmatrix} = \mathcal{O}_\beta \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}, \quad (3.69)$$

where the basis rotation matrix is given in Eq. (3.15). As was discussed earlier, in the Higgs alignment limit the state H_0 completely overlaps with SM Higgs boson h and thus has SM-like Yukawa couplings at the tree level. Provided that h generates no tree-level FCNCs in the SM, so does the H_0 state in the case of alignment, while the heavier non-SM states will, in general, feature tree-level FCNCs. This would necessarily exclude wide domains of the parameter space unless a strong decoupling of $H_{1,2}$ mass states occurs in the spectrum and/or there is an additional mechanism to suppress their tree-level FCNC contributions due to a specific flavor symmetry like in the considered model.

For the purpose of estimating the size of those FCNCs, let us write down the Yukawa couplings of H'_1 and H'_2 states with the physical down-type quarks as follows:

$$-\mathcal{L}_Q^{H'_1, H'_2} = \frac{H'_1}{v} \bar{d}_L N_{d1} \hat{d}_R + \frac{H'_2}{v} \bar{d}_L N_{d2} \hat{d}_R + \text{h.c.}, \quad (3.70)$$

where the matrices N_{d1} and N_{d2} are given by

$$N_{d1} = \frac{v}{\sqrt{2}v_{13}} U_{L,d}^\dagger (\Gamma_1 v_3 - \Gamma_3 v_1) U_{R,d}, \quad (3.71a)$$

$$N_{d2} = U_{L,d}^\dagger \left[\frac{v_2}{v_{13}} \frac{1}{\sqrt{2}} (\Gamma_1 v_1 + \Gamma_3 v_3) - \frac{v_{13}}{v_2} \frac{1}{\sqrt{2}} \Gamma_2 v_2 \right] U_{R,d}, \quad (3.71b)$$

in terms of the down-type unitary transformation matrices $U_{L,d}$ and $U_{R,d}$.

In order to simplify the expressions for the N_{d1} and N_{d2} matrices, let us get back to the mass textures in Eq. (3.60). Due to the block-diagonal structure of M_u , the corresponding

bi-diagonalizing matrices $U_{L,u}$ and $U_{R,u}$ are expected to have block-diagonal structures as well. Actually, it is possible to choose

$$U_{L,u} = \begin{pmatrix} \checkmark & \checkmark & 0 \\ \checkmark & \checkmark & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.72)$$

One could add a phase to the $(3,3)$ element of $U_{L,u}$ that would contribute to the total phase of $(M_u)_{33}$, but such a phase could be absorbed into $(U_{R,u})_{33}$, and is therefore not necessary in this choice of $U_{L,u}$. By definition of the CKM matrix, we find

$$(U_{L,d})^{3j} = V_{\text{CKM}}^{3j}, \quad (3.73)$$

meaning that the third row of $U_{L,d}$ is identical to that of the CKM matrix.

Now, it is instructive to define the following projection matrix

$$P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.74)$$

Then, using the specific structure of the Yukawa textures, Eqs. (3.57) and (3.58), we arrive at the following relations in the down-quark sector:

$$\Gamma_3 = (\Gamma_3)_{33}P, \quad \frac{1}{\sqrt{2}}(\Gamma_1 v_1 + \Gamma_3 v_3) = P.M_d. \quad (3.75)$$

With Eqs. (3.73) and (3.75), we can straightforwardly rewrite N_{d1} and N_{d2} given in Eq. (3.71) as follows

$$(N_{d1})_{ij} = \frac{v v_3}{v_1 v_{13}} (V_{\text{CKM}}^*)_{3i} V_{\text{CKM}}^{3j} (M_d^{\text{diag}})_{jj} - \frac{1}{\sqrt{2}} \frac{v v_{13}}{v_1} (\Gamma_3)_{33} (V_{\text{CKM}}^*)_{3i} (U_{R,d})_{3j}, \quad (3.76a)$$

$$(N_{d2})_{ij} = \frac{v_{13}}{v_2} (M_d^{\text{diag}})_{jj} \delta_{ij} + \left(\frac{v_{13}}{v_2} + \frac{v_2}{v_{13}} \right) (V_{\text{CKM}}^*)_{3i} V_{\text{CKM}}^{3j} (M_d^{\text{diag}})_{jj}. \quad (3.76b)$$

in terms of the diagonal down-type mass matrix M_d^{diag} found in Eq. (2.16).

Note that if $(\Gamma_1)_{31}$ and $(\Gamma_1)_{32}$ were zero in Eq. (3.57), then we could choose $U_{R,d}$ to be block-diagonal as well. However, due to the smallness of the $(3,j)$ elements in the CKM matrix (for $j \neq 3$), the elements of Γ_1 are expected to be small. In this case, it is reasonable to assume that the $(U_{R,d})_{3j}$ elements ($j \neq 3$) are also small. Thus, the off-diagonal elements of N_{d1} and N_{d2} matrices indeed exhibit a BGL-like suppression due to being proportional, at least, to a single power of small V_{CKM}^{3j} ($j \neq 3$) elements. So, we expect that the FCNC couplings in the down-quark sector will be under control.

A similar procedure applied to the up-quark sector would show that, due to special structure of the up-type Yukawa textures dictated by the flavor symmetry, there are no scalar boson mediated FCNC couplings at the tree level there. As in the usual BGL formulation, it is possible, in fact, to redefine the flavor symmetry charges of quark fields in Eq. (3.2) such that the tree-level FCNC couplings would reside entirely in the up-quark sector instead of the down-quark one. However, we choose to work with the current version because, relatively light BSM scalars possessing tree-level FCNC couplings in the down-quark sector have a better chance to be discovered in the near future.

4 An overview of Artificial Neural Networks

Artificial Neural Networks (ANNs) are a popular method in ML that has its origin in creating a simple mathematical model of a brain. These represent a structure of artificial neurons connected into networks that can learn how to solve various problems, approximate functions or find certain patterns. The basis of the ANN is the artificial neuron, the perceptron [55], often called a node in a larger network. A perceptron consists of a number of weights w_i , often a bias weight b and an activation function ϕ , that are used with inputs x_i to calculate an output y as

$$y = \phi \left(b + \sum_i w_i x_i \right). \quad (4.1)$$

An visual example of a perceptron is presented in Fig. 3. The bias weight b is often visualized as an extra input node that always has the value one.

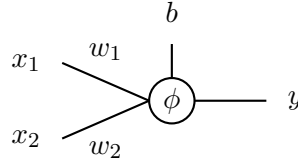


Figure 3: An example perceptron.

When several of these perceptrons are connected into a network, it becomes a Multi-Layer Perceptron (MLP), which is the simplest type of ANN. If all nodes are connected to every node in the previous layer, or to every input node, the network is said to be fully connected and fully connected MLPs are the simplest type of ANN. An example can be seen in Fig. 4. A node that is not in the input layer or output layer is called a hidden node.

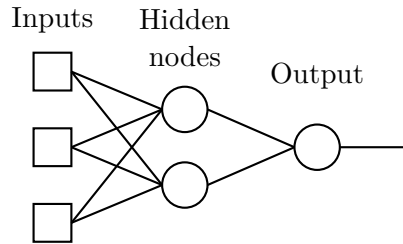


Figure 4: An example of a fully connected MLP.

There are several different ways to train ANNs, the most common being backpropagation, a method based on gradient descent where the weights of the network are changed so that the network follows the error function towards a minimum. With complicated outputs that switch depending on the set of targets, getting a hold of gradients for the outputs can be difficult. Another common group of methods for training neural networks, that need no knowledge of the output calculations, are Genetic Algorithms (GAs). As with the networks themselves, GAs are based in mimicking nature, specifically evolution, utilizing reproduction and survival of the fittest to find better networks.

To train with a GA, a population of networks is created and then the population reproduces, while bad networks are killed off, over generations until a stop criterion is met and the training is done. In the view of biological evolution, there are two common types of reproduction: crossover and mutation. These work on the chromosomes, or the list of weights, of the networks. The crossover creates children using two parent networks by mixing their chromosomes, whereas the mutation creates children by randomly changing the weights in the chromosome of one parent. The specifics of how both the crossover and mutation work depends on the GA used. The choice of what bad networks are killed off also depends on the method, but usually better children replace their parents in the population or the worst networks after a number of reproductions get killed. A visual example of the training process using a simple GA can be seen in Fig. 5.

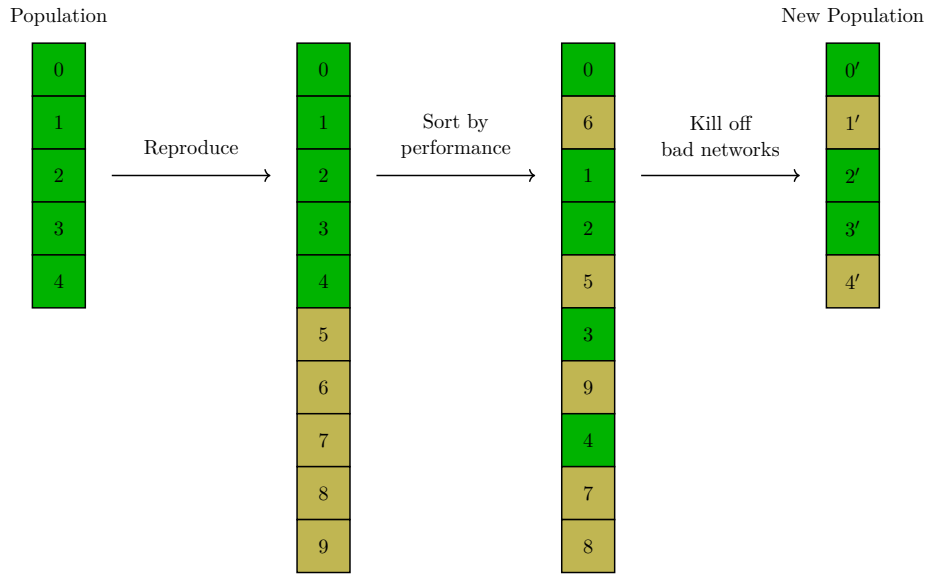


Figure 5: An example of the evolution of a small population of five networks over one generation with a genetic algorithm.

What determines if a network is good or bad is the error, or loss, function. The error function calculates how good or bad the prediction of an ANN is depending on the output and a target value. This calculation is also often done for a batch of inputs. Some common error functions are the Mean Square Error (MSE) and the Mean Absolute Error (MAE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (o_i - t)^2, \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n |o_i - t|, \quad (4.2)$$

where t is the target, o_i is the output for input i and the error is calculated with a batch of n inputs.

When training neural networks, the goal is to make a network with the smallest error, a network where the outputs are as close to the targets as possible, but this should hold for the general problem and not only for a specific set of input points. If too much training is done on a specific set of points, one might start to fit better to that specific set of points, but worse for the general, underlying function one tries to approximate. This is called overfitting. The opposite is underfitting, which happens when the problem is too complex for the specified

architecture of the networks. Examples of over- and underfitting are shown in Fig 6, for regression of a $y = x^2$ with added noise. Here, one can see that the green line, that represents underfitting, is not complex enough to find the underlying x^2 distribution and the blue curve, that represents overfitting, has become a lot more complicated than an x^2 function. It bends back and forth as much as it can trying to hit this specific set of points and because of this, new predictions based off the blue function will probably be worse than the red fit, which is in general closer to the true function.

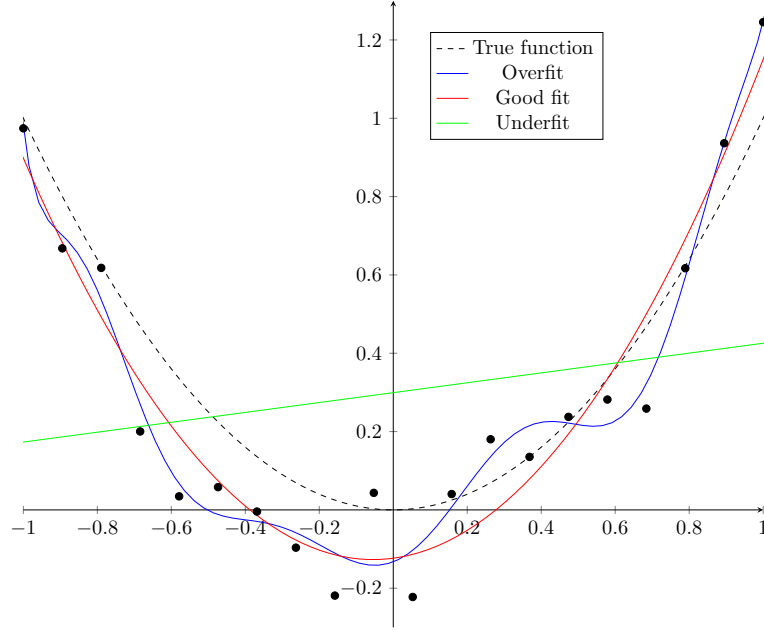


Figure 6: Showcase of underfitting and overfitting for x^2 with added noise. The green curve is a first order polynomial regression, the red is a second order and the blue is a 15th order.

Underfitting is not a problem if the architecture fits the problem and there is enough training. Overfitting, on the other hand, is trickier, but there are several different ways to deal with it. Some common methods are regularization and early stop.

Regularization is done by adding a regularization term $+\lambda R(w_i)$ to the error function, where R is a function of the weights w_i and λ is a parameter that determines the magnitude of the regularization. This R is often made to penalize large weights and therefore penalize complex answers. For the regression example, this would mean that the type of overfitting seen in the blue curve in Fig 6 would be forced towards the less complex red curve. Two common types of regularization are L1- and L2 regularization, also known as lasso- and ridge regression, where $R = \sum_i |w_i|$ and $R = \sum_i w_i^2$ respectively.

The early stop method requires another set of data, usually called validation data, on top of the training data. When training, the error calculated from training data will decrease over time, even when overfitting, but since overfitting by definition is training too much on a specific set, the error on another valid data set will instead increase. This means that looking at how the error of the validation data changes over time, one can notice when the method starts to overfit. The early stop method works by seeing if the error of a validation data set increases during training. If it does, stop training at that moment.

5 Methods and Algorithms

One of the main goals of this project was to create a computer program that can probe large and complicated parameter spaces, like in the new physics model discussed above, in a smarter and faster way than a random scan. This was achieved by creating a program that lets the ANNs transform the entire parameter input space by training them such that a significant portion of the transformed space passes a given set of theoretical and phenomenological constraints. Here, the ANNs have been trained by a GA made to converge in input spaces with possible epistasis among parameters [56], discussed more in Section 5.2. The training and validation sets used in this project are sampled according to another overarching algorithm inspired by the one made for large input and/or output dimensions [57] as discussed in Sections 5.3 and 5.4. The program has been written in Java 8.

5.1 Neural Network

The ANNs used in this project are fully connected MLPs that have the inputs connected to one layer of hidden nodes that connect to the output node, where the number of nodes in the hidden layer is the same as the number of input nodes, as seen in Fig. 7a. This architecture was chosen, since it is the smallest architecture possible that will work with the idea of transforming the entire input space. The smallest possible was chosen because bigger networks would make calculations in the training process slower, though this does not necessarily mean that larger architectures would converge slower. In further studies, it would be interesting to test different architectures, to see if larger architectures can possibly converge faster or find better, but more complicated, transformations.

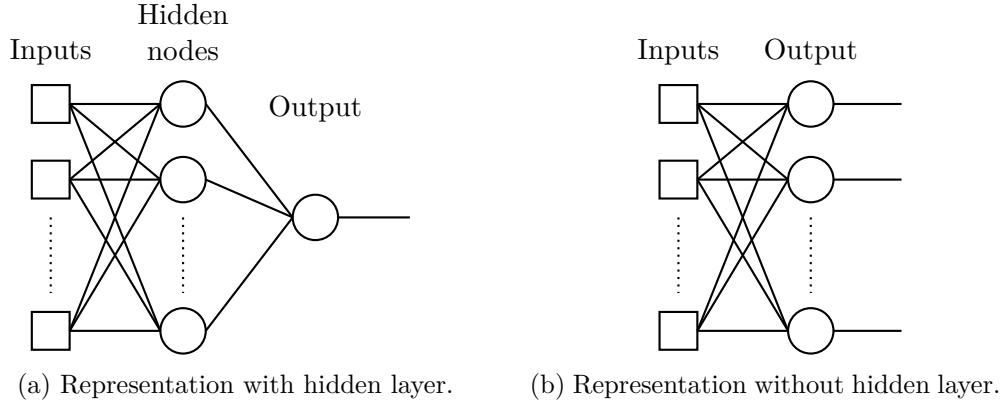


Figure 7: A fully connected MLP of the type of network used in this project.

For a network that trains on several different targets, one often has several output nodes, one for each target. This does not work here, since the networks are used to transform the input parameters. If there were different output nodes for different targets, and those output nodes had different sets of weights, they would represent a different set of input parameters for every node. This would mean the network would not find one transformation that creates good sets of input parameters that solve all the targets, but it would instead create different transformations for every target.

All input values used are in the interval of $[-1, 1]$, calculated from Halton sequences discussed in Sec. 5.3, and are scaled and translated to fit the defined range of the input when

necessary for calculations. The values need to be kept in this range to ensure that the different input values do not escape their defined boundaries. For this reason, the hyperbolic tangent has been used as the activation function for the hidden layer, as it has a range of $[-1, 1]$. If the output weights were trained on like the weights in the hidden layer, values could fluctuate out of this range, and therefore output weights are fixed to 1. With this in mind, the network can be thought of as an ANN with an input layer, no hidden layer, and an output layer, where the output layer calculates the transformation of the input space, which is then used to calculate errors outside of the ANN structure, as seen in Fig. 7b.

Network weights are initialized with normalized initialization [58], often called Xavier initialization, where weights are initialized uniformly within the interval $[-a, a]$ where

$$a = \sqrt{\frac{6}{k+1}}. \quad (5.1)$$

Here, k are the number of input parameters. This prevents the weighted sum to saturate the hyperbolic tangent function independently of the number of input parameters.

The error function used is a modified MAE, which calculates the error not depending on one target for every output, but an allowed interval. If the output is in the target interval, the error is zero. If the output is outside of the interval, the error is calculated as the absolute value between the output and the closest interval endpoint. The total error is

$$E_{\text{MA}} = \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{j=1}^m E(o_j^i, t_j^L, t_j^U) \quad (5.2)$$

where n is the number of points, m is the number of outputs and targets per point and $E(o, t^L, t^U)$ is calculated as

$$E(o, t^L, t^U) = \begin{cases} t_j^L - o_j^i & \text{if } o_j^i < t_j^L, \\ 0 & \text{if } t_j^L < o_j^i < t_j^U, \\ o_j^i - t_j^U & \text{if } t_j^U < o_j^i, \end{cases} \quad (5.3)$$

, where o are outputs, and t^L and t^U are the lower- and upper bounds on the targets, respectively.

When numerically calculating Eq. (5.2), one has to carefully make sure that the values of the sums do not become larger than the largest possible real value represented in the program and therefore turn into some representation of infinity. Doing the divisions inside the sums will require more operations for the program, and therefore will be slower, but it is also safer, since it protects from this type of overflowing. Here, this is the reason the factor $1/m$ is kept within the first sum.

If an output has a value of Not a Number (NaN), from an undefined mathematical operation e.g. a negative argument in a square root, the largest absolute error a_j for the specific target t_j is used. This modifies $E(o, t^L, t^U)$ with

$$|o_j^i - t_j^{U/L}| \rightarrow \begin{cases} a_j & \text{if } o_j^i = \text{NaN}, \\ |o_j^i - t_j^{U/L}| & \text{otherwise}, \end{cases} \quad (5.4)$$

where

$$a_j = \text{Max} [|o_j^1 - t_j^U|, \dots, |o_j^n - t_j^U|, |o_j^1 - t_j^L|, \dots, |o_j^n - t_j^L|]. \quad (5.5)$$

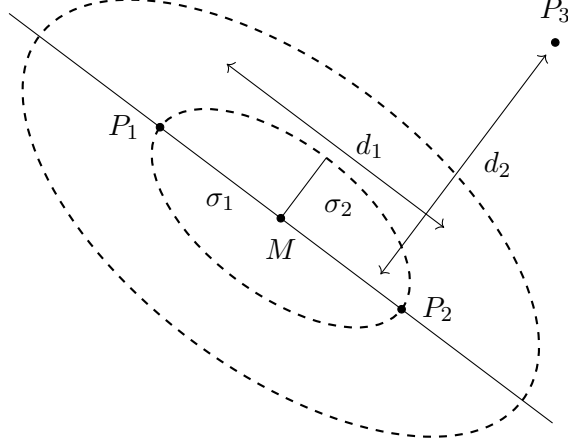


Figure 8: A visualisation of how UNDX creates children. The ellipses show the 1σ and 2σ cut-offs for the normal distributions, with $\alpha = 0.5$, $\beta = 0.35$ and $n = 2$.

5.2 Genetic Algorithm

In this project, a robust GA that utilizes both Unimodal Normal Distribution Crossover (UNDX) [59] and Uniform Crossover (UX) was implemented to train the ANNs. This method has been shown to converge on data sets with strong epistasis between parameters [56], meaning that it works on the data where the influence of some parameters might overpower others, which is important for models with unknown or complicated interactions between the inputs. Algorithm 1 describes the setup used in the current project. To speed up the process, the program uses multiple threads to run this algorithm, with slight modifications to make it work safely and properly with multiple threads. The number of threads used should be the same as the amount of cores in the computer [60].

Here, the UX reproduction randomly splits the weights from the two parents between the two children,

$$c_{1i} = p_{xi}, \quad c_{2i} = p_{yi} \text{ with} \quad (5.6)$$

$$\vec{P}_1 = (p_{11}, \dots, p_{1n}) \text{ and} \quad (5.7)$$

$$\vec{P}_2 = (p_{21}, \dots, p_{2n}), \quad (5.8)$$

where \vec{P}_1 and \vec{P}_2 are the chromosomes, or weight vectors, of the parents, n is the number of weights in a chromosome, and x is uniformly randomized to take values 1 or 2, while y takes the other value, for $i = 1, 2, \dots, n$.

The UNDX creates children using the normal distributions defined by three parents, see Fig. 8. The normal distributions are centered around the midpoint \vec{m} between the chromosomes of the two main parents, \vec{P}_1 and \vec{P}_2 . The standard deviation for the first distribution is $\sigma_1 = \alpha d_1$, where $\alpha = 0.5^3$ is a constant and d_1 is the distance between \vec{P}_1 and \vec{P}_2 . The second standard deviation is $\sigma_2 = \beta d_2 / \sqrt{n}$, where $\beta = 0.35^3$, and d_2 is the shortest distance between the chromosome of the third parent \vec{P}_3 and the line connecting \vec{P}_1 and \vec{P}_2 . The children \vec{C}_1

³ $\alpha = 0.5$ and $\beta = 0.35$ are the recommended values [56, 59]

Algorithm 1: Algorithm for UNDX with UX [56] reproduction. n_{kids} is a parameter (here: $n_{kids} = 1$), n_{pop} is the size of the population. P_{undx}^S is limited to $[0.05, 0.95]$. Every RANDOM is a new uniform random number in $[0, 1)$. The child closest to Parent 1(2) is Child 1(2).

Input: Population, Training set, Validation set.

Initialization

└ Initialize P_{undx} to chosen value (Here 0.1).

repeat

└ Set $N_{undx}, N_{ux}, N_{undx}^S, N_{ux}^S = 0$.

└ **while** $i < n_{kid}(n_{pop}/2)$ **do**

1 **Selection for reproduction**

└ Chose Parent 1 and Parent 2 by random sampling without replacement from the population.

2 **Reproduction**

└ **if** RANDOM $< P_{undx}$ **then**

└ Chose Parent 3 by random sampling. Create two children according to UNDX. If either child is better than both parents, increment N_{undx}^S . Increment N_{undx} .

└ **else**

└ Create two children according to UX. If either child is better than both parents, increment N_{ux}^S . Increment N_{ux} .

└ **Selection for survival**

└ **if** Child 1 $>$ Parent 1 and 2 **then**

└ Replace Parent 1 with Child 1.

└ **if** Child 2 $>$ Parent 2 **then**

└ └ Replace Parent 2 with Child 2.

└ **else if** Child 2 $>$ Parent 1 and 2 **then**

└ Replace Parent 2 with Child 2.

└ **if** Child 1 $>$ Parent 1 **then**

└ └ Replace Parent 1 with Child 1.

└ i++

if $n_{pop}/2 > N_{undx}^S + N_{ux}^S$ **then**

└ **while** $j < n_{pop}/2 - (N_{undx}^S + N_{ux}^S)$ **do**

└ **Selection for reproduction:** Same as block 1.

└ **Reproduction:** Same as block 2.

└ **Selection for survival**

└ The first individual that survives is the best individual from the family of parents and children. The second one is selected from the other three by rank-based selection [61]. These individuals replace the parents.

└ j++

└ **Update probabilities**

└ Calculate $P_{undx}^S = N_{undx}^S / N_{undx}$ and $P_{ux}^S = N_{ux}^S / N_{ux}$

└ Update $P_{undx} = P_{undx}^S / (P_{undx}^S + P_{ux}^S)$.

until Stop criterion satisfied, according to Section 5.2.1;

and \vec{C}_2 can now be defined as

$$\vec{C}_1 = \vec{m} + z_1 \vec{e}_1 + \sum_{k=2}^n z_k \vec{e}_k \text{ and} \quad (5.9)$$

$$\vec{C}_2 = \vec{m} - z_1 \vec{e}_1 - \sum_{k=2}^n z_k \vec{e}_k, \text{ where} \quad (5.10)$$

$$\vec{e}_1 = \frac{\vec{P}_2 - \vec{P}_1}{|\vec{P}_2 - \vec{P}_1|} \quad (5.11)$$

and \vec{e}_k are the orthonormal basis vectors spanning the weight-space. In this project, \vec{e}_k are generated through the modified Gram-Schmidt process with \vec{e}_1 as a starting point and z_1 is taken from the normal distribution $N(0, \sigma_1^2)$ and z_k – from $N(0, \sigma_2^2)$. The children are named such that the child closest to Parent 1 is Child 1.

GAs often use both crossover and mutation. However, since the UNDX is a type of extended-box crossover, where the children can be created in a volume of weight-space not strictly bounded by the weights of the parents, mutation is not necessary and can even worsen the performance [62], and is therefore not used in this project.

5.2.1 Training stop criteria

There are two stop criteria for the UNDX with UX training algorithm: early stop and maximum generations.

The early stop criterion is set so that the training stops if the validation error increases. This is to counteract overfitting to the training data. Overfitting means training so much on a specific data set that the networks perform great on that data set, but would perform worse on most of other valid data sets. The validation error used is E_{MA} calculated from the validation data and the four best networks in the generation.

The training will also stop if a maximum number of generations is reached. Here, this maximum is set to 20 in order to encourage training with different sets of data.

5.3 Sampling points

With large input spaces and complex error landscapes, how one chooses to randomize or to calculate the points to be used for validation or training on that input space becomes more important. A good set of points should be able to represent the underlying error landscape while also uniformly sampling the input space. The frequently used pseudo-random generation is a fast and simple method, but it does not generate a uniform spread of points if the input space is too large or the number of points is too small. Similarly, the generation according to a uniform grid does not work well for large input spaces or a low number of points. In this project, we employ the Halton sequences [57, 63, 64] to generate points.

Any integer n , can be written in base b as

$$n = n_0 + n_1 b + n_2 b^2 + \dots + n_M b^M, \quad (5.12)$$

where M is the integer part of $\ln n / \ln b$. From here a unique fraction $\phi_b(n)$ that lies between 0 and 1 can be constructed as

$$\phi_b(n) = \frac{n_0}{b} + \frac{n_1}{b^2} + \frac{n_2}{b^3} + \dots + \frac{n_M}{b^{M+1}}. \quad (5.13)$$

A k -dimensional Halton sequence can then be created as

$$(\phi_{b_1}(n), \phi_{b_2}(n), \phi_{b_3}(n), \dots, \phi_{b_k}(n)) , \quad (5.14)$$

where $n = 1, 2, \dots, N$ and b_1, b_2, \dots, b_k are only necessarily coprime, but are commonly the first k primes. Since N is not used in the sequence, it is simple to create more points in the sequence by gradually increasing N . The Halton sequences generate numbers in $[0, 1]$, which are here stretched to $[-1, 1]$. The points in the sequence can be calculated using the Algorithm 2.

Algorithm 2: Algorithm for generating the next point in a Halton sequence [63] of dimension k with bases b_j , modified for using a leap value L . The counter c is initialized outside the function, as to keep its value over several function calls. Brackets here denote the integer part of a number. The last line stretches f from $[0, 1]$ to $[-1, 1]$.

Initialization

```

  c++
  H = empty list of length k
  n = cL
  for j = 0; j < k; j++ do
    f = 0
    m = [ln n / ln b_j]
    for i = m; i >= 0; i-- do
      B = [(b_j)^i]
      d = n/B
      f = f + d/(B × b_j)
    n = n mod B
  H_j = 2f - 1

```

The Halton algorithm calculates a sequence of points, but those points are only uniformly distributed on unit square projections of the hypercube for small dimensions. For projections on larger dimensions, the sampling clusters into bands, nullifying the uniform distribution in those dimensions. There are different ways to mitigate this effect, for example, by permuting the values ϕ within the points or using a leap value L to skip values of n , such that $n = mL$, where $m = 1, 2, \dots, N/L$. $L = 409$ is considered to be a good leap value [64], and is used here. When implementing a leap value, it should not be used as one of the bases b .

5.4 Incremental learning

The algorithm that controls how the networks are trained is the so-called incremental learning algorithm, specified in Algorithm 3, which takes inspiration from Diego G. et al [57]. Here, S new Halton points are generated at every iteration, to be able to explore a larger number of points without an extreme increase in memory usage. The fact that Halton sequences are used to generate points means that the points will be uniformly spread out over the input space regardless of the number of points. Even with uniformly distributed points, if S is small enough, one set of points will be very different from another set of points and the algorithm will have a hard time training and will quickly hit the early stop criterion. Even in these

cases, after some slow training during the first iterations or after it finds a training set that is a good enough representation of most points, it might start to train and converge like normal, but a larger number of points is recommended.

Algorithm 3: Algorithm for sampling and incremental learning. Halton points are generated according to Algorithm 2 in Section 5.3.

Initialization

Initialize a population of n_{pop} networks.

Validation = empty set.

Train = input set with S Halton points

repeat

Validation = *Train*.

Train = input set with S Halton points.

Train the population according to Algorithm 1 with *Train* and *Validation* data.

until *Convergence according to Section 5.4.1;*

5.4.1 Convergence criterion

The convergence criterion used for the incremental learning algorithm is based upon the so-called pass rate. Since targets are defined in terms of intervals, it is simple to check if a point passes all the target constraints. Since finding input spaces where points pass all targets is the goal of the algorithm, this convergence criteria is used. The pass rate is calculated as the ratio of the number of points in a set that pass over the total number of points in a set. The set of points used is the set created using the validation data and the four best networks in the iteration. Here, convergence was defined if the pass rate is greater than 0.5, meaning at least 50% of points generated from the validation data and the four best networks pass all target constraints.

5.5 Input and target files

In order to generalize the program for it to be usable with different models and different inversions, an input file was created for the inversion procedure in which the input parameters, Lagrangian parameters, parameters that are both input and Lagrangian ones and extra necessary equations are defined in blocks. In these blocks, variables can be defined with a name and a type. For input parameters, a range of values to search over also has to be defined, while for the Lagrangian parameters and extra equations, the equations that compute them have to be specified. The variable names used in the equations should be defined prior in the input file. The equations are entered in Mathematica form. The program then translates the variable definitions, input ranges and Mathematica equations into the code that can be executed. The number of inputs specified in this file also defines the dimensionality of the network and the Halton sequences.

Some example rows of an input file are:

```
- InputParameters -
a1;      Angle;      0;      2*pi
b;       Real;       4;      15
```

```
- LagrangianParameters -
out;      Real;      b Sin[a1]
```

Here two inputs are defined: an angle called $a1 \in [0, 2\pi]$ and a real value called $b \in [4, 15]$. One output out is defined as $out = b \sin(a1)$. During the inversion procedure, a set of numbers in a specific range (here, $[-1, 1]$) would be assigned to $a1$ and b and the numbers would then be stretched and translated to conform the specified ranges of $a1$ and b . Then, out would be calculated with these values of $a1$ and b and returned.

Target files are also defined with different blocks, e.g. unitarity constraints and oblique parameters, which are all calculated in the program. The necessary variables for these target calculations have to be defined in the input file. There are also blocks for targets on **SPheno** [37, 38] and **flavio** [33] outputs. Some example rows from the target file with the boundedness from below conditions are:

```
11;      0;
12;      0;
13;      0;
Sqrt[4 12 13] + 16 + Min[0, 19];      0;
Sqrt[4 11 12] + 14 + Min[0, 17];      0;
Sqrt[4 11 13] + 15 + Min[0, 18 + 2 110, 18 - 2 110];      0;
```

5.6 Modular training

The training was made modular, where it is done on a sequence of modules, each containing a set of targets. This was done both to make it easier to train by reducing the size of the target space that the networks need to search through and to make it easy to add more constraints for a more precise analysis later. The program first trained the networks on unitarity constraints in the scalar sector of the considered 3HDM, then retrained with the BFB constraints added, then again retrained with bounds on oblique parameters added.

Since all of the targets need to overlap anyways, this would find the minimum MAE, as defined in Eq. (5.2), for one batch of targets, then, when a new set of targets are added, the algorithm would work around the minimum of the first batch to find an overlapping (or close-by) minimum for the second batch, and so on. If one wants to change the second batch of targets, retraining on the first batch would not be necessary; one would only need to train on the new second batch in the same minimum of the first batch.

If the algorithm cannot fully find a new global minimum after the addition of new targets, since it was already close to a previous minimum, it could instead fall into a local minimum. If this happens, what areas of the parameter space are found might depend on the order in which the different modules are trained. This has not been explored in this project, but might be interesting for further studies.

6 Numerical analysis

Here, we would like to present the results of a first numerical analysis of the parameter space of the BGL-like \mathcal{CP} -conserving 3HDM using the ML techniques and algorithms described above in detail.

Errors bars listed are 1σ deviations. In this work, we force all values into these error bars, but for future work one could consider sampling the values according to the normal

distributions defined by the means and standard deviations listed, for a more statistically accurate picture.

6.1 Parameter search specifications

In the scalar sector, the SM-like lightest \mathcal{CP} -even scalar mass took values in 125.10 ± 0.14 GeV suggested by the experimental uncertainties, Eq. (2.9), while the other heavier scalar masses have been allowed to take values in the 200 GeV to 2 TeV range. The angles $\gamma_1, \gamma_2, \alpha_3, \beta_1$ and β_2 have been distributed over $[0, 2\pi]$, while α_1 and α_2 took small values in $[0, 10^{-2}]$, in a close proximity to the Higgs alignment limit. Besides, λ_{10} has been taken from $[-5, 5]$ interval consistent with perturbativity. In the Yukawa sector, the mixing angles and phases in U_L and U_R took values in the interval $[0, 2\pi]$. The CKM elements were constrained to be within the experimental ranges given in Eq. (2.21). The SM fermion masses, that are used in our analysis as part of the input parameter space, were determined at a suitable top-quark mass scale $Q^2 = m_t^2$ in Ref. [65] and read as follows

$$\begin{aligned} m_u &= 1.22^{+0.48}_{-0.40} \text{ MeV}, & m_c &= 0.590 \pm 0.080 \text{ GeV}, & m_t &= 162.9 \pm 2.8 \text{ GeV}, \\ m_d &= 2.76^{+1.19}_{-1.14} \text{ MeV}, & m_s &= 52 \pm 15 \text{ MeV}, & m_b &= 2.75 \pm 0.09 \text{ GeV}, \\ m_e &= 0.485289396 \text{ MeV}, & m_\mu &= 102.4673155 \text{ MeV}, & m_\tau &= 1742.15 \pm 0.20 \text{ MeV}. \end{aligned}$$

While rather tight constraints on the input parameters have been imposed due to experimentally well measured values such as the SM-like Higgs boson mass, in principle, they can be relaxed somewhat due to potentially sizable effects from radiative corrections which are not accounted for in our analysis. We, however, were mostly concerned here by the proof-of-concept in this first use of the ML algorithm to explore the complex BGL-3HDM parameter space and a diverse set of constraints in the most efficient manner, even within the tightest possible bounds. Later on, provided that the code is well-tested and accounts for the most relevant constraints, some of these ranges can be relaxed in order to explore larger potentially viable regions of the model parameter space.

6.2 Results

In the scalar sector, the ML training has been done with a population of 200 networks and $S = 10\,000$ points in the training and validation sets. (Note: The minimum that is searched for here refers to the minimum MAE, as defined in Eq. (5.2).)

For the Yukawa sector part, because the **SPheno** plus **flavio** setup has appeared to be a lot slower to run than the other calculations, about two orders of magnitude, it became extremely slow to train. Since a GA is used, the only way to speed up the process would be to lower the population and/or the number of points in the training and validation sets. For the setup used, to train the Yukawa sector in a reasonable time on one computer, the population and number of points had to be reduced too much, such that the algorithm did not seem to converge properly.

For these reasons, the Yukawa sector has not been trained on in the current results presented here. For future studies, it would be better to implement calculation of specific Wilson coefficients and flavor observables directly in the program instead of using external packages, as was done with the other targets, to speed up the process. Another possibility would be to

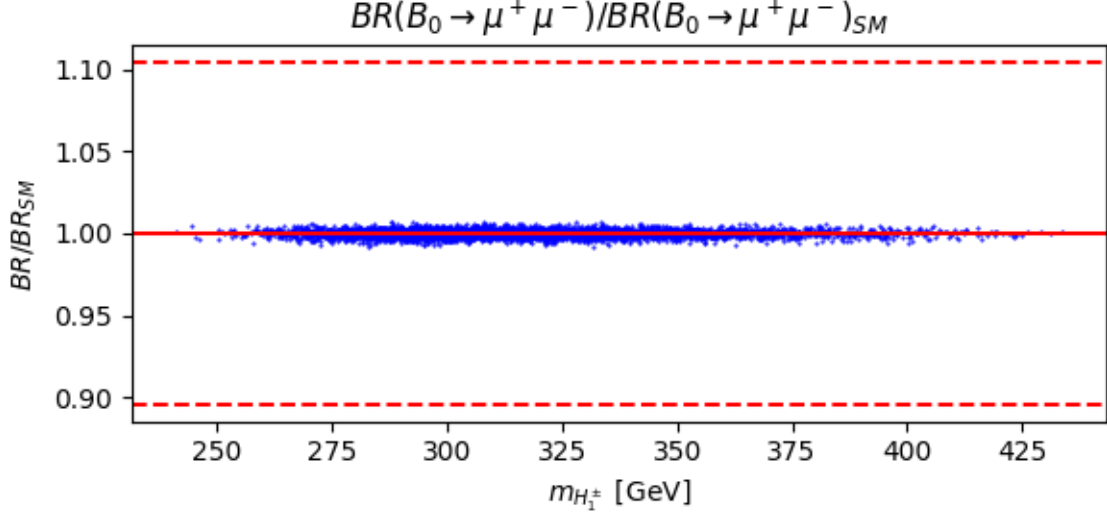


Figure 9: The branching ratio $\text{BR}(B_0 \rightarrow \mu^+ \mu^-)$ plotted against the charged H_1^\pm Higgs mass $m_{H_1^\pm}$. Dashed lines are 1σ SM errors.

use a computer cluster for training on the Yukawa sector, but since one of the goals of this program is to be able to find results in a reasonable time without a cluster if possible, the internal calculations would be preferred. Instead of training, points were passed through the trained scalar sector and tested against randomized input parameters in the BGL-limit of Yukawa textures for several flavor physics constraints such as the branching ratios $\text{BR}(B_0 \rightarrow \mu^+ \mu^-)$, $\text{BR}(B_s \rightarrow \mu^+ \mu^-)$, and $\text{BR}(B \rightarrow X_s \gamma)$ and the frequencies for $B_s^0 - \bar{B}_s^0$ and $B_d^0 - \bar{B}_d^0$ oscillations, ΔM_s and ΔM_d . Some other similar FCNC branching ratios were also computed, but were found less interesting to be presented here. The SM values for these FCNC observables, with errors, as given by `flavio` are as follows (with 3 decimals)

$$\begin{aligned}
 \text{BR}(B_0 \rightarrow \mu^+ \mu^-)_{SM} &= (1.143 \pm 0.119) \cdot 10^{-10}, \\
 \text{BR}(B_s \rightarrow \mu^+ \mu^-)_{SM} &= (3.668 \pm 0.155) \cdot 10^{-9}, \\
 \text{BR}(B \rightarrow X_s \gamma)_{SM} &= (3.291 \pm 0.218) \cdot 10^{-4}, \\
 \Delta M_{sSM} &= (1.250 \pm 0.075) \cdot 10^{-11}, \\
 \Delta M_{dSM} &= (3.978 \pm 0.516) \cdot 10^{-13}.
 \end{aligned} \tag{6.1}$$

The uncertainties are calculated in `flavio` by drawing random values according to the probability distribution of the observable. The relative error of the uncertainty $\Delta\sigma/\sigma = 1/\sqrt{2N}$ where N is the number values drawn. Here, $N = 100\,000$ was used, giving a relative uncertainty $\Delta\sigma/\sigma \approx 0.00224$.

Figs. 9 and 10 show the branching ratios of the $\text{BR}(B_0 \rightarrow \mu^+ \mu^-)$ and $\text{BR}(B_s \rightarrow \mu^+ \mu^-)$ processes, respectively, over the corresponding SM values, plotted against the charged mass $m_{H_1^\pm}$. In Fig. 9, a “cloud” of points are very evenly spread out around the SM value, while in Fig. 10, almost all of the points form two distinct clouds around the SM value, at ratios of 0.99 and 1.01, approximately. The same pattern emerges in $\text{BR}(B_{0/s} \rightarrow e^+ e^-)$ and $\text{BR}(B_{0/s} \rightarrow$

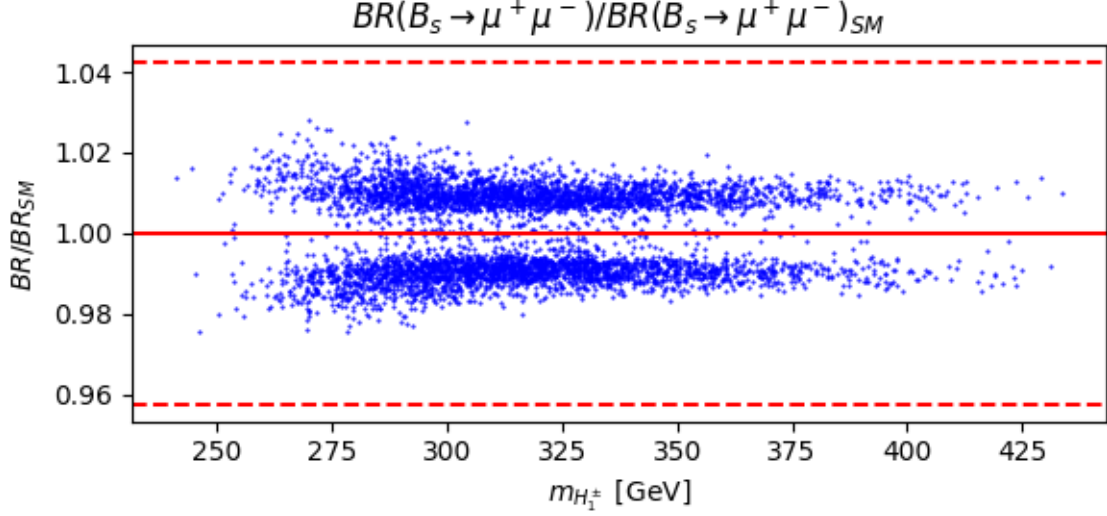


Figure 10: The branching ratio $\text{BR}(B_s \rightarrow \mu^+ \mu^-)$ plotted against the charged H_1^\pm Higgs mass $m_{H_1^\pm}$. Dashed lines are 1σ SM errors.

$\tau^+ \tau^-$) observables. The splitting into distinct lines for $\text{BR}(B_s \rightarrow \mu^+ \mu^-)$ resembles a similar finding in the previous work of Ref. [27], but the allowed regions here are more spread and further apart than the ones in Ref. [27]. For $\text{BR}(B_{0/s} \rightarrow e^+ e^-)$ there is the splitting seen in the previous work that is not seen here. That splitting was also not centered around the SM value, lines forming around 1.03 and 1.07, which is a sharp contrast to the cloud centered around the SM value obtained in the present work. Besides, $\text{BR}(B \rightarrow X_s \gamma)$, as seen in Fig. 11, also has a split but seems to move towards the SM value for larger $m_{H_1^\pm}$, which is in agreement with the previous work. The ΔM_s and ΔM_d oscillation frequencies are depicted in Fig. 12 and are plotted against each other, as ratios with their SM values. Fig. 12 shows the same plot with the SM errors introduced in Eq. 6.1. They stay in a very concentrated spot, but they are not completely aligned with the measured values in Eq. (2.24), being shifted by about $0.64 \cdot 10^{-13}$ in the ΔM_d direction and $0.08 \cdot 10^{-11}$ in the ΔM_s direction from their respective central values. As seen in Fig. 12

The observed differences between the numerical results obtained in this project and the previous similar work might occur due to differences in the formulation of the BGL-3HDM, but also because of the differences in constraints. Indeed, we depart from the exact Higgs alignment limit used in the previous work, as well as employ the most generic soft flavour symmetry breaking sector. In addition, in Ref. [27], the Higgs sector phenomenology constraints were considered while they were not incorporated in the numerical results at the moment due to a slow computation of the corresponding observables (although they are implemented in the code already). This would be a well-defined next step for further studies, to improve on the phenomenological accuracy, but also to compare this analysis to the one used in Ref. [27]. These results show that in the BGL-limit of the Yukawa sector a lot of points appear to be close to the SM predictions for the key FCNC observables which means that the training should also work well on this sector if the speed of the calculations of observables are optimized or a computer cluster is used for such a training.

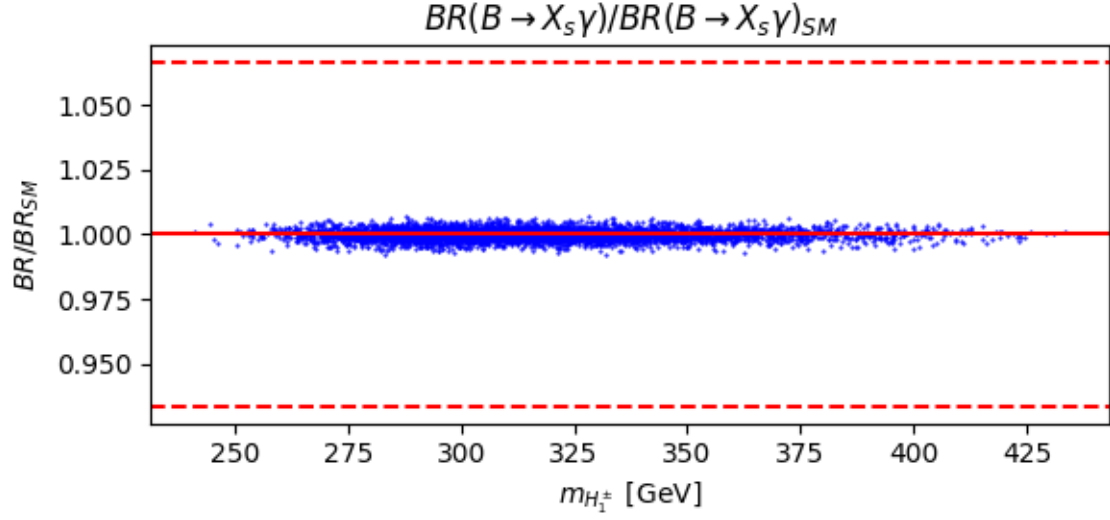


Figure 11: The branching ratio $\text{BR}(B \rightarrow X_s \gamma)$ plotted against the charged H_1^\pm Higgs mass $m_{H_1^\pm}$. Dashed lines are 1σ SM errors.

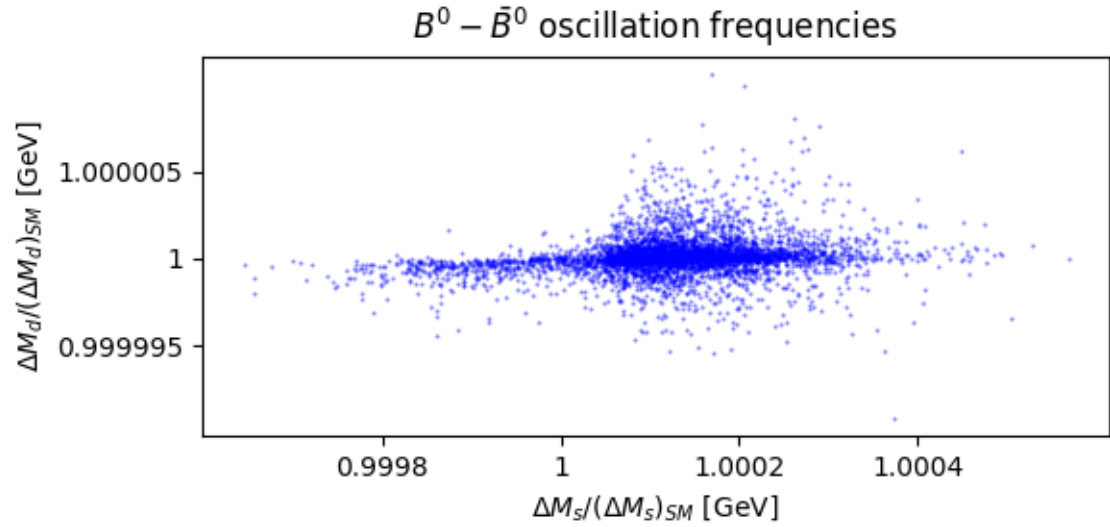


Figure 12: The $B - \bar{B}$ oscillation frequencies ΔM_s and ΔM_d , as ratios with their SM values, plotted against each other.

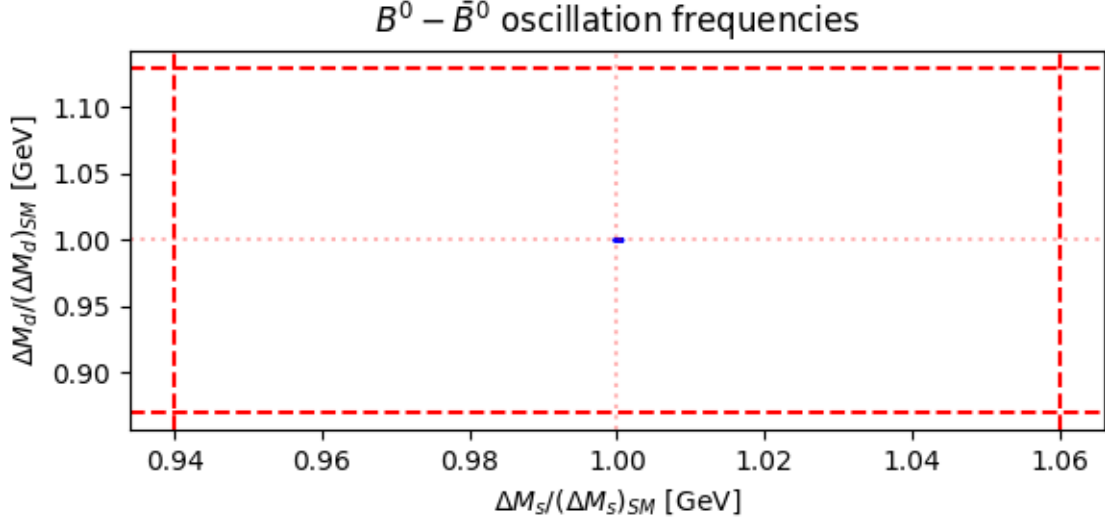


Figure 13: The $B - \bar{B}$ oscillation frequencies ΔM_s and ΔM_d , as ratios with their SM values, plotted against each other. Dashed lines are 1σ SM errors.

A scan of 10 000 000 Halton points was also made, and the number of points passed each individual set of scalar constraints was counted. About 1.7% of all points passed all of the oblique parameter constraints, 0.42% passed the BFB constraints and no points passed the unitarity constraints. This shows that it would be very difficult to get any points with just a random scan, since these are the pass rates for the individual constraints, but for a point to pass, it would need to pass all of these constraints. After the full training of the neural networks, over 50% of points passed all constraints.

The evolution of the errors for training on the set of unitarity constraints is shown in Fig. 14. It shows the errors evolved over iterations. It shows this for AveTrain error, which is the average training error over the entire population, BestTrain error, which is the average training error for the best four networks, and the Val error, which is the average validation error for the best four networks. The spikes in AveTrain are because the UNDX algorithm does not always put better networks into the population, to keep the gene pool varied, it sometimes keeps worse networks. This is the reason an average over the entire population was not used to validate for the early stop condition, but instead only a few of the best networks. The spikes in BestTrain are because the size of the training and validation sets, $S = 10\,000$, was too small. It needs to train with the data sets that were bad representations of the actual input space, but after a while of training, the networks transformed the input space in such a way that even these smaller data sets were representative. When training on the bad representations, it often hits early stop conditions and often switches from one badly representative data set to another, which changed the error wildly. A larger S should be considered for future use to get a smoother curve.

Figs. 15 and 16 show how the \mathcal{CP} -even Higgs masses H_1, H_2 , the \mathcal{CP} -odd Higgs masses A_1, A_2 and the charged Higgs masses H_1^\pm, H_2^\pm change with the different stages of training. The masses are ordered such that $H_1 < H_2$, $A_1 < A_2$ and $H_1^\pm < H_2^\pm$. The left side of Fig. 15 shows the distributions without any training, sampled directly with Halton sequences. The

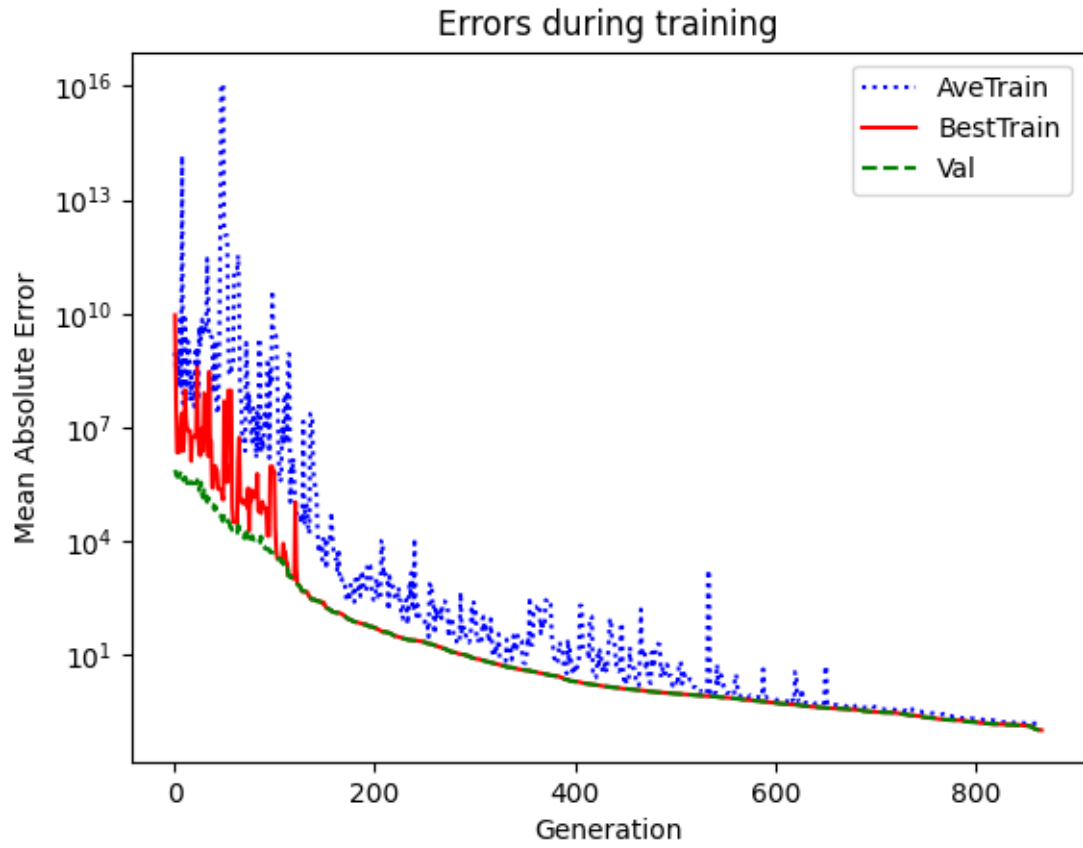


Figure 14: The average training error over the entire population (AveTrain), the average training error of the best four networks (BestTrain) and the average validation error of the best four networks (Val) are shown for every generation of training for unitarity constraints.

plots do not seem uniform and show a bias for larger masses, this is because the squared masses were uniformly sampled instead of the masses. The right side of Fig. 15 shows the distributions after the networks have been trained on the unitarity constraints. The points after training are generated by the best four networks. Fig. 16 similarly shows the distributions after the BFB constraints on the left and the distributions of the fully trained networks after the oblique parameter constraints on the right.

Before training, the masses are spread out over the entire input space, between 200 GeV and 2 TeV. The global minimum for the unitarity constraints lead to all the masses being in the 300 – 800 GeV range, with $m_{H_1} \approx m_{H_2}$, $m_{A_1} \approx m_{A_2}$ and $m_{H_1^\pm} \approx m_{H_2^\pm}$. After the BFB constraints, the range moved to 200 – 450 GeV for \mathcal{CP} -even Higgs masses, 200 – 350 GeV for \mathcal{CP} -odd Higgs masses and around 300 – 600 GeV for charged Higgs masses. Here, the charged Higgs particles’ masses are still mostly close to each other, but the neutral scalar states have a larger spread. After the oblique parameter constraints, the ranges have increased slightly again for neutral scalar masses, to between 200 – 500 GeV for \mathcal{CP} -even Higgs masses, mostly 200 – 400 GeV for \mathcal{CP} -odd Higgs masses. For charged Higgs masses, the range was constrained to 250 – 475 GeV. The \mathcal{CP} -odd scalar masses seemed to be constrained the most for all stages. The typical masses found in the analysis were also close to the lower limit of 200 GeV, such that reducing this limit might be very interesting for future studies.

These plots show a big difference between the ML approach and the random scan approach. If one compares the plots passing all constraints in Fig. 16 and compare to the previous work with a similar model [27], it is clear that the points generated by the ML algorithm are much less spread apart. The learning pushes networks to a global minimum, and the networks will then be able to generate points around that minimum, but the points in the global minimum are not the only points that can pass the constraints. Even in a local minimum, the points might also be able to pass constraints, which means that the random scans will hit the points that pass in the entire input space, while the ML will only find points around a global minimum. The range of the parameters found in the global minimum will continue to shrink with the continued training, since the networks need to more efficiently transform the entire input space into points that pass. Because of this, it might be better to train for a smaller convergence pass rate, to possibly get values in a more spread out range. Generating points after the network has been trained goes very quickly, so one could get a number of points even with a pass rate in the small percentages. In this project, a pass rate of 50% was used, but for future studies, it would be interesting to see if smaller pass rates give a better spread of parameters. To find more spread out points, one could also split the input space into different, unexplored areas to see where the minimum would lie there. Here, we start with the scalar boson masses within 200 – 2000 GeV range, but mostly ended up in the 200 – 450 GeV range after training. If the masses were forced to be larger in e.g. 1000 – 2000 GeV interval, then the minimum could end up close to 1000 GeV, possibly indicating that the model works better with lower masses over-all, or it could lie in the ranges of larger masses, suggesting that the error space is more complicated than that.

Fig. 17 shows the calculated normal distributions of the VEVs after the full training. The mean of v_1 was at 130.5 GeV with a standard deviation of 7.5 GeV. For v_2 and v_3 , the means were found to be –122.7 and 167.0 GeV, respectively, with standard deviations of 20.9 and 12.6 GeV, respectively. A specific hierarchy between the generations of the VEVs could have been a nice explanation for the hierarchy between the fermions, but these VEVs do not seem to show any particularly strong hierarchy in the current results.

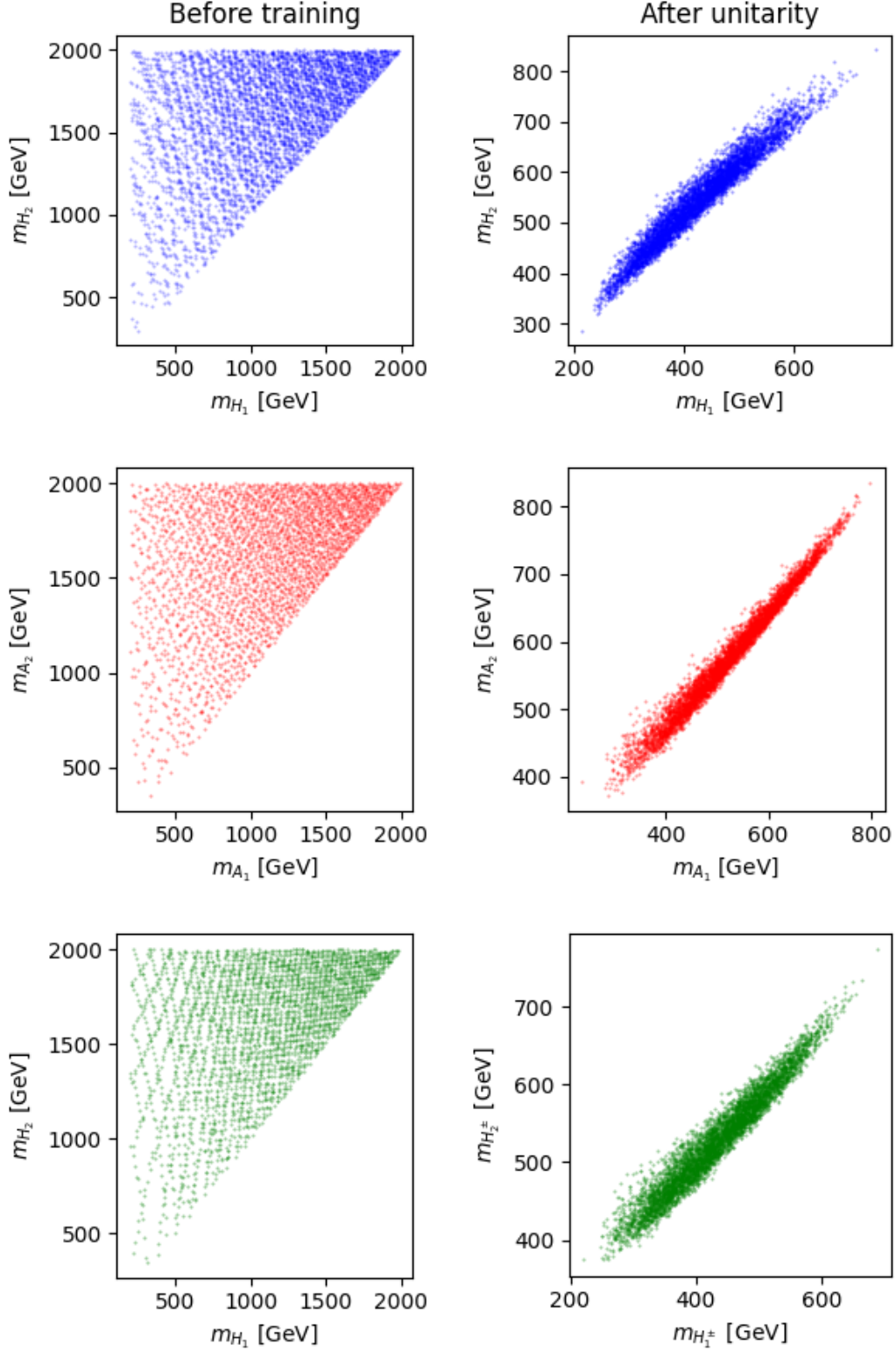


Figure 15: The distribution of \mathcal{CP} -even Higgs masses H_1, H_2 , \mathcal{CP} -odd Higgs masses A_1, A_2 and charged Higgs masses H_1^\pm, H_2^\pm . Left: Sampled before training, from Halton sequence; no constraints. Right: After unitarity constraints.

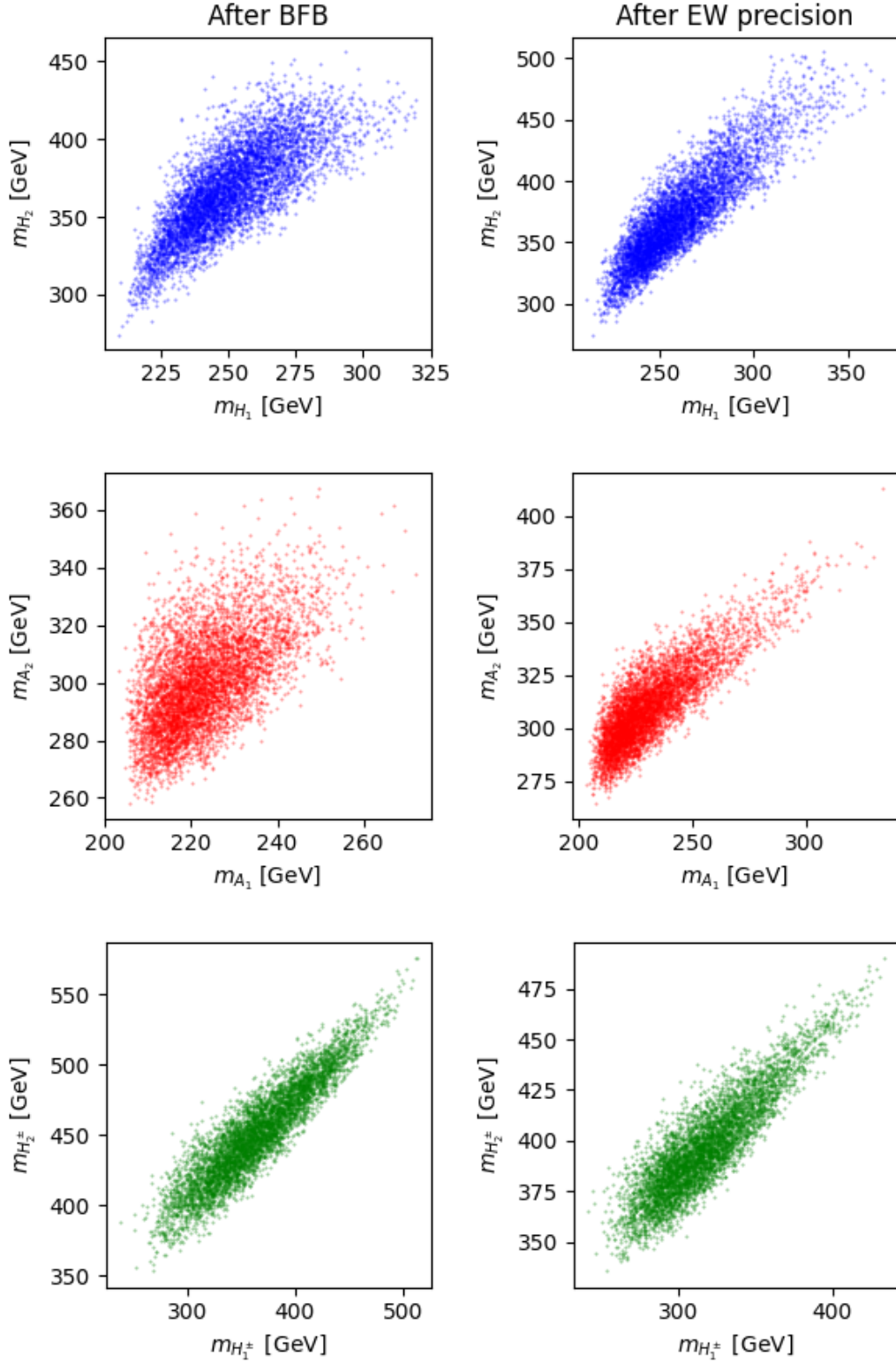


Figure 16: The distribution of \mathcal{CP} -even Higgs masses H_1, H_2 , \mathcal{CP} -odd Higgs masses A_1, A_2 and charged Higgs masses H_1^\pm, H_2^\pm . Left: After unitarity and BFB constraints. Right: After unitarity, BFB and EW precision constraints.

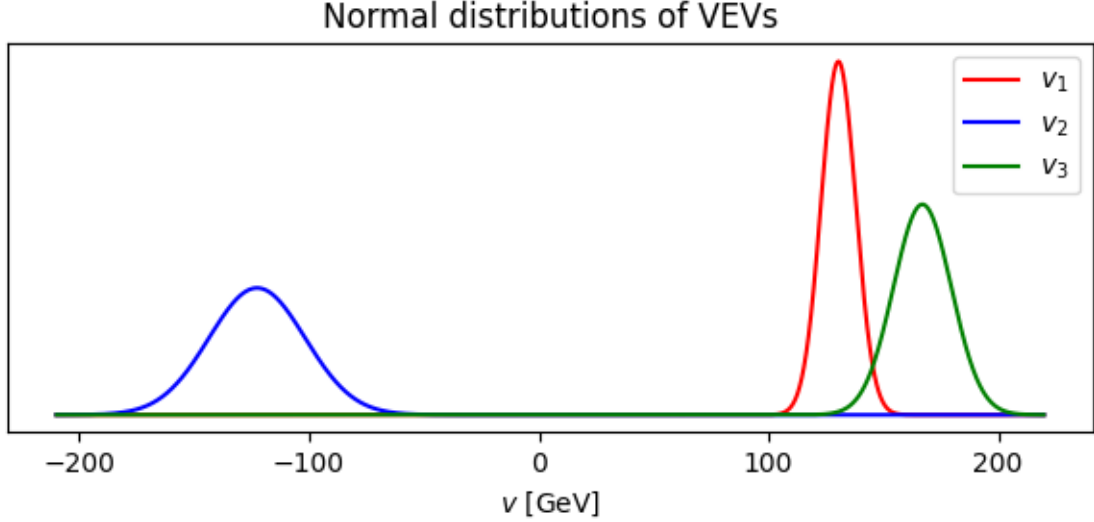


Figure 17: Calculated normal distributions of the VEVs v_1 , v_2 and v_3 found after the full training.

Fig. 18 shows the oblique parameters S , T and U plotted against each other. Most points for both S and U were found comfortably inside their bounds, but T seemed to prefer values towards its upper bound. In general, S seems to increase with larger T , while U seems to decrease with larger T .

7 Conclusion

Machine Learning tools can be very powerful for parameter searches in complicated new physics models. In this thesis, we have developed such a tool to explore the parameter space of a recently proposed Three Higgs Doublet Model with a $U(1) \times \mathbb{Z}_2$ flavor symmetry implementing the unitarity and boundedness from below constraints as well as the experimental bounds on oblique corrections. We have employed a sophisticated inversion procedure for both the scalar and Yukawa sectors of the model enabling us to select the points already within the experimentally allowed ranges of physical masses and measured mixing angles, as well as with small off-Higgs-alignment corrections.

For the scalar sector, our algorithm has been trained well and quickly, in total the training procedure over three sequentially applied sets of the scalar sector constraints took less than one day on a personal computer. A similar analysis for a random scan that could have revealed a sizable amount of points passing the same sets of constraints would take weeks of running time at a large computer cluster. Our trained networks can be used to produce millions of points that pass all the implemented constraints in a matter of minutes. Since the Machine Learning is made to force the neural networks into the global minimum of the error, the points are created much closer to each other than the points generated by a random scan as the latter can find points that pass into local minima as well.

The program was fast when the calculations of observables that are the subject of constraints were fast. However, when using the external packages, in our case, **SPheno** and

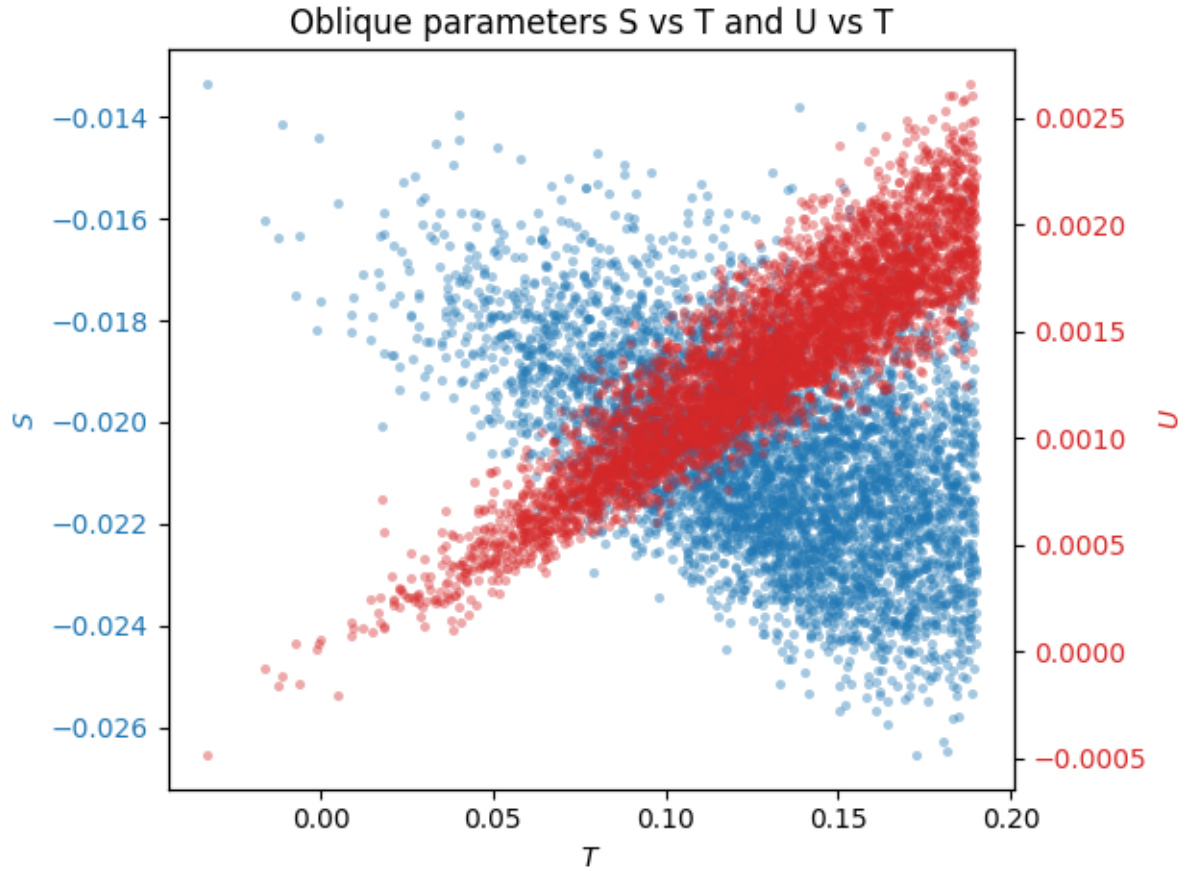


Figure 18: Oblique parameters S and U plotted against T for points that passed constraints.

`flavio` that are used to calculate the flavor constraints, the training has naturally become a lot slower. In this project, a good midway point between the time spent on training and possible convergence has not been found suggesting that the calculations of observables should be made internal to the program and more time-efficient, or if that is still slow, a more powerful computer, or a cluster of computers, would become necessary. The cluster would only need to be used on the slowest training steps though, which would drastically reduce the overall cluster running time, and would almost certainly still be a lot faster than a random scan.

The minimum found by the algorithm has preferred the smaller scalar boson masses. i.e. close to the lower bound imposed on them in the input space. While the search for all the non-SM scalar boson masses was done between 200 GeV and 2000 GeV, the ranges found by the algorithm were reduced down to the 200 – 450 GeV after training. Since 200 GeV was the lowest permitted mass in the input, this might suggest that even smaller masses should be considered. However, that might contradict phenomenological exclusion bounds on new physics scalars, which are yet to be implemented.

One possible upside with a 3HDM is that a hierarchy in the Higgs VEVs could be related to the mass hierarchies in the fermion sector, but such a hierarchy was not found here. Instead, the VEVs were all of the relatively similar size, with mean values found to be 130.5, -122.7 and 167.0 GeV for $v_{1,2,3}$, respectively.

The FCNCs seem to be well constrained and are under control as they should for a BGL-like model. The ΔM_d and ΔM_s observables seem to be constrained to very specific values, but they differ from the measurements. The branching ratios of various B -mesons show split clouds of points around the SM value, similar to what has been found in the previous work in Ref. [27], but B_0 shows no such split, in contradiction to it. This might be due some differences in the model structure and/or in phenomenological treatment since additional Higgs physics constraints were included in Ref. [27] and are not included in the present work.

For future development, more constraints should be added to the scalar sector, such as constraints on Higgs physics and vacuum stability, if possible calculated internally in the program for a better efficiency. Other steps would be to expand the model by introducing lepton flavor violation and \mathcal{CP} violation. Adding calculations for specific Wilson coefficients and FCNC constraints in the program would make searches in non-BGL multi-Higgs models possible. If this is implemented, this could for simplicity be tested for a near-BGL limit, and one could investigate if training on a few FCNC constraints would constrain the rest of the FCNCs, or if one would need to train on all of them for the full flavor constraints.

Acknowledgements

I am very grateful to my supervisor Roman Pasechnik for the guidance and support during the project, I have learned a great deal over the past year. I also want to thank António Morais, João Pedro and Felipe Freitas for helpful discussions and advice. Huge thank you to my family for listening to all my problems, even when they do not understand, and to my roommates for their continuous support.

A Higgs mechanism: a toy model

The Higgs mechanism for spontaneous symmetry breaking (SSB) is best illustrated in a toy-model, with a complex (charged) scalar field χ whose potential invariant under global $U(1)$

symmetry reads

$$V(\chi) = \mu^2 \chi^\dagger \chi + \lambda (\chi^\dagger \chi)^2, \quad (\text{A.1})$$

in terms of the mass term of the field, μ^2 , and its quartic interaction strength, λ . If $\mu^2 > 0$, it corresponds to the physical mass squared of χ field, and the minimum of the potential energy is identified with $\chi = 0$. Such a potential preserves the original Abelian U(1) symmetry of the theory, see Fig. 19a. However, if $\mu^2 < 0$, the field χ acquires a constant VEV, $\langle \chi \rangle = \text{const}$, in the minimum of the scalar potential found by means of the extremum condition, also known as the tadpole equation,

$$\left\langle \frac{\partial V(\chi)}{\partial \chi} \right\rangle \equiv \frac{\partial V(\chi)}{\partial \chi} \Big|_{\chi=\langle \chi \rangle} = 0 \quad \rightarrow \quad \langle \chi \rangle = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}}. \quad (\text{A.2})$$

In this case, the U(1) symmetry appears to be broken such that a particle ends up in one of the two possible minima, $\chi = v$ and $\chi = -v$, illustrated in Fig. 19b. A positive frequency of oscillations around the new stable minimum, and hence a positive mass of the physical scalar boson is ensured by the minimization condition

$$\left\langle \frac{\partial^2 V(\chi)}{\partial \chi^2} \right\rangle > 0, \quad (\text{A.3})$$

applied for solutions of Eq. (A.2). Clearly, a particular minimum that is chosen by the particle appears to be more energetically favourable than unstable configuration with $\chi = 0$.

The Goldstone theorem states that massless scalar (Goldstone) bosons appear in the theory when continuous symmetries are spontaneously broken – one for every broken generator [66, 67]. Upon gauging the U(1) symmetry and parameterising the complex scalar field as

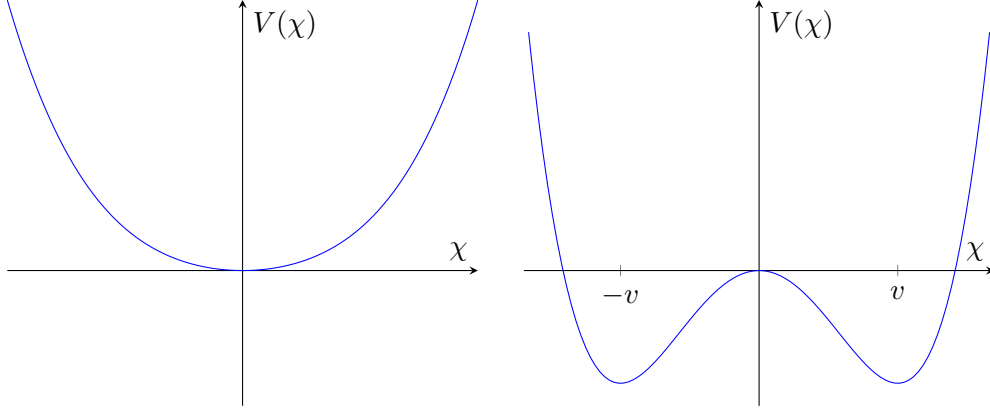
$$\chi(x) = \frac{v + h(x)}{\sqrt{2}} e^{i\theta(x)/v}, \quad (\text{A.4})$$

in terms of the two real scalar fields – the Higgs field $h(x)$ with mass $m_h = \sqrt{2}\mu$, and the massless Goldstone field, $\theta(x)$ – one can easily demonstrate that the initially massless and transversely polarised photon acquires a non-zero mass given by a product of the Higgs VEV v and the U(1) gauge coupling and gains a third (longitudinal) polarisation through effectively “absorbing” the Goldstone field $\theta(x)$. One also finds that the Goldstone field $\theta(x)$ can be removed by a gauge transformation of the U(1) gauge field in a specific choice of the gauge known as the unitary gauge. In this sense, the Goldstone mode completely disappears from the physical spectrum of the theory, being “eaten up” to provide a non-zero mass to the photon.

This illustration makes a rather pedagogical introduction into the non-Abelian Higgs mechanism of spontaneous EW symmetry breaking realised the framework of the SM.

B Inverted procedure equations list

Here all the equations defining the Lagrangian parameters in the inverted procedure of the scalar sector are listed. Here, $s_x = \sin(x)$ and $c_x = \cos(x)$.



(a) Scalar field potential with $\mu^2 > 0$. (b) Scalar field potential with $\mu^2 < 0$.

Figure 19: The two possible scenarios for the scalar field potential $V(\chi)$.

$$\begin{aligned} \mu_{13}^2 = & \frac{1}{2(v_{13}^2)(v_{13}^2 + v_2^2)^2} \left\{ (m_{A_1}^2 - m_{A_2}^2) \left[vv_2 s_{2\gamma_1} (v_1 - v_3)(v_1 + v_3)(v_{13}^2 + v_2^2) \right. \right. \\ & \left. \left. - v_1 v_3 c_{2\gamma_1} \left\{ v^2 v_2^2 + (v_{13}^2 + v_2^2)^2 \right\} \right] \right. \\ & \left. + v_1 v_3 (-m_{A_1}^2 - m_{A_2}^2) [v_2(v_2 - v) + v_{13}^2] [v_2(v + v_2) + v_{13}^2] \right\} - 2\lambda_{10} v_1 v_3 \end{aligned} \quad (\text{B.1})$$

$$\mu_{23}^2 = -\frac{v}{(v_{13}^2 + v_2^2)^2} \left\{ v_1 (m_{A_1}^2 - m_{A_2}^2) s_{\gamma_1} c_{\gamma_1} (v_{13}^2 + v_2^2) + m_{A_1}^2 v v_2 v_3 s_{\gamma_1}^2 + m_{A_2}^2 v v_2 v_3 c_{\gamma_1}^2 \right\} \quad (\text{B.2})$$

$$\mu_{12}^2 = -\frac{v}{(v_{13}^2 + v_2^2)^2} \left\{ v_3 (m_{A_2}^2 - m_{A_1}^2) s_{\gamma_1} c_{\gamma_1} (v_{13}^2 + v_2^2) + m_{A_1}^2 v v_1 v_2 s_{\gamma_1}^2 + m_{A_2}^2 v v_1 v_2 c_{\gamma_1}^2 \right\} \quad (\text{B.3})$$

$$\lambda_7 = \frac{1}{v_1 v_2} \left\{ -2\mu_{12}^2 + \frac{v}{(v_{13}^2 + v_2^2)^2} \left[v_3(m_{H_1^\pm}^2 - m_{H_2^\pm}^2) s_{2\gamma_2} (v_{13}^2 + v_2^2) \right. \right. \\ \left. \left. - 2vv_1 v_2 (m_{H_1^\pm}^2 s_{\gamma_2}^2 + m_{H_2^\pm}^2 c_{\gamma_2}^2) \right] \right\}. \quad (\text{B.4})$$

$$\lambda_8 = \frac{2}{v_1 v_3} \left\{ \frac{1}{2(v_{13}^2)(v_{13}^2 + v_2^2)^2} \left((m_{H_1^\pm}^2 - m_{H_2^\pm}^2) \left[vv_2 s_{2\gamma_2} (v_1 - v_3)(v_1 + v_3)(v_{13}^2 + v_2^2) \right. \right. \right. \\ \left. \left. - v_1 v_3 c_{2\gamma_2} \{v^2 v_2^2 + (v_{13}^2 + v_2^2)^2\} \right] + v_1 v_3 (-m_{H_1^\pm}^2 - m_{H_2^\pm}^2) \left[v_2 (v_2 - v) \right. \right. \\ \left. \left. + v_{13}^2 \right] \left[v_2 (v + v_2) + v_{13}^2 \right] \right) - \lambda_{10} v_1 v_3 - \mu_{13}^2 \right\} \quad (\text{B.5})$$

$$\lambda_9 = \frac{1}{v_1 v_2} \left\{ -2\mu_{23}^2 + \frac{v}{(v_{13}^2 + v_2^2)^2} \left[v_1 (m_{H_2^\pm}^2 - m_{H_1^\pm}^2) s_{2\gamma_2} (v_{13}^2 + v_2^2) \right. \right. \\ \left. \left. - 2vv_2 v_3 (m_{H_1^\pm}^2 s_{\gamma_2}^2 + m_{H_2^\pm}^2 c_{\gamma_2}^2) \right] \right\} \quad (\text{B.6})$$

$$\lambda_1 = \frac{1}{2v_1^3} \left\{ v_1 c_{\alpha_1}^2 \left[s_{\alpha_2}^2 (m_h^2 s_{\alpha_3}^2 + m_{H_2}^2 c_{\alpha_3}^2) + m_h^2 c_{\alpha_2}^2 \right] + v_1 s_{\alpha_1} c_{\alpha_1} s_{\alpha_2} s_{2\alpha_3} (m_h^2 - m_{H_2}^2) \right. \\ \left. + v_1 s_{\alpha_1}^2 (m_h^2 c_{\alpha_3}^2 + m_{H_2}^2 s_{\alpha_3}^2) + \mu_{12}^2 v_2 + \mu_{13}^2 v_3 \right\} \quad (\text{B.7})$$

$$\lambda_2 = \frac{1}{2v_2^3} \left\{ v_2 s_{\alpha_1}^2 \left[s_{\alpha_2}^2 (m_h^2 s_{\alpha_3}^2 + m_{H_2}^2 c_{\alpha_3}^2) + m_h^2 c_{\alpha_2}^2 \right] + v_2 s_{\alpha_1} c_{\alpha_1} s_{\alpha_2} s_{2\alpha_3} (m_{H_2}^2 - m_h^2) \right. \\ \left. + v_2 c_{\alpha_1}^2 (m_h^2 c_{\alpha_3}^2 + m_{H_2}^2 s_{\alpha_3}^2) + \mu_{12}^2 v_1 + \mu_{23}^2 v_3 \right\} \quad (\text{B.8})$$

$$\lambda_3 = \frac{1}{2v_3^3} \left\{ v_3 c_{\alpha_2}^2 (m_h^2 s_{\alpha_3}^2 + m_{H_2}^2 c_{\alpha_3}^2) + m_h^2 v_3 s_{\alpha_2}^2 + \mu_{13}^2 v_1 + \mu_{23}^2 v_2 \right\} \quad (\text{B.9})$$

$$\lambda_4 = \frac{1}{v_1 v_2} \left\{ -\mu_{12}^2 - \lambda_7 v_1 v_2 + c_{\alpha_1}^2 \left[s_{\alpha_2}^2 (m_h^2 s_{\alpha_3}^2 + m_{H_2}^2 c_{\alpha_3}^2) + m_h^2 c_{\alpha_2}^2 \right] \right. \\ \left. + s_{\alpha_1} c_{\alpha_1} s_{\alpha_2} s_{2\alpha_3} (m_h^2 - m_{H_2}^2) + s_{\alpha_1}^2 (m_h^2 c_{\alpha_3}^2 + m_{H_2}^2 s_{\alpha_3}^2) \right\} \quad (\text{B.10})$$

$$\lambda_5 = -\frac{1}{v_1 v_3} \left\{ \mu_{13}^2 + v_1 v_3 (2\lambda_{10} + \lambda_8) + c_{\alpha_1} s_{\alpha_2} c_{\alpha_2} (m_h^2 s_{\alpha_3}^2 + m_{H_2}^2 c_{\alpha_3}^2 - m_h^2) \right. \\ \left. + s_{\alpha_1} c_{\alpha_2} s_{\alpha_3} c_{\alpha_3} (m_h^2 - m_{H_2}^2) \right\} \quad (\text{B.11})$$

$$\lambda_6 = -\frac{1}{v_2 v_3} \left\{ \mu_{23}^2 + \lambda_9 v_2 v_3 + c_{\alpha_2} \left[s_{\alpha_1} s_{\alpha_2} (m_h^2 s_{\alpha_3}^2 + m_{H_2}^2 c_{\alpha_3}^2 - m_h^2) \right. \right. \\ \left. \left. + c_{\alpha_1} s_{\alpha_3} c_{\alpha_3} (m_{H_2}^2 - m_h^2) \right] \right\} \quad (\text{B.12})$$

C Oblique parameters

All the functions for the oblique parameters not mentioned in Section 3.2.4 are mentioned here, as found in Ref. [32]. First, the oblique parameters and then the functions used in their calculations.

$$\bar{S} = \frac{g^2}{384\pi^2} \left[\sum_{a=2}^n [2s_W^2 - (\mathcal{U}^\dagger \mathcal{U})_{aa}]^2 G(m_a^2, m_a^2, m_Z^2) \right. \\ + 2 \sum_{a=2}^{n-1} \sum_{a'=a+1}^n |(\mathcal{U}^\dagger \mathcal{U})_{aa'}|^2 G(m_a^2, m_{a'}^2, m_Z^2) \\ + \sum_{b=2}^{m-1} \sum_{b'=b+1}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{bb'}]^2 G(M_b^2, M_{b'}^2, m_Z^2) \\ - 2 \sum_{a=2}^n (\mathcal{U}^\dagger \mathcal{U})_{aa} \ln m_a^2 + \sum_{b=2}^m (\mathcal{V}^\dagger \mathcal{V})_{bb} \ln M_b^2 - \ln m_h^2 \\ \left. + \sum_{b=2}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{1b}] \hat{G}(M_b^2, m_Z^2) - \hat{G}(m_h^2, m_Z^2) \right] \quad (\text{C.1})$$

$$\begin{aligned}
\bar{U} = & \frac{g^2}{384\pi^2 c_W^2} \left[\sum_{a=2}^n \sum_{b=2}^m |(\mathcal{U}^\dagger \mathcal{V})_{ab}|^2 G(m_a^2, M_b^2, m_W^2) \right. \\
& - \sum_{a=2}^n [2s_W^2 - (\mathcal{U}^\dagger \mathcal{U})_{aa}]^2 G(m_a^2, m_a^2, m_Z^2) \\
& - 2 \sum_{a=2}^{n-1} \sum_{a'=a+1}^n |(\mathcal{U}^\dagger \mathcal{U})_{aa'}|^2 G(m_a^2, m_{a'}^2, m_Z^2) \\
& - \sum_{b=2}^{m-1} \sum_{b'=b+1}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{bb'}]^2 G(M_b^2, M_{b'}^2, m_Z^2) \\
& \left. + \sum_{b=2}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{1b}]^2 \left[\hat{G}(M_b^2, m_W^2) - \hat{G}(M_b^2, m_Z^2) \right] \right] \quad (C.2)
\end{aligned}$$

$$\bar{X} = - \frac{g^2 s_W}{192\pi^2 c_W} \sum_{a=2}^n [2s_W^2 - (\mathcal{U}^\dagger \mathcal{U})_{aa}]^2 G(m_a^2, m_a^2, m_Z^2) \quad (C.3)$$

$$\begin{aligned}
\bar{V} = & \frac{g^2}{384\pi^2 c_W^2} \left[\sum_{a=2}^n [2s_W^2 - (\mathcal{U}^\dagger \mathcal{U})_{aa}]^2 H(m_a^2, m_a^2, m_Z^2) \right. \\
& + 2 \sum_{a=2}^{n-1} \sum_{a'=a+1}^n |(\mathcal{U}^\dagger \mathcal{U})_{aa'}|^2 H(m_a^2, m_{a'}^2, m_Z^2) \\
& + \sum_{b=2}^{m-1} \sum_{b'=b+1}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{bb'}]^2 H(M_b^2, M_{b'}^2, m_Z^2) \\
& \left. + \sum_{b=2}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{1b}]^2 \hat{H}(M_b^2, m_Z^2) - \hat{H}(M_b^2, m_Z^2) \right] \quad (C.4)
\end{aligned}$$

$$\begin{aligned}
\bar{W} = & \frac{g^2}{384\pi^2} \left[\sum_{a=2}^n \sum_{b=2}^m |(\mathcal{U}^\dagger \mathcal{V})_{ab}|^2 H(m_a^2, M_b^2, m_W^2) \right. \\
& \left. + \sum_{b=2}^m [\text{Im}(\mathcal{V}^\dagger \mathcal{V})_{1b}]^2 \hat{H}(M_b^2, m_W^2) - \hat{H}(M_b^2, m_W^2) \right] \quad (C.5)
\end{aligned}$$

$$\begin{aligned}
G(I, J, Q) = & -\frac{16}{3} + \frac{5(I+J)}{Q} - \frac{2(I-J)^2}{Q^2} \\
& + \frac{3}{Q} \left[\frac{I^2 + J^2}{I-J} - \frac{I^2 - J^2}{Q} + \frac{(I-J)^3}{3Q^2} \right] \ln \frac{I}{J} + \frac{r}{Q} f(t, r) \quad (C.6)
\end{aligned}$$

$$\tilde{G}(I, J, Q) = -2 + \left(\frac{I-J}{Q} - \frac{I+J}{I-J} \right) \ln \frac{I}{J} + \frac{f(t, r)}{Q} \quad (C.7)$$

$$\hat{G}(I, Q) = G(I, Q, Q) + 12\tilde{G}(I, Q, Q) \quad (C.8)$$

$$\begin{aligned}
H(I, J, Q) = & 2 - \frac{9(I+J)}{Q} + \frac{6(I-J)^2}{Q^2} \\
& + \frac{3}{Q} \left[-\frac{I^2+J^2}{I-J} + 2\frac{I^2-J^2}{Q} - \frac{(I-J)^3}{Q^2} \right] \ln \frac{I}{J} \\
& + \left[I+J - \frac{(I-J)^2}{Q} \right] \frac{3f(t, r)}{Q^2}
\end{aligned} \tag{C.9}$$

$$\begin{aligned}
\tilde{H}(I, J, Q) = & 4 + \left(\frac{I+J}{I-J} - 2\frac{I-J}{Q} \right) \ln \frac{I}{J} \\
& + \frac{-Q^2 + 3Q(I-J) - 2(I-J)^2}{rQ} f(t, r)
\end{aligned} \tag{C.10}$$

$$\hat{H}(I, Q) = H(I, Q, Q) + 12\tilde{H}(I, Q, Q) \tag{C.11}$$

$$t = I + J - Q \tag{C.12}$$

$$r = Q^2 - 2Q(I+J) + (I-J)^2 \tag{C.13}$$

$$f(t, r) = \begin{cases} \sqrt{r} \ln \left| \frac{t - \sqrt{r}}{t + \sqrt{r}} \right| & \text{if } r > 0, \\ 0 & \text{if } r = 0, \\ 2\sqrt{-r} \arctan \frac{\sqrt{-r}}{t} & \text{if } r < 0. \end{cases} \tag{C.14}$$

References

- [1] G. Aad *et al.* [ATLAS Collaboration], *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1 [arXiv:1207.7214 [hep-ex]].
- [2] S. Chatrchyan *et al.* [CMS Collaboration], *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30 [arXiv:1207.7235 [hep-ex]].
- [3] J. Ellis, *Outstanding questions: Physics beyond the Standard Model*, Phil. Trans. Roy. Soc. Lond. A **370** (2012) 818.
- [4] D. E. Morrissey and M. J. Ramsey-Musolf, *Electroweak baryogenesis*, New J. Phys. **14** (2012) 125003 [arXiv:1206.2942 [hep-ph]].
- [5] G. Aad *et al.* [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** (2020) no.1, 012002 [arXiv:1909.02845 [hep-ex]].
- [6] A. P. Morais, R. Pasechnik and W. Porod, *Prospects for new physics from gauge Left-Right-Colour-Family Grand Unification* [arXiv:2001.06383 [hep-ph]].

- [7] G. C. Branco, L. Lavoura and J. P. Silva, *CP Violation*, Int. Ser. Monogr. Phys. **103** (1999) 1.
- [8] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1 [arXiv:1106.0034 [hep-ph]].
- [9] I. P. Ivanov, *Building and testing models with extended Higgs sectors*, Prog. Part. Nucl. Phys. **95** (2017) 160 [arXiv:1702.03776 [hep-ph]].
- [10] I. P. Ivanov and C. C. Nishi, *Abelian symmetries of the N-Higgs-doublet model with Yukawa interactions*, JHEP **1311** (2013) 069 [arXiv:1309.3682 [hep-ph]].
- [11] T. D. Lee, *A Theory of Spontaneous T Violation*, Phys. Rev. D **8** (1973) 1226.
- [12] T. D. Lee, *CP Nonconservation and Spontaneous Symmetry Breaking*, Phys. Rept. **9** (1974) 143.
- [13] R. González Felipe, I. P. Ivanov, C. C. Nishi, H. Serôdio and J. P. Silva, *Constraining multi-Higgs flavour models*, Eur. Phys. J. C **74** (2014) no.7, 2953 [arXiv:1401.5807 [hep-ph]].
- [14] S. L. Glashow and S. Weinberg, *Natural Conservation Laws for Neutral Currents*, Phys. Rev. D **15** (1977) 1958.
- [15] E. A. Paschos, *Diagonal Neutral Currents*, Phys. Rev. D **15** (1977) 1966. doi:10.1103/PhysRevD.15.1966
- [16] F. J. Botella, G. C. Branco, A. Carmona, M. Nebot, L. Pedro and M. N. Rebelo, *Physical Constraints on a Class of Two-Higgs Doublet Models with FCNC at tree level*, JHEP **1407** (2014) 078 [arXiv:1401.6147 [hep-ph]].
- [17] H. E. Haber and D. O’Neil, *Basis-independent methods for the two-Higgs-doublet model. II. The Significance of $\tan\beta$* , Phys. Rev. D **74** (2006) 015018 Erratum: [Phys. Rev. D **74** (2006) no.5, 059905] [hep-ph/0602242].
- [18] A. Pich and P. Tuzon, *Yukawa Alignment in the Two-Higgs-Doublet Model*, Phys. Rev. D **80** (2009) 091702 [arXiv:0908.1554 [hep-ph]].
- [19] P. M. Ferreira, I. P. Ivanov, E. Jiménez, R. Pasechnik and H. Serôdio, *CP4 miracle: shaping Yukawa sector with CP symmetry of order four*, JHEP **1801** (2018) 065 [arXiv:1711.02042 [hep-ph]].
- [20] J. E. Camargo-Molina, T. Mandal, R. Pasechnik and J. Wessén, *Heavy charged scalars from $c\bar{s}$ fusion: A generic search strategy applied to a 3HDM with $U(1) \times U(1)$ family symmetry*, JHEP **1803** (2018) 024 [arXiv:1711.03551 [hep-ph]].
- [21] V. Keus, S. F. King and S. Moretti, *Three-Higgs-doublet models: symmetries, potentials and Higgs boson masses*, JHEP **1401** (2014) 052 [arXiv:1310.8253 [hep-ph]].
- [22] I. P. Ivanov and E. Vdovin, *Discrete symmetries in the three-Higgs-doublet model*, Phys. Rev. D **86** (2012) 095030 [arXiv:1206.7108 [hep-ph]].

- [23] G. C. Branco, W. Grimus and L. Lavoura, *Relating the scalar flavor changing neutral couplings to the CKM matrix*, Phys. Lett. B **380** (1996) 119 [hep-ph/9601383].
- [24] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.
- [25] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.
- [26] F. J. Botella, G. C. Branco, M. Nebot and M. N. Rebelo, *Two-Higgs Leptonic Minimal Flavour Violation*, JHEP **1110** (2011) 037 [arXiv:1102.0520 [hep-ph]].
- [27] I. Padilla, *Phenomenology of a Three Higgs Doublet Model with a $U(1) \times \mathbb{Z}_2$ Flavour Symmetry*, (2019)
- [28] D. Bourilkov, *Machine and Deep Learning Applications in Particle Physics*, Int. J. Mod. Phys. A **34** (2020) no.35, 1930019 [arXiv:1912.08245 [physics.data-an]].
- [29] D. Guest, K. Cranmer and D. Whiteson, *Deep Learning and its Application to LHC Physics*, Ann. Rev. Nucl. Part. Sci. **68** (2018) 161 [arXiv:1806.11484 [hep-ex]].
- [30] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, Eur. Phys. J. C **76** (2016) no.5, 235 [arXiv:1601.07913 [hep-ex]].
- [31] M. E. Peskin and T. Takeuchi, *Estimation of oblique electroweak corrections*, Phys. Rev. D **46** (1992) 381.
- [32] W. Grimus, L. Lavoura, O. M. Ogreid and P. Osland, *The Oblique parameters in multi-Higgs-doublet models*, Nucl. Phys. B **801** (2008) 81 [arXiv:0802.4353 [hep-ph]].
- [33] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, (2018) [arXiv:1810.08132 [hep-ph]].
- [34] F. Staub, *Sarah*, (2008) [arXiv:0806.0538 [hep-ph]].
- [35] F. Staub, *Exploring new models in all detail with SARAH*, Adv. High Energy Phys. **2015** (2015) 840780 [arXiv:1503.04200 [hep-ph]].
- [36] F. Staub, *SARAH 4 : A tool for (not only SUSY) model builders*, Comput. Phys. Commun. **185** (2014) 1773 [arXiv:1309.7223 [hep-ph]].
- [37] W. Porod, *SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at $e^+ e^-$ colliders*, Comput. Phys. Commun. **153** (2003) 275 [hep-ph/0301101].
- [38] W. Porod and F. Staub, *SPheno 3.1: Extensions including flavour, CP-phases and models beyond the MSSM*, Comput. Phys. Commun. **183** (2012) 2458 [arXiv:1104.1573 [hep-ph]].
- [39] M. Tanabashi *et al.* [Particle Data Group], *Review of Particle Physics*, Phys. Rev. D **98** (2018) no.3, 030001.

- [40] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.
- [41] A. J. Buras and R. Fleischer, *Quark mixing, CP violation and rare decays after the top quark discovery*, Adv. Ser. Direct. High Energy Phys. **15** (1998) 65 [hep-ph/9704376].
- [42] M. Antonelli *et al.*, *Flavor Physics in the Quark Sector*, Phys. Rept. **494** (2010) 197 [arXiv:0907.5386 [hep-ph]].
- [43] L. Silvestrini, *Effective Theories for Quark Flavour Physics*, arXiv:1905.00798 [hep-ph].
- [44] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.
- [45] A. J. Buras, W. Slominski and H. Steger, *B0 anti-B0 Mixing, CP Violation and the B Meson Decay*, Nucl. Phys. B **245** (1984) 369.
- [46] A. J. Buras, W. Slominski and H. Steger, *B Meson Decay, CP Violation, Mixing Angles and the Top Quark Mass*, Nucl. Phys. B **238** (1984) 529.
- [47] D. Das, *Implications Of The Higgs Discovery On Physics Beyond The Standard Model*, [arXiv:1511.02195 [hep-ph]].
- [48] H. E. Haber and O. Stål, *New LHC benchmarks for the CP -conserving two-Higgs-doublet model* Eur. Phys. J. C **75** (2015) no.10, 491 Erratum: [Eur. Phys. J. C **76** (2016) no.6, 312] [arXiv:1507.04281 [hep-ph]].
- [49] S. Moretti and K. Yagyu, *Constraints on Parameter Space from Perturbative Unitarity in Models with Three Scalar Doublets*, Phys. Rev. D **91** (2015) 055022 [arXiv:1501.06544 [hep-ph]].
- [50] M. P. Bento, H. E. Haber, J. C. Romão and J. P. Silva, *Multi-Higgs doublet models: physical parametrization, sum rules and unitarity bounds*, JHEP **1711** (2017) 095 [arXiv:1708.09408 [hep-ph]].
- [51] I. Maksymyk, C. Burgess and D. London, *Beyond S, T and U*, Phys. Rev. D **50** (1994), 529-535 [arXiv:hep-ph/9306267 [hep-ph]].
- [52] R. Coimbra, M. O. Sampaio and R. Santos, *ScannerS: Constraining the phase diagram of a complex scalar singlet at the LHC*, Eur. Phys. J. C **73** (2013), 2428 [arXiv:1301.2599 [hep-ph]].
- [53] L. Calibbi and G. Signorelli, *Charged Lepton Flavour Violation: An Experimental and Theoretical Introduction*, Riv. Nuovo Cim. **41** (2018) no.2, 71 [arXiv:1709.00294 [hep-ph]].
- [54] F. D. Murnaghan, *The unitary and rotation groups*, volume 3. Spartan Books (1962)
- [55] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review **65** (1958) no.6 386

- [56] I. Ono, H. Kita and S. Kobayashi, *A Robust Real-Coded Genetic Algorithm Using Unimodal Normal Distribution Crossover Augmented by Uniform Crossover: Effects of Self-Adaptation of Crossover Probabilities* GECCO'99 **1** (1999)
- [57] D. G. Loyola R, M. Pedernana and S. Gimeno García, *Smart sampling and incremental function learning for very large high dimensional data*, Neural Networks **78** (2016)
- [58] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*, PMLR **9** (2010)
- [59] H. Kita, I. Ono and S. Kobayashi, *Theoretical analysis of the unimodal normal distribution crossover for real-coded genetic algorithms*, Cat. No.98TH8360 (1998)
- [60] O. Schuessler and D. Loyola, *Parallel Training of Artificial Neural Networks Using Multithreaded and Multicore CPUs*, ICANNGA'11 **1** (2011)
- [61] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st edition. Addison-Wesley Longman Publishing Co., Inc. (1989)
- [62] Y. Yoon and Y-H. Kim, *The Roles of Crossover and Mutation in Real-Coded Genetic Algorithms*, IntechOpen (2012)
- [63] J. H. Halton, *On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals*, Numer. Math. **2** (1960) no.1
- [64] L. Kocis and W. J. Whiten, *Computational Investigations of Low-Discrepancy Sequences*, ACM Trans. Math. Softw. **23** (1997) no.2
- [65] Z. z. Xing, H. Zhang and S. Zhou, *Updated Values of Running Quark and Lepton Masses*, Phys. Rev. D **77** (2008) 113016 [arXiv:0712.1419 [hep-ph]].
- [66] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19** (1961) 154.
- [67] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (1962) 965.