

# Quark masses and mixing in string-inspired models

**Andrei Constantin** <sup>a,d</sup> **Cristoforo S. Fraser-Taliente** <sup>a</sup> **Thomas R. Harvey** <sup>b,c</sup>  
**Lucas T.Y. Leung** <sup>a</sup> and **Andre Lukas** <sup>a</sup>

<sup>a</sup>*Rudolf Peierls Centre for Theoretical Physics, University of Oxford,  
Parks Road, Oxford OX1 3PU, U.K.*

<sup>b</sup>*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Massachusetts  
Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.*

<sup>c</sup>*Center for Theoretical Physics, Massachusetts Institute of Technology,  
77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.*

<sup>d</sup>*School of Mathematics, University of Birmingham,  
Watson Building, Edgbaston, Birmingham B15 2TT, U.K.*

*E-mail:* [andrei.constantin@physics.ox.ac.uk](mailto:andrei.constantin@physics.ox.ac.uk),  
[cristoforo.fraser-taliente@physics.ox.ac.uk](mailto:cristoforo.fraser-taliente@physics.ox.ac.uk), [trharvey@mit.edu](mailto:trharvey@mit.edu),  
[lucas.leung@physics.ox.ac.uk](mailto:lucas.leung@physics.ox.ac.uk), [andre.lukas@physics.ox.ac.uk](mailto:andre.lukas@physics.ox.ac.uk)

**ABSTRACT:** We study a class of supersymmetric Froggatt-Nielsen (FN) models with multiple U(1) symmetries and Standard Model (SM) singlets inspired by heterotic string compactifications on Calabi-Yau threefolds. The string-theoretic origin imposes a particular charge pattern on the SM fields and FN singlets, dividing the latter into perturbative and non-perturbative types. Employing systematic and heuristic search strategies, such as genetic algorithms, we identify charge assignments and singlet VEVs that replicate the observed mass and mixing hierarchies in the quark sector, and subsequently refine the Yukawa matrix coefficients to accurately match the observed values for the Higgs VEV, the quark and charged lepton masses and the CKM matrix. This bottom-up approach complements top-down string constructions and our results demonstrate that string FN models possess a sufficiently rich structure to account for flavour physics. On the other hand, the limited number of distinct viable charge patterns identified here indicates that flavour physics imposes tight constraints on string theory models, adding new constraints on particle spectra that are essential for achieving a realistic phenomenology.

**KEYWORDS:** String and Brane Phenomenology, String Models, Superstrings and Heterotic Strings

**ARXIV EPRINT:** [2410.17704](https://arxiv.org/abs/2410.17704)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>String Froggatt-Nielsen models</b>	<b>7</b>
2.1	Multiplet content of string FN models	7
2.2	Anomaly cancellation	9
2.3	SUGRA Lagrangian of string FN models	10
2.4	Masses and mixing	11
<b>3</b>	<b>The landscape of string FN models</b>	<b>12</b>
3.1	A first estimate of the size	12
3.2	Naturalness conditions	14
3.3	The search strategy	16
<b>4</b>	<b>Genetic algorithms and the string FN environment</b>	<b>17</b>
4.1	Genetic algorithms: a brief review	17
4.2	The string FN environment	18
4.3	The fitness function	19
4.4	Implementation of the search strategy	21
<b>5</b>	<b>Models containing only perturbative singlets</b>	<b>22</b>
5.1	Example model with $\mathbf{n} = (1, 1, 1, 1, 1)$	23
5.2	Example model with $\mathbf{n} = (1, 1, 3)$	26
<b>6</b>	<b>Models with non-perturbative singlets</b>	<b>28</b>
6.1	Models with $\mathbf{n} = (1, 1, 1, 1, 1)$	28
6.2	Models with $\mathbf{n} = (1, 2, 2)$	31
<b>7</b>	<b>Conclusion</b>	<b>34</b>
<b>A</b>	<b>Naturalness conditions and charge-fixing constraints</b>	<b>36</b>
A.1	Effects of order-one coefficients	36
A.2	Off-diagonal entries in the up-Yukawa matrix	37
<b>B</b>	<b>Leading-order operators</b>	<b>38</b>
<b>C</b>	<b>The fitness function</b>	<b>40</b>
<b>D</b>	<b>Crude renormalisation scheme</b>	<b>43</b>

---

## 1 Introduction

Over the past four decades, the development of string theory models of particle physics has gradually advanced, leading to the successful construction of models with the correct (supersymmetric) particle content of the Standard Model. Various approaches have contributed to this progress, and many such models are now known. For comprehensive reviews and detailed discussions, see [1–3] and the references therein. The next crucial step toward achieving a fully realistic string-derived Standard Model is to obtain the correct masses and mixing of quarks and leptons. Recently, significant progress has been made in this direction within top-down string constructions [4], specifically in the context of  $E_8 \times E_8$  heterotic string compactifications on smooth Calabi-Yau threefolds with abelian internal gauge fluxes.<sup>1</sup> The calculations presented in ref. [4] relied extensively on neural network techniques and represented a substantial undertaking in numerical differential geometry. A limited exploration of the complex structure moduli space was also included.

In this paper, we study fermion masses and mixing from a complementary, bottom-up perspective in the context of string-inspired four-dimensional theories with  $\mathcal{N} = 1$  supersymmetry. In string models, matter fields are tied to branes, singularities, and internal gauge flux within the underlying manifold. Mathematically, gauge flux is represented by vector bundles over the internal manifold or its submanifolds. Often, at certain loci in their moduli space, these vector bundles can ‘split’, meaning they exhibit a product structure group. At these split loci, Green-Schwarz anomalous U(1) symmetries emerge, which manifest in the low-energy effective field theory as global U(1) symmetries. Moving away from the split loci corresponds to switching on bundle moduli VEVs; in this process, the effects of the U(1) symmetries persist in the moduli space, constraining the allowed operators in the effective field theory [7]. Quarks and leptons, along with various geometric moduli (bundle moduli and Kähler moduli), acquire charges under these symmetries, imposing constraints on the allowed operators in the effective theory. In this paper, we analyse the implications of these symmetries for fermion masses and mixing. The specific form of the four-dimensional supergravity theories considered here is derived from heterotic Calabi-Yau compactifications with split vector bundles [8–13]. The calculation of physical Yukawa couplings in ref. [4] has been performed in the same setting and one of the reasons behind the present approach is to guide such calculations towards phenomenologically viable regions of the moduli space, including the bundle moduli space. Furthermore, it is known that related effective theories arise in F-theory compactifications [14–16], where similar methods can also be applied.

From a four-dimensional model-building perspective, our approach is similar to the Froggatt-Nielsen (FN) mechanism for the origin of the mass hierarchy [17–25]. However, it incorporates several important differences stemming from the presumed stringy origin of our models. For standard FN models of the simplest type, one assumes the presence of a U(1) flavour symmetry in addition to the Standard Model (SM) gauge group, typically referred to as horizontal U(1) symmetry. Both the SM fields and an additional SM singlet scalar field (flavon or FN scalar)  $\phi$  carry charge under the U(1) flavour symmetry. The presence of U(1)-invariant operators of the form  $\phi^p \times (\text{Yukawa interaction term})$ , where  $p$  is

---

<sup>1</sup>See also refs. [5, 6] for similar calculations in the context of standard embedding models.

a non-negative integer, implies that, once  $\phi$  develops a vacuum expectation value (VEV)  $\langle\phi\rangle$ , a Yukawa coupling constant of order  $\langle\phi\rangle^p$  is generated. The hope is then that the hierarchy of fermion masses can be explained by a somewhat small value of  $\langle\phi\rangle$  (measured in units of an appropriate mass scale) and different integers  $p$  for the various entries of the Yukawa matrices, as dictated by the U(1) charges.

Our string-inspired FN models (or simply “string FN models”) deviate from the standard FN mechanism in several ways, being more general in some aspects, whilst more specific in others. The string FN models considered in this paper are characterised by:

**Presence of supersymmetry.** We assume that our models are MSSM-like with Standard Model gauge group. In addition, the structure of the underlying string models implies that the horizontal U(1) charges of the MSSM multiplets can be organised in terms of SU(5) multiplets. The form of the Yukawa couplings is dictated by supersymmetry, the SM gauge symmetry and the horizontal symmetry.

**Underlying GUT symmetry.** While our models have a standard model gauge group and are by no means field theory GUTs there is an underlying SU(5) GUT symmetry at the string level which leaves certain traces in the low-energy theory. In particular, the FN U(1) symmetries commute with the SU(5) GUT symmetry so that U(1) charges can be organised by GUT multiplets.

**Several horizontal U(1)s.** We consider symmetry groups of the form  $\mathcal{G} \cong \text{U}(1)^{f-1}$ , where  $f = 2, 3, 4$  or  $5$ . It is convenient to parametrize  $\mathcal{G}$  by starting with a partition  $\mathbf{n} = (n_1, \dots, n_f)$  of 5 which, in the underlying string model, describes the splitting type of rank 5 vector bundles. In terms of this partition, the group  $\mathcal{G}$  is given by

$$\mathcal{G} = \left\{ \left( e^{i\epsilon^1}, \dots, e^{i\epsilon^f} \right) \mid \epsilon^a \in \mathbb{R}, \mathbf{n} \cdot \boldsymbol{\epsilon} = 0 \right\} \cong \text{U}(1)^{f-1}, \tag{1.1}$$

where  $\boldsymbol{\epsilon} = (\epsilon^1, \dots, \epsilon^f)$  are the group parameters. Representations of  $\mathcal{G}$  are labelled by integer vectors  $\mathbf{q} = (q_1, \dots, q_f) \in \mathbb{Z}^f$ , subject to the identification

$$\mathbf{q} \sim \mathbf{q}' \iff \mathbf{q} - \mathbf{q}' \in \mathbb{Z}\mathbf{n}. \tag{1.2}$$

**Several FN scalars.** Instead of a single FN scalar, we introduce multiple  $\mathcal{G}$ -charged scalars (chiral superfields) which come in two varieties. We denote the first type of FN scalars (chiral SM singlets) as  $\phi^{(\alpha)}$ , where  $\alpha$  is now an index, not a power. In a string theory context, these are interpreted as bundle moduli and contribute to the superpotential and the Kähler potential at perturbative level. Denoting by  $\mathbf{q}(\phi^{(\alpha)})$  the corresponding  $\mathcal{G}$ -charge vector, the singlet  $\phi^{(\alpha)}$  transforms linearly as

$$\phi^{(\alpha)} \mapsto e^{-i\mathbf{q}(\phi^{(\alpha)}) \cdot \boldsymbol{\epsilon}} \phi^{(\alpha)}, \tag{1.3}$$

where the vector  $\boldsymbol{\epsilon}$  contains the U(1) group parameters. Additionally, we introduce FN scalars  $\Phi^{(i)} = e^{-T^i}$  which feature in non-perturbative string contributions to the

superpotential and the Kähler potential.<sup>2</sup> These are associated with geometrical Kähler moduli fields  $T^i$  which transform non-linearly under  $\mathcal{G}$ :

$$T^i \mapsto T^i + i\mathbf{k}^i \cdot \epsilon, \tag{1.4}$$

where  $\mathbf{k}^i$  are integral vectors dictated, in the string context, by the topological properties of the split bundle. This non-linear transformation for the Kähler moduli  $T^i$  implies a linear transformation for the FN scalars  $\Phi^{(i)}$ . Thus, both types of FN scalars are  $\mathcal{G}$ -charged and their presence as singlet insertions is constrained.<sup>3</sup>

So far, the discussion has been fairly general. However, our models are special due to the particular structure of the allowed  $\mathcal{G}$  charges dictated by the presumed underlying string setting.

**Pattern of charges.** The SM fields and the perturbative FN scalars  $\phi^{(\alpha)}$  can carry charges  $\pm 1$  under at most two U(1) symmetries (and are uncharged under the others), according to a certain pattern suggested by the underlying string models and described in detail in the following section.

**Anomaly cancellation.** The charges for the non-perturbative FN scalars  $\Phi^{(i)}$  are constrained by anomaly cancellation, which imposes a restriction on the allowed integral vectors  $\mathbf{k}^i$  in the transformation (1.4).

**Yukawa unification.** In contrast to traditional field theory GUTs, the string GUT symmetry underlying our models do not automatically impose Yukawa unification [26] since the SM multiplets, which are unified in traditional field theory GUTs, originate from different multiplets in the underlying string theory GUT. However, as mentioned above, these multiplets necessarily share the same U(1) charges. This means our models have a crude form of Yukawa unification where allowed singlet insertions between the d-quark and the charged lepton sector are the same but the coefficients multiplying these insertions do not respect unification and will generically break it.

**Example model.** Before outlining our strategy for identifying such string FN models, it is useful to present an example. The model is characterised by a horizontal symmetry  $\mathcal{G} \cong \text{U}(1)^4$ , corresponding to the uniform partition  $\mathbf{n} = (1, 1, 1, 1, 1)$  of 5. As mentioned above, it is convenient to describe  $\mathcal{G}$  in terms of 5 group parameters with one relation, such that the charge vectors  $\mathbf{q} = (q_1, \dots, q_5) \in \mathbb{Z}^5$  are subject to the identification

$$\mathbf{q} \sim \mathbf{q}' \iff \mathbf{q} - \mathbf{q}' \in \mathbb{Z}(1, 1, 1, 1, 1). \tag{1.5}$$

---

<sup>2</sup>These contributions are associated with worldsheet instantons and arise as effects of the string worldsheet wrapping holomorphic curves inside the Calabi-Yau threefold. The instantons contribute exponentially suppressed terms, where the exponent is proportional to the volume of the wrapped curves, as measured by the Kähler moduli.

<sup>3</sup>One could be tempted to think of the U(1)-charges in this context by analogy with the modular weights that arise in orbifold compactifications, however, this is not helpful. The crucial difference is that in orbifold compactifications the modular weights of low-energy fields (including the Kähler moduli) are related to the orbifold geometry and its discrete symmetries. The would-be analogue in smooth Calabi-Yau compactifications corresponds to large diffeomorphisms, e.g. in the form of discrete symmetries. By contrast, the U(1)-symmetries used here arise from gauge transformations in the presence of line bundles.

The spectrum contains the Standard Model matter fields, written in SU(5) GUT notation:

$$\mathbf{10}_1 \quad \mathbf{10}_2 \quad \mathbf{10}_5 \quad \bar{\mathbf{5}}_{1,2} \quad \bar{\mathbf{5}}_{1,2} \quad \bar{\mathbf{5}}_{1,2} \quad H_{4,5}^u \quad H_{4,5}^d, \quad (1.6)$$

as well as five types of perturbative FN scalars

$$\phi_{5,1} \quad \phi_{3,5} \quad \phi_{4,5} \quad \phi_{1,2} \quad \phi_{4,1}, \quad (1.7)$$

and no non-perturbative FN scalars. The subscripts indicate the U(1) charges as follows:

$$\mathbf{q}(\mathbf{10}_a) = \mathbf{e}_a, \quad \mathbf{q}(\bar{\mathbf{5}}_{a,b}) = -\mathbf{q}(H_{a,b}^u) = \mathbf{q}(H_{a,b}^d) = \mathbf{e}_a + \mathbf{e}_b, \quad \mathbf{q}(\phi_{a,b}) = \mathbf{e}_a - \mathbf{e}_b, \quad (1.8)$$

where  $\mathbf{e}_a$  is the  $a$ -th standard basis vector in 5 dimensions. Our task now is to determine the structure of the Yukawa matrices as dictated by the U(1) symmetries.

The generic form of the Yukawa couplings is the following:

$$\begin{aligned} \text{up sector:} & \quad (\text{singlet insertions}) \times H_{-\mathbf{e}_a - \mathbf{e}_b}^u \mathbf{10}_{\mathbf{e}_c}^i \mathbf{10}_{\mathbf{e}_d}^j \\ \text{down sector:} & \quad (\text{singlet insertions}) \times H_{\mathbf{e}_a + \mathbf{e}_b}^d \bar{\mathbf{5}}_{\mathbf{e}_c + \mathbf{e}_d}^i \mathbf{10}_{\mathbf{e}_e}^j, \end{aligned} \quad (1.9)$$

The leading-order singlet insertions that lead to  $\mathcal{G}$ -invariant operators are:

$$\begin{aligned} \text{up sector:} & \quad \begin{pmatrix} \phi_{5,1}\phi_{4,1} & \phi_{5,1}\phi_{1,2}\phi_{4,1} & \phi_{4,1} \\ \phi_{5,1}\phi_{1,2}\phi_{4,1} & \phi_{5,1}\phi_{1,2}^2\phi_{4,1} & \phi_{1,2}\phi_{4,1} \\ \phi_{4,1} & \phi_{1,2}\phi_{4,1} & \phi_{4,5} \end{pmatrix} \\ \text{down sector:} & \quad \begin{pmatrix} \phi_{5,1}\phi_{3,5} & \phi_{5,1}\phi_{3,5} & \phi_{5,1}\phi_{3,5} \\ \phi_{5,1}\phi_{3,5}\phi_{1,2} & \phi_{5,1}\phi_{3,5}\phi_{1,2} & \phi_{5,1}\phi_{3,5}\phi_{1,2} \\ \phi_{3,5} & \phi_{3,5} & \phi_{3,5} \end{pmatrix} \end{aligned} \quad (1.10)$$

with the convention that the ordering of the three families follows the sequence in which they are listed in (1.6).

As noted above, Yukawa unification is not automatically enforced in string-derived GUT models, which implies that while the structure of singlet insertions in the Yukawa couplings can be described using SU(5) GUT language, the pre-factors, corresponding to order-one coefficients, cannot. In the underlying string context these coefficients depend on the geometrical moduli (complex structure and Kähler moduli) of the compactification. They also take into account the effects of a non-trivial Kähler potential for the matter fields, as well as contributions from the RG running.

Before optimising the order-one coefficients, a crucial step that defines the FN paradigm is to assign generic VEVs to the FN scalars such that the observed mass hierarchies and mixing parameters are reproduced qualitatively. For the model at hand, we claim that the following VEV assignments represent a viable choice:

$$\langle \phi_{5,1} \rangle \sim \epsilon^3, \quad \langle \phi_{3,5} \rangle \sim \epsilon^4, \quad \langle \phi_{4,5} \rangle \sim \epsilon, \quad \langle \phi_{1,2} \rangle \sim \epsilon^5, \quad \langle \phi_{4,1} \rangle \sim \epsilon^4,$$

with  $\epsilon \approx 0.4$ . This choice of VEVs gives the following Yukawa textures:

$$\text{up sector} \sim \begin{pmatrix} \epsilon^7 & \epsilon^{12} & \epsilon^4 \\ \epsilon^{12} & \epsilon^{17} & \epsilon^9 \\ \epsilon^4 & \epsilon^9 & \epsilon \end{pmatrix}, \quad \text{down sector} \sim \begin{pmatrix} \epsilon^7 & \epsilon^7 & \epsilon^7 \\ \epsilon^{12} & \epsilon^{12} & \epsilon^{12} \\ \epsilon^4 & \epsilon^4 & \epsilon^4 \end{pmatrix} \quad (1.11)$$

leading (once order-one coefficients are included) to the following patterns for the up-quark, down-quark and charged lepton masses:

$$(m_u, m_c, m_t) \sim (\epsilon^{17}, \epsilon^7, \epsilon), \quad (m_d, m_s, m_b), \quad (m_e, m_\mu, m_\tau) \sim (\epsilon^{12}, \epsilon^7, \epsilon^4). \quad (1.12)$$

Our choice of VEVs as powers of the same parameter  $\epsilon$  is inspired by the approach taken in, for example, in refs. [18, 19]. Since  $\epsilon$  is sufficiently close to 1, this choice does not introduce a large hierarchy between the various FN scalar VEVs.

To compare the model against the experimental values for the quark masses, charged lepton masses and CKM mixing parameters, we then optimise the order-one coefficients and the scale  $\epsilon$ . A modest random search over this  $(3 \times 9 + 1)$ -dimensional parameter space leads to values that are correct within percentage level.

We now present our strategy for identifying viable string-inspired FN models. The first step is to specify the horizontal symmetry group  $\mathcal{G}$ , for which there are six possible choices. For each of these, we systematically explore all inequivalent charge assignments for the MSSM multiplets. The number of such assignments is small in each case. Next, we specify the allowed number of FN scalar fields, both perturbative (up to 5) and non-perturbative (up to 4), that can appear in the spectrum.

With these settings fixed, we proceed to search for viable  $G$ -charge assignments for the FN scalars and corresponding VEV patterns, expressed as integer powers of a single parameter  $\epsilon \lesssim 1$ . The range of powers of  $\epsilon$  is constrained so that the VEVs of the perturbative singlets satisfy  $\langle \phi \rangle \lesssim 1$  (in Planck units) while those of the non-perturbative singlets satisfy  $\langle \Phi \rangle \ll 1$ . We construct the leading  $\mathcal{G}$ -invariant terms contributing to the Yukawa couplings (using the method presented in appendix B) and check if the resulting Yukawa texture leads to promising values for the quark masses and CKM parameters. At this stage we do not check the masses in the charged lepton sector, as general arguments indicate that this can be achieved with order-one coefficients, provided that the down-quark Yukawa sector is taken care of.

Finally, we optimize the order-one coefficients in the quark and charged lepton sectors, along with the parameter  $\epsilon$ . The goal is to ensure that all measurable quantities align with their experimental values to within a few percent accuracy. The order-one coefficients can, in principle, be calculated within string models, as demonstrated in the analysis of ref. [4], which shows that these coefficients can vary by at least an order of magnitude depending on the choice of complex structure. As such, they can contribute to the observed mass and mixing hierarchies. However, in this paper, we adopt the approach of generating the hierarchies directly through powers of FN scalars, excluding cases where hierarchies arise purely from tuning the coefficients. A detailed analysis shows that certain naturalness conditions must be imposed, which in particular constrain the allowed charge assignments for the Standard Model multiplets, as discussed in section 3.

Even after taking into account such naturalness considerations, the landscape of string FN models is far too large to be searched systematically for successful models, particularly when multiple U(1) symmetries and many scalars are present. For this reason, we resort to heuristic search methods, in particular genetic algorithms (GAs). Another approach would be to use reinforcement learning which proved successful in the search of standard FN models in ref. [27]).

So far, we have made no mention of the neutrino sector. The chiral multiplets associated with the FN scalars include fermions that could, in principle, serve as right-handed neutrinos. However, the generation of appropriate neutrino masses (such as through the see-saw mechanism) is more complicated and tied to the problem of moduli stabilization, which complicates the overall dynamics. A detailed treatment of these issues is beyond the scope of this paper and will be explored in a future publication. We emphasize, however, that finding successful FN models that can reproduce the quark and charged lepton sectors has important implications for top-down string model building. The successful  $\mathcal{G}$  charge patterns identified in this way can be translated into, likely strong, topological constraints on string compactifications.

It is important to note that the FN mechanism explored here, based on U(1) symmetries from string compactifications, is not the only approach for generating mass and mixing hierarchies in string models. Discrete symmetries can also play a role, both in smooth Calabi-Yau settings and, more extensively, within the framework of modular flavour symmetries (for reviews, see [28–31] and references therein). These modular symmetries can also have a string-theoretic origin [32–34].

The rest of the paper is structured as follows. In section 2 we present the class of four-dimensional  $\mathcal{N} = 1$  supergravity theories that underlie the string FN models studied here. We discuss the flavour symmetry  $\mathcal{G}$  and its phenomenological consequences and summarise our strategy for exploring the string FN landscape. After a short review of GAs in section 4 and a discussion of the implementation used for the present purpose, we present our results in sections 5 and 6. We conclude in section 7.

## 2 String Froggatt-Nielsen models

In the first part of this section, we describe the general structure of string FN models, within the framework of four-dimensional  $\mathcal{N} = 1$  supergravity. We also briefly review the relevant phenomenological constraints and discuss a number of general features of string FN models which will be helpful in the search.

### 2.1 Multiplet content of string FN models

Our models have a symmetry  $G_{\text{SM}} \times \mathcal{G}$ , where  $G_{\text{SM}} = \text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$  is the Standard Model gauge group and  $\mathcal{G}$  contains the additional Green-Schwarz anomalous U(1) factors. Its precise structure is determined by a (non-trivial) partition of five described by an integer vector  $\mathbf{n} = (n_1, \dots, n_f)$ , where  $f = 2, 3, 4$  or  $5$ , with all  $n_a \geq 1$  and  $\sum_{a=1}^f n_a = 5$ . In terms of this partition, the group  $\mathcal{G}$  is given by

$$\mathcal{G} = \left\{ \left( e^{i\epsilon^1}, \dots, e^{i\epsilon^f} \right) \mid \epsilon^a \in \mathbb{R}, \mathbf{n} \cdot \boldsymbol{\epsilon} = 0 \right\} \cong \text{U}(1)^{f-1}, \quad (2.1)$$

field	SM rep	name	SU(5)	$\mathcal{G}$ charge pattern	SU(5) $\times\mathcal{G}$
$Q$	$(\mathbf{3}, \mathbf{2})_1$	LH quark	$\mathbf{10}$	$\mathbf{e}_a$	$\mathbf{10}_a$
$u$	$(\bar{\mathbf{3}}, \mathbf{1})_{-4}$	RH $u$ -quark			
$e$	$(\mathbf{1}, \mathbf{1})_6$	RH electron			
$d$	$(\bar{\mathbf{3}}, \mathbf{1})_2$	RH $d$ -quark	$\bar{\mathbf{5}}$	$\mathbf{e}_a + \mathbf{e}_b$	$\bar{\mathbf{5}}_{a,b}$
$L$	$(\mathbf{1}, \mathbf{2})_{-3}$	LH lepton			
$H^d$	$(\mathbf{1}, \mathbf{2})_{-3}$	down-Higgs	$\bar{\mathbf{5}}^{H^d}$	$\mathbf{e}_a + \mathbf{e}_b$	$\bar{\mathbf{5}}_{a,b}^{H^d}$
$H^u$	$(\mathbf{1}, \mathbf{2})_3$	up-Higgs	$\mathbf{5}^{H^u}$	$-\mathbf{e}_a - \mathbf{e}_b$	$\mathbf{5}_{a,b}^{H^u}$
$\phi$	$(\mathbf{1}, \mathbf{1})_0$	pert. FN scalar	$\mathbf{1}$	$\mathbf{e}_a - \mathbf{e}_b$	$\mathbf{1}_{a,b}$
$\Phi$	$(\mathbf{1}, \mathbf{1})_0$	non-pert. FN scalar	$\mathbf{1}$	$\mathbf{k} = (k_1, \dots, k_f)$	$\mathbf{1}$

**Table 1.** The field content and multiplet charges of string FN models. In addition to the  $G_{\text{SM}}$  representations, we have also included the associated SU(5) GUT representations for the usual embedding  $G_{\text{SM}} \subset \text{SU}(5)$ . The fifth column provides the  $\mathcal{G}$  charge pattern, where  $\{\mathbf{e}_a\}$  are the  $f$ -dimensional standard basis vectors. The indices  $a$  and  $b$  indicate under which U(1) symmetries multiplets carry charge and depend on the family. For multiplets with charge  $\pm(\mathbf{e}_a + \mathbf{e}_b)$ , the choice  $a = b$  is permitted only if the partition vector  $\mathbf{n}$  satisfies  $n_a > 1$ . The  $\mathcal{G}$  charges for the non-perturbative FN scalars  $\Phi^{(i)}$  are integer vectors  $\mathbf{k}^i = (k^i_a) \in \mathbb{Z}^f$  satisfying  $\sum_{a=1}^f k^i_a = 0$ .

where  $\epsilon = (\epsilon^1, \dots, \epsilon^f)$  are the group parameters and  $\mathcal{G}$  is isomorphic to a product of  $f - 1$  U(1) factors. These are gauge symmetries but the associated gauge bosons are super-massive and at low energies they appear, effectively, as global symmetries. It is often convenient to describe  $\mathcal{G}$  by using  $f$  U(1) symmetries, subject to the constraint in eq. (2.1), and we will denote these by  $U_a(1)$ , where  $a = 1, \dots, f$ . From this viewpoint, representations of  $\mathcal{G}$  can be labelled by integer vectors  $\mathbf{q} = (q_1, \dots, q_f) \in \mathbb{Z}^f$ , subject to identifying two such vectors  $\mathbf{q}$  and  $\mathbf{q}'$  according to the relation

$$\mathbf{q} \sim \mathbf{q}' \iff \mathbf{q} - \mathbf{q}' \in \mathbb{Z}\mathbf{n} . \tag{2.2}$$

Hence, the  $\mathcal{G}$  charge lattice is given by  $\mathbb{Z}^f / \mathbb{Z}\mathbf{n}$ . In particular, this means that the  $\mathcal{G}$ -invariant operators are precisely those with charge  $\mathbf{q} \in \mathbb{Z}\mathbf{n}$ .

The multiplet content of the  $\mathcal{N} = 1$  supergravity theory for string FN models consists of gauge multiplets with gauge group  $G_{\text{SM}}$  (while the  $\mathcal{G}$  vector multiplets are super-massive and have been integrated out) and chiral multiplets. The latter include the standard MSSM multiplets  $Q^I, u^I, d^I, L^I, e^I$ , where  $I = 1, 2, 3$  is a family index, the two Higgs multiplets  $H^u, H^d$ , as well as a number  $N_\phi + N_\Phi$  of chiral multiplets containing the perturbative FN scalars (bundle moduli)  $\phi^{(\alpha)}$ , where  $\alpha = 1, \dots, N_\phi$ , and the non-perturbative FN scalars (associated with Kähler moduli)  $\Phi^{(i)} = e^{-T^i}$ , where  $i = 1, \dots, N_\Phi$ .

The list of fields and their representations are given explicitly in table 1, where we have not only provided the  $G_{\text{SM}}$  representations but also the SU(5) GUT representations for the usual embedding  $G_{\text{SM}} \subset \text{SU}(5)$ . The appearance of the SU(5) multiplet structure is related

to the structure of the underlying string models. The group  $\mathcal{G}$  can be viewed as a collection of  $U(1)$  flavour symmetries with a specific pattern of charges indicated in the fifth column in table 1, which depends only on the  $SU(5)$  representation. For this reason, it will be convenient to use  $SU(5)$  GUT notation for the fields and operators. However, this does not mean that we consider GUTs — our models are MSSM-like with the Standard Model gauge group.<sup>4</sup> While the charges and, hence, the singlet insertions follow an  $SU(5)$  GUT pattern this is not, in general the case for the coefficients in front of these insertions. This means we have a crude form of Yukawa unification at the level of allowed singlet insertions which is generically broken by the coefficients multiplying these insertions [35].

Any of the Standard Model multiplets and the perturbative FN scalars from table 1, denoted generically by  $C$  and having  $\mathcal{G}$  charge  $\mathbf{q}(C)$ , transform linearly as

$$C \mapsto e^{-i\mathbf{q}(C)\cdot\epsilon} C. \tag{2.3}$$

On the other hand, the Kähler moduli  $T^i$  with  $\mathcal{G}$  charge  $\mathbf{k}^i \in \mathbb{Z}^f$  transform non-linearly as:

$$T^i \mapsto T^i + i\mathbf{k}^i \cdot \epsilon \quad \Leftrightarrow \quad \Phi^{(i)} = e^{-T^i} \mapsto e^{-i\mathbf{k}^i \cdot \epsilon} \Phi^{(i)}, \tag{2.4}$$

where the charge vectors  $\mathbf{k}^i = (k^i_a)$  are subject to the constraint

$$\sum_{a=1}^f k^i_a = 0, \quad \forall i \in \{1, \dots, N_\Phi\}. \tag{2.5}$$

(In a string context, this condition corresponds to the vanishing first Chern class of the underlying split vector bundle.)

## 2.2 Anomaly cancellation

As mentioned above, the  $U(1)$  symmetries in  $\mathcal{G}$  are typically anomalous. In particular, the mixed triangle anomalies of the form  $U_a(1) \times SU(3)^2$ ,  $U_a(1) \times SU(2)^2$ ,  $U_a(1) \times U(1)^2$  which involve one  $U_a(1)$  and two Standard Model gauge bosons, are non-vanishing. For the three factors in  $G_{\text{SM}}$  these anomalies are proportional to

$$\begin{aligned} \mathbf{A}^{(3)} &= \mathbf{A}^{(5)} \\ \mathbf{A}^{(2)} &= \mathbf{A}^{(5)} + \mathbf{q}(H^u) + \mathbf{q}(H^d) \\ \mathbf{A}^{(1)} &= \mathbf{A}^{(5)} + \frac{3}{5} \left( \mathbf{q}(H^u) + \mathbf{q}(H^d) \right) \end{aligned} \tag{2.6}$$

where  $\mathbf{A}^{(5)}$  is the mixed GUT  $U_a(1)SU(5)^2$  anomaly,

$$\mathbf{A}^{(5)} = \sum_{I=1}^3 \left( \mathbf{q}(\bar{\mathbf{5}}^I) + 3\mathbf{q}(\mathbf{10}^I) \right), \tag{2.7}$$

evaluated only on the families  $\bar{\mathbf{5}}^I$  and  $\mathbf{10}^I$  (but not on the Higgs multiplets which are incomplete from an  $SU(5)$  GUT perspective).

---

<sup>4</sup>These models are based on rank five vector bundles and, initially, lead to  $SU(5)$  GUT models. However, the  $SU(5)$  group is broken to the Standard Model group by a Wilson-line bundle and the GUT symmetry is never realised at the level of four-dimensional field theories.

These triangle anomalies have to be cancelled by the Green-Schwarz mechanism, facilitated by a non-trivial  $\mathcal{G}$  transformation of the gauge kinetic function for the  $G_{\text{SM}}$  gauge bosons. In principle, this gauge kinetic function can be different for the three factors in  $G_{\text{SM}}$ . However, for models derived from heterotic compactifications, the gauge kinetic function is universal and we adopt this feature for our string FN models. From eq. (2.6), this means we require, in a first instance, that the two Higgs doublets have opposite  $\mathcal{G}$ -charge, that is,  $\mathbf{q}(H^u) + \mathbf{q}(H^d) \sim \mathbf{0}$ , so that all three anomaly coefficients are the same. The universal gauge kinetic function for  $G_{\text{SM}}$  is given by

$$f = S + \beta_i T^i, \tag{2.8}$$

where  $\beta_i \in \mathbb{Z}$  are integers and  $S$  is the dilaton multiplet with  $\mathcal{G}$  transformation

$$S \mapsto S + i\beta_i \epsilon \cdot \mathbf{k}^i. \tag{2.9}$$

For Green-Schwarz anomaly cancellation to work, we require that

$$\mathbf{A}^{(5)} \sim \beta_i \mathbf{k}^i, \tag{2.10}$$

and this condition will be imposed on our models.

### 2.3 SUGRA Lagrangian of string FN models

The  $R$ -parity conserving<sup>5</sup> Kähler potential and superpotential for our models are given by

$$K = \hat{k}^u H^u \bar{H}^u + \hat{k}^d H^d \bar{H}^d + \hat{K}_{IJ}^Q Q^I \bar{Q}^J + \hat{K}_{IJ}^u u^I \bar{u}^J + \hat{K}_{IJ}^d d^I \bar{d}^J + \hat{K}_{IJ}^L L^I \bar{L}^J + \hat{K}_{IJ}^e e^I \bar{e}^J + \dots \tag{2.11}$$

$$W = \hat{\Lambda}_{IJ}^u H^u Q^I u^J + \hat{\Lambda}_{IJ}^d H^d Q^I d^J + \hat{\Lambda}_{IJ}^e H^d L^I e^J + \dots, \tag{2.12}$$

where the dots stand for terms that involve only the chiral multiplets in which the FN scalars live. The Kähler potential for the gravitational moduli (which exhibits a no-scale structure at tree level) has been omitted as it only depends on moduli quantities uncharged under  $\mathcal{G}$ , and hence places no further constraints relevant to our discussion. All hatted quantities are coupling functions which contain the possible singlet insertions in terms of  $\phi^{(\alpha)}$  and  $\Phi^{(i)}$  which lead to  $\mathcal{G}$ -invariant operators. More concretely, the Yukawa coupling functions are schematically given by

$$\hat{\Lambda}_{IJ}^d = \sum_r a_{IJ,r}^d \mathcal{O}_{IJ,r}^d, \quad \hat{\Lambda}_{IJ}^u = \sum_r a_{IJ,r}^u \mathcal{O}_{IJ,r}^u, \quad \hat{\Lambda}_{IJ}^e = \sum_r a_{IJ,r}^e \mathcal{O}_{IJ,r}^e, \tag{2.13}$$

where  $a_{IJ,r}^d$  and  $a_{IJ,r}^u$  are order-one coefficients which are not fixed by any of the symmetries under consideration. The terms  $\mathcal{O}_{IJ,r}^d$ ,  $\mathcal{O}_{IJ,r}^u$ ,  $\mathcal{O}_{IJ,r}^e$  are  $\phi^{(\alpha)}$  and  $\Phi^{(i)}$  singlet insertions (counted by the index  $r$ ) which lead to  $\mathcal{G}$ -invariant operators, that is, they satisfy

$$\mathbf{q}(\mathcal{O}_{IJ,r}^u) \sim -\mathbf{q}(H^u Q^I u^J), \quad \mathbf{q}(\mathcal{O}_{IJ,r}^d) \sim -\mathbf{q}(H^d Q^I d^J), \tag{2.14}$$

$$\mathbf{q}(\mathcal{O}_{IJ,r}^e) \sim -\mathbf{q}(H^d L^I e^J).$$

---

<sup>5</sup>Here we assume the existence of a symmetry, such as the standard R-Parity or a matter parity, which originates from the presumed underlying string model and forbids R-parity violating terms.

The matter field metrics in the Kähler potential are constructed in a similar way, for example

$$K_{IJ}^Q = b_I^Q \delta_{IJ} + \sum_r b_{IJ,r}^Q \mathcal{O}_{IJ,r}^Q \tag{2.15}$$

Here,  $b_I^Q$  and  $b_{IJ,r}^Q$  are order-one coefficients<sup>6</sup> and the singlet insertions  $\mathcal{O}_{IJ,r}^Q$  are monomials in the singlet fields (and their complex conjugates) with

$$\mathbf{q}(\mathcal{O}_{IJ,r}^Q) \sim -\mathbf{q}(Q^I \bar{Q}^J) . \tag{2.16}$$

Note that  $\mathbf{q}(Q^I \bar{Q}^I) \sim \mathbf{0}$  so that the diagonal kinetic terms are always allowed without any singlet insertion, as indicated by the first term on the r.h.s. of eq. (2.15). Similar expressions and statements apply to the other field space metrics in eq. (2.11).

In order to illustrate how to find the allowed singlet insertions, we consider an explicit example with split pattern  $\mathbf{n} = (1, 1, 1, 1, 1)$  and a multiplet content which includes  $\bar{\mathbf{5}}_{1,2}$ ,  $\mathbf{10}_2$ ,  $\mathbf{10}_5$ ,  $\bar{\mathbf{5}}_{4,5}^H$  and  $\mathbf{5}_{4,5}^H$  (using GUT language for simplicity). Then, the up-Yukawa term  $\mathbf{5}_{4,5}^H \mathbf{10}_5 \mathbf{10}_5$  carries  $\mathcal{G}$ -charge  $\mathbf{q}(\mathbf{5}_{4,5}^H \mathbf{10}_5 \mathbf{10}_5) = (0, 0, 0, -1, 1)$ . This means possible singlet insertions are, for example,  $\phi_{4,5}$  or  $\phi_{4,1} \phi_{1,5}$  (provided singlets with these charges appear in the theory). On the other hand, the down-Yukawa coupling  $\bar{\mathbf{5}}_{4,5}^H \bar{\mathbf{5}}_{1,2} \mathbf{10}_5$  has  $\mathcal{G}$ -charge  $\mathbf{q}(\bar{\mathbf{5}}_{4,5}^H \bar{\mathbf{5}}_{1,2} \mathbf{10}_5) = (1, 1, 0, 1, 2)$  so that a singlet insertion  $\phi_{3,5}$  is possible. Note, this leads to  $\mathbf{q}(\phi_{3,5} \bar{\mathbf{5}}_{4,5}^H \bar{\mathbf{5}}_{1,2} \mathbf{10}_5) = (1, 1, 1, 1, 1)$  and, from eq. (2.2), this charge is indeed equivalent to charge  $\mathbf{0}$ . In summary, we are following standard FN lore, but with multiple U(1) symmetries and scalar fields and specific charge patterns, as suggested by string theory.

## 2.4 Masses and mixing

After having written down the most general  $\mathcal{G}$ -invariant theory (up to a certain maximal operator dimension), as explained in the previous sub-section, we consider singlet VEVs  $\langle \phi^{(\alpha)} \rangle < 1$  and  $\langle \Phi^{(i)} \rangle \ll 1$ , thereby converting the various coupling functions into coupling constants

$$K^u = \hat{K}^u(\langle \phi^{(\alpha)} \rangle, \langle \Phi^{(i)} \rangle) , \quad \Lambda_{IJ}^u = \hat{\Lambda}_{IJ}^u(\langle \phi^{(\alpha)} \rangle, \langle \Phi^{(i)} \rangle) , \quad \dots \tag{2.17}$$

and similarly for the other couplings. To work out quark masses and mixing we should first canonically normalise kinetic terms, that is, we should introduce  $\text{GL}(3, \mathbb{C})$  matrices  $P^Q$ ,  $P^u$  and  $P^d$  such that

$$(P^Q)^\dagger K^Q P^Q = \mathbf{1}_3 , \quad (P^u)^\dagger K^u P^u = \mathbf{1}_3 , \quad (P^d)^\dagger K^d P^d = \mathbf{1}_3 . \tag{2.18}$$

Then, the ‘physical’ mass matrices  $M^u$  and  $M^d$ , relative to canonically normalised kinetic terms, are given by

$$M^u = \frac{v^u}{\sqrt{k^u}} (P^Q)^\dagger \Lambda^u P^u , \quad M^d = \frac{v^d}{\sqrt{k^d}} (P^Q)^\dagger \Lambda^d P^d , \tag{2.19}$$

---

<sup>6</sup>The overall scale for the matter field Kähler potential  $\sim (T + \bar{T})^{-1}$  can be accounted for by an appropriate redefinition of the order-one coefficients as well as the reference scale for the VEVs of the singlet fields. To simplify the discussion, we have absorbed the moduli-dependence into the order-one coefficients and definition of the reference scales. Ultimately, deciding whether the moduli-dependent coefficients  $b_I^Q$  and  $b_{IJ,r}^Q$  are indeed of order one in any given compactification, can be accomplished by performing numerical calculations of the matter field Kähler metric analogous to those carried out in ref. [4].

where  $v^u$  and  $v^d$  are the up- and down-Higgs VEVs, respectively, satisfying

$$\sqrt{(v^u)^2 + (v^d)^2} = v \simeq 174 \text{ GeV} , \quad (2.20)$$

with the electro-weak VEV  $v$ . The discussion can be somewhat simplified owing to the structure of the matter field metrics as, for example, given in eq. (2.15). The order-one terms along the diagonal dominate over the additional singlet insertions, so that the matrices  $P^Q$ ,  $P^u$  and  $P^d$  are approximately diagonal with order-one diagonal entries. Rotating with these matrices amounts to another order-one effect which we can think of as being absorbed by the order-one coefficients in  $\hat{\Lambda}^u$  and  $\hat{\Lambda}^d$ . Therefore, in practice, since we are only concerned about the leading-order terms in the singlet fields, we only need to construct the singlet insertions for the Yukawa functions  $\hat{\Lambda}_{IJ}^d$  and  $\hat{\Lambda}_{IJ}^u$  and can omit the sub-leading order effects from canonically normalising the kinetic terms. Similar considerations apply to the charged leptons and the Higgs fields.

It follows that we can simply write

$$M^u = v^u \Lambda^u , \quad M^d = v^d \Lambda^d , \quad M^e = v^d \Lambda^e . \quad (2.21)$$

As usual, these mass matrices need to be diagonalised to obtain the quark and charged lepton masses so we introduce unitary matrices  $U^u, V^u, U^d, V^d, U^e, V^e$  such that

$$\begin{aligned} M^u &= U^u \hat{M}^u (V^u)^\dagger & \hat{M}^u &= \text{diag}(m_u, m_c, m_t) \\ M^d &= U^d \hat{M}^d (V^d)^\dagger & \hat{M}^d &= \text{diag}(m_d, m_s, m_b) \\ M^e &= U^e \hat{M}^e (V^e)^\dagger & \hat{M}^e &= \text{diag}(m_e, m_\mu, m_\tau) . \end{aligned} \quad (2.22)$$

The Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix  $V_{\text{CKM}}$  is given by

$$V_{\text{CKM}} = (U^u)^\dagger U^d . \quad (2.23)$$

The above Yukawa couplings are those just below the compactification scale and, in principle, they have to be RG evolved down to low energies for a comparison with measured masses and mixing. It is well-known that this leads to order-one rescalings of the Yukawa couplings, and these will be taken into account using the approximate solution to the RG equations described in appendix D.

For later reference, the measured masses for the quarks and charged leptons are listed in table 2.

The CKM matrix, taken from ref. [36], is given by

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 0.97373 \pm 0.00031 & 0.2243 \pm 0.0008 & 0.00382 \pm 0.00020 \\ 0.221 \pm 0.004 & 0.975 \pm 0.006 & 0.0408 \pm 0.0014 \\ 0.0086 \pm 0.0002 & 0.0415 \pm 0.0009 & 1.014 \pm 0.029 \end{pmatrix} . \quad (2.24)$$

### 3 The landscape of string FN models

#### 3.1 A first estimate of the size

Given a horizontal symmetry  $\mathcal{G} \cong \text{U}(1)^{f-1}$ , associated with a partition  $\mathbf{n}$  of 5 with  $f$  parts, we estimate the number of charge assignments for the SM multiplets and the FN singlets.

<b>Quark</b>	$m_u$ (GeV)	$m_c$ (GeV)	$m_t$ (GeV)
<b>Mass</b>	$0.00216^{+0.00049}_{-0.00026}$	$1.27 \pm 0.02$	$172.4 \pm 0.07$
<b>Quark</b>	$m_d$ (GeV)	$m_s$ (GeV)	$m_b$ (GeV)
<b>Mass</b>	$0.00467^{+0.00048}_{-0.00017}$	$0.093^{+0.011}_{-0.005}$	$4.18^{+0.03}_{-0.02}$
<b>Lepton</b>	$m_e$ (GeV)	$m_\mu$ (GeV)	$m_\tau$ (GeV)
<b>Mass</b>	$0.000511 \pm 0.000001$	$0.1057 \pm 0.0001$	$1.77682 \pm 0.00016$

**Table 2.** Measured masses of quarks and charged leptons in GeV, as reported in ref. [36].

For the three  $\mathbf{10}_a$  multiplets, there are  $f^3$  choices. For each  $\bar{\mathbf{5}}_{a,b}$  or  $\mathbf{5}_{a,b}$  multiplet there are, roughly speaking,  $f^2/2$  choices (depending on the partition  $\mathbf{n}$ , charges with  $a = b$  may or may not be allowed), giving a total of  $f^8/16$  choices. Recall that the two Higgs multiplets come in a vector-like pair, so their  $\mathcal{G}$  charges are correlated. For each perturbative singlet  $\phi^{(\alpha)}$  there are  $f^2$  possibilities, giving a total of  $f^{2N_\phi}$  choices. For each non-perturbative singlet  $\Phi^{(i)}$ , taking into account the constraint (2.5), there are  $\Delta k^{f-1}$  possibilities, giving a total of  $\Delta k^{(f-1)N_\Phi}$  choices, where we assume a range of charges  $k^i_a \in \{k_{\min}, \dots, k_{\max}\}$  and  $\Delta k = k_{\max} - k_{\min} + 1$ .

Combining all this gives an approximate total number of charge choices of

$$N_{\text{charges}} \simeq \frac{f^{11+2N_\phi} \Delta k^{(f-1)N_\Phi}}{16} . \tag{3.1}$$

For instance, for  $f = 5$ ,  $\Delta k = 10$ ,  $N_\phi = 4$  and  $N_\Phi = 2$  this leads to approximately  $N_{\text{charges}} \simeq 10^{20}$  choices. However, not all of these configurations are physically distinct. The  $\mathbf{10}$  and  $\bar{\mathbf{5}}$  families can be permuted under an  $S_3 \times S_3$  symmetry, while the singlet fields can be permuted under an  $S_{N_\phi} \times S_{N_\Phi}$  symmetry. Additionally, there is a permutation symmetry for the  $U_a(1)$  factors, governed by the subgroup of  $S_f$  that stabilizes the vector  $\mathbf{n}$ . These symmetries reduce the number of inequivalent charge assignments by the order of the combined symmetry group. In the above example, this gives  $\simeq 10^{15}$  inequivalent charge assignments. However, for the heuristic search method described below it will not be practical to eliminate the permutation symmetry so we will be dealing effectively with the number in eq. (3.1).

The choice of singlet VEVs also needs to be included in the estimation of the total size of the string FN landscape. The reference scale for the singlet VEVs is the four-dimensional Planck scale which, in a string context, can be verified by a straightforward reduction calculation from ten dimensions.<sup>7</sup> The VEVs can, of course, take values in a continuous range but, in the spirit of FN model building, they will be discretised by writing

$$\langle \phi^{(\alpha)} \rangle = \epsilon^{p_\alpha} , \quad \langle \Phi^{(i)} \rangle = \epsilon^{\tilde{p}_i} , \tag{3.2}$$

where  $0 < \epsilon < 1$  is a model-dependent number which we refer to as the ‘ $\epsilon$ -scale’ and  $p_\alpha \in \{1, \dots, p_{\max}\}$ ,  $\tilde{p}_i \in \{\tilde{p}_{\min}, \dots, \tilde{p}_{\max}\}$  are non-negative integers. Since the VEVs for the non-perturbative singlets  $\Phi^{(i)}$  are exponentially suppressed, their typical values will be taken

<sup>7</sup>Together with an appropriate rescaling to account for overall factors in the Kähler potential and superpotential due to canonically-normalising the kinetic terms, as discussed in section 2.4.

to be smaller than the  $\phi^{(\alpha)}$  VEVs. This will be implemented by different ranges for the integers  $p_\alpha$  and  $\tilde{p}_i$ , setting  $\tilde{p}_{\min} > 1$ . The total number of models in the string FN landscape, including the number of VEV choices, is

$$N_{\text{models}} = N_{\text{charges}} \Delta p^{N_\phi} \Delta \tilde{p}^{N_\Phi} , \tag{3.3}$$

where  $\Delta p = p_{\max}$  and  $\Delta \tilde{p} = \tilde{p}_{\max} - \tilde{p}_{\min} + 1$ . For the example  $f = 5$ ,  $\Delta k = 10$ ,  $N_\phi = 4$  and  $N_\Phi = 2$  and choosing  $\Delta p = \Delta \tilde{p} = 10$  this leads to, roughly,  $N_{\text{models}} \simeq 10^{26}$  models. It's therefore clear that, at least for the cases with many singlet fields, the landscape of string FN models is large enough that heuristic search algorithms are a sensible strategy.

### 3.2 Naturalness conditions

Another aspect of FN model building is that family mass hierarchies should result from either charge choices or different VEVs (or a combination of both), but not from fine-tuning of the  $\mathcal{O}(1)$  coefficients in eq. (2.13). In order to formulate such a non-fine-tuning condition, consider a mass matrix  $M$  ( $M^u$  or  $M^d$  in our case) and define  $\mathcal{M} = MM^\dagger$ . From our parametrisation of VEVs in eq. (3.2) and controlled by the  $\mathcal{G}$ -symmetry and charge choices, the entries of  $\mathcal{M}$  are organised by powers of  $\epsilon$ , that is,

$$\mathcal{M}_{IJ} = c_{IJ} \epsilon^{\nu_{IJ}} , \tag{3.4}$$

where  $|c_{IJ}| = \mathcal{O}(1)$  with  $c_{JI} = c_{IJ}^*$  and  $\nu_{IJ} = \nu_{JI} \in \{0, 1, \dots\}$ . For illustration, consider a simple two-family example with matrix

$$\mathcal{M} = \begin{pmatrix} c_{11} \epsilon^{\nu_{11}} & c_{12} \epsilon^{\nu_{12}} \\ c_{12}^* \epsilon^{\nu_{12}} & c_{22} \epsilon^{\nu_{22}} \end{pmatrix} . \tag{3.5}$$

We want to require that the two eigenvalues of  $\mathcal{M}$  are proportional to different powers of  $\epsilon$ . It is easy to show, by considering  $\text{tr}(\mathcal{M})$  and  $\det(\mathcal{M})$ , that this is equivalent to the ‘naturalness condition’

$$\left. \begin{array}{l} \mathcal{M} \text{ eigenvalues with} \\ \text{different powers of } \epsilon \end{array} \right\} \Leftrightarrow \min \left( \frac{\nu_{11} + \nu_{22}}{2}, \nu_{12} \right) > \min(\nu_{11}, \nu_{22}) . \tag{3.6}$$

For example, if  $\nu_{11} = \nu_{22}$  this condition cannot be satisfied so, in this case, the eigenvalues of  $\mathcal{M}$  are given by the same power of  $\epsilon$ . Then, the only way to obtain eigenvalues with a numerical hierarchy is by tuning the coefficients  $c_{IJ}$  — and this is precisely what we would like to exclude. For this reason, we impose the condition (3.6) on the powers  $\nu_{IJ}$  or, rather, we impose its three-family analogue which can be easily derived by considering

$$\lambda_3 + \lambda_2 + \lambda_1 = \text{tr}(\mathcal{M}) = \mathcal{O}(\epsilon^{r_3}) \tag{3.7}$$

$$\lambda_3 \lambda_2 + \lambda_3 \lambda_1 + \lambda_2 \lambda_1 = \frac{1}{2} \left( \text{tr}(\mathcal{M})^2 - \text{tr}(\mathcal{M}^2) \right) = \mathcal{O}(\epsilon^{r_2}) \tag{3.8}$$

$$\lambda_1 \lambda_2 \lambda_3 = \det(\mathcal{M}) = \mathcal{O}(\epsilon^{r_1}) , \tag{3.9}$$

where  $\lambda_1 \ll \lambda_2 \ll \lambda_3$  are the eigenvalues of  $\mathcal{M}$  and the  $r_I$  specify the leading  $\epsilon$  powers of each of the three invariants, computed directly from the matrix (3.4). Clearly,  $\lambda_3 = \mathcal{O}(\epsilon^{r_3})$ ,

$\lambda_2 = \mathcal{O}(\epsilon^{r_2-r_3})$  and  $\lambda_1 = \mathcal{O}(\epsilon^{r_1-r_2-r_3})$  so, in order to have three eigenvalues with different orders in  $\epsilon$ , we require the naturalness conditions

$$2r_3 < r_2, \quad 2r_2 < r_1. \tag{3.10}$$

For example, consider the matrix

$$\mathcal{M} = \mathcal{O} \begin{pmatrix} \epsilon^5 & \epsilon^3 & \epsilon^2 \\ \epsilon^3 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon & 1 \end{pmatrix} \tag{3.11}$$

for which  $r_3 = 0$ ,  $r_2 = 2$  and  $r_1 = 6$ . Then  $\lambda_3 = \mathcal{O}(1)$ ,  $\lambda_2 = \mathcal{O}(\epsilon^2)$  and  $\lambda_1 = \mathcal{O}(\epsilon^4)$ , so we have three eigenvalues with an  $\epsilon$  hierarchy and the model passes the naturalness conditions. On the other hand, for

$$\mathcal{M} = \mathcal{O} \begin{pmatrix} \epsilon^3 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon & 1 \end{pmatrix} \tag{3.12}$$

it follows that  $r_3 = 0$ ,  $r_2 = 2$  and  $r_1 = 4$ . Hence,  $\lambda_3 = \mathcal{O}(1)$ ,  $\lambda_2 = \mathcal{O}(\epsilon^2)$  and  $\lambda_1 = \mathcal{O}(\epsilon^2)$ , so this model fails the naturalness test.

The above naturalness test significantly constraints the range of viable charge patterns. A detailed analysis, summarised in appendix A, leads to the following two necessary conditions:

- The spectrum should contain no pair of **10**-multiplets with identical  $\mathcal{G}$ -charge.
- If  $\bar{\mathbf{5}}_{a,b}^H$  is the multiplet containing the down-Higgs, the spectrum cannot contain both a **10<sub>a</sub>** and a **10<sub>b</sub>** multiplet.

Combining these constraints with the  $S_3$  permutation symmetry in the **10**-sector leads to a small number of possible charge choices for the **10** multiplets and the  $\bar{\mathbf{5}}^H$  multiplet which are listed in table 3. These charge choices are selected such that any of the other charge choices for a fixed vector  $\mathbf{n}$  will be equivalent to one of them up to the redundant permutation symmetry  $S_3 \times S_3 \times S_{N_\phi} \times S_{N_\Phi} \times S_f$ . In picking these charge choices, the redundant symmetry is broken down into a smaller subgroup as described in table 3 for each of the charge choices. It is notable that the split patterns  $\mathbf{n} = (1, 4)^T$  and  $\mathbf{n} = (2, 3)^T$  are ruled out completely. For the remaining cases with  $|\mathbf{n}| \geq 3$ , a short calculation shows that in order to generate the hierarchy in the up-Yukawa sector, at least two FN singlets are required.

For each of the 7 rows in table 3, the remaining number of charge assignments and the overall number of model, including VEV choices is given by

$$\tilde{N}_{\text{charges}} \simeq \frac{f^{6+2N_\phi} \Delta k^{(f-1)N_\Phi}}{8}, \quad \tilde{N}_{\text{models}} \simeq \tilde{N}_{\text{charges}} \Delta p^{N_\phi} \Delta \tilde{p}^{N_\Phi}. \tag{3.13}$$

For the example, with  $f = 5$ ,  $\Delta k = 10$ ,  $N_\phi = 4$ ,  $N_\Phi = 2$  and  $\Delta p = \Delta \tilde{p} = 10$ , this leads to  $\tilde{N}_{\text{models}} \simeq 10^{23}$ , still a very large landscape of string FN models.

$\mathbf{n}$	Charges				Symmetry after $q$ -fixing	$ G_{\text{red}} $
(1, 1, 1, 1, 1)	$\mathbf{10}_1$	$\mathbf{10}_2$	$\mathbf{10}_5$	$\bar{\mathbf{5}}_{4,5}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi} \times S_2$	$12N_\phi!N_\Phi!$
(1, 1, 1, 2)	$\mathbf{10}_1$	$\mathbf{10}_2$	$\mathbf{10}_4$	$\bar{\mathbf{5}}_{4,4}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi} \times S_2$	$12N_\phi!N_\Phi!$
	$\mathbf{10}_2$	$\mathbf{10}_3$	$\mathbf{10}_4$	$\bar{\mathbf{5}}_{1,4}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi} \times S_2$	$12N_\phi!N_\Phi!$
	$\mathbf{10}_1$	$\mathbf{10}_2$	$\mathbf{10}_3$	$\bar{\mathbf{5}}_{1,4}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi} \times S_2$	$12N_\phi!N_\Phi!$
	$\mathbf{10}_1$	$\mathbf{10}_3$	$\mathbf{10}_4$	$\bar{\mathbf{5}}_{1,2}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi}$	$6N_\phi!N_\Phi!$
(1, 1, 3)	$\mathbf{10}_1$	$\mathbf{10}_2$	$\mathbf{10}_3$	$\bar{\mathbf{5}}_{3,3}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi} \times S_2$	$12N_\phi!N_\Phi!$
(1, 2, 2)	$\mathbf{10}_1$	$\mathbf{10}_2$	$\mathbf{10}_3$	$\bar{\mathbf{5}}_{3,3}^H$	$S_3 \times S_{N_\phi} \times S_{N_\Phi}$	$6N_\phi!N_\Phi!$

**Table 3.** List of charge choices for  $\mathbf{10}^I$  and  $\bar{\mathbf{5}}^H$  required for a non-fine-tuned model. For the last two split patterns,  $\mathbf{n} = (1, 4)$  and  $\mathbf{n} = (2, 3)$ , no viable charge choices remain. The third column lists the order of the permutation symmetry  $G_{\text{red}}$  left in each sector.

The actual number of physically distinct models is obtained by dividing the total number of models  $\tilde{N}_{\text{models}}$  in eq. (3.13) by the order of the remaining permutation symmetry. For cases with a number of FN singlets less than four, this reduction may lead to a reasonably small environment size which facilitates a complete scan of the model space. For example, for the case  $f = 3$ ,  $\Delta k = 8$ ,  $N_\phi = 3$ ,  $N_\Phi = 0$ ,  $\Delta p = \Delta \tilde{p} = 8$ , leads to  $\tilde{N}_{\text{models}}/|G_{\text{red}}| \simeq 10^6$  models after removing symmetry redundancies, so that a complete scan is computationally feasible. On the other hand, the choice  $f = 5$ ,  $\Delta k = 10$ ,  $N_\phi = 4$ ,  $N_\Phi = 2$  and  $\Delta p = \Delta \tilde{p} = 10$  leads to  $\tilde{N}_{\text{models}}/|G_{\text{red}}| \simeq 10^{21}$  models and heuristic search methods have to be used.

### 3.3 The search strategy

Our approach to exploring the string FN landscape has two main phases. First, we perform a GA search (or, for cases with a sufficiently small number of models, a systematic search) to identify charge assignments for the SM multiplets and FN singlets, together with  $\epsilon$ -powers for the singlet VEVs that generate promising Yukawa textures which lead to roughly correct quark mass hierarchies and mixing. The Yukawa textures are obtained by constructing the leading  $\mathcal{G}$ -invariant operators using the method presented in appendix B. This step focuses on the quark sector only, as the charged lepton mass hierarchy approximately follows the down-quark sector.

In the second phase, we optimize the order-one coefficients in all three Yukawa matrices as well as the  $\epsilon$ -scale, ensuring the Higgs VEV, quark and lepton masses, and CKM matrix match their experimentally measured values at the electroweak scale  $M_Z$  within a few percent. A few further remarks are in order.

The string FN models studied here involve two types of singlets, perturbative and non-perturbative, each coming with a distinct pattern of  $\mathcal{G}$ -charges. For any model containing only perturbative singlet fields, it is straightforward to construct (typically many) physically equivalent models by enhancing the spectrum with additional non-perturbative singlet fields whose contributions to the Yukawa matrices are sub-leading.

The converse direction is not entirely straightforward and, in general, it is not possible to associate to any given model including both perturbative and non-perturbative singlets a physically equivalent model that is entirely perturbative. The charges of non-perturbative singlet fields can, in principle, be reproduced by suitable combinations of perturbative singlets. However, the VEVs of the charge-equivalent combinations of perturbative singlets must also match the VEV of the non-perturbative singlet. This may not be compatible with the allowed VEV ranges and, moreover, the new operator insertion obtained by replacing a non-perturbative singlet may not remain leading-order. An example illustrating the relation between perturbative and non-perturbative models is given in section 6.2.

Finally, we note that models with a smaller number of singlets can be included as special cases of models with a higher number of singlets (perturbative, non-perturbative or both). This can be done either by adding perturbative singlets of the form  $\phi_{a,a}$  which are  $\mathcal{G}$ -neutral and hence do not contribute to the leading order operators or by adding non-perturbative fields  $\Phi^{(i)}$  with high charges that do not contribute to the leading order operators and setting the corresponding coefficient  $\beta_i = 0$  in eq. (2.10) so the anomaly cancellation conditions remain satisfied.

It is also worth noting that, since the charge patterns for the perturbative singlets are much more restrictive than those for the non-perturbative singlets, the space of entirely perturbative models is much smaller. For this reason, it will be useful to search in section 5 the space of models containing only perturbative singlets before tackling the full string FN models.

## 4 Genetic algorithms and the string FN environment

As we have seen in the previous section, for many cases of interest the space of string FN models is too large to be amenable to systematic searches, so heuristic methods are required. Genetic algorithms (GAs) (see refs. [37–39] for reviews) have already proved to be extremely powerful in identifying physically viable models in a variety of contexts, including inflationary cosmology [40, 41], BSM physics [42–44], and string phenomenology [45–50]. We start this section with a short review of GAs and then discuss the environment of string FN models.

### 4.1 Genetic algorithms: a brief review

Genetic Algorithms (GA) belong to a class of search algorithms motivated by natural evolution. Using mathematical parlance, the basic ingredients of a GA are summarised in the following diagram.

$$[0, 1] \xleftarrow{\pi_0} \mathbb{Z}_2^{N_{\text{bits}}} \xrightarrow{\psi} E \xrightarrow{f} \mathbb{R}_{\leq 0} \quad (4.1)$$

Here,  $E$  is the environment, also referred to as the space of phenotypes. Roughly speaking, for our applications,  $E$  is the space of string FN models. On this space, we have a real-valued function  $f$ , the fitness function, which measures the ‘quality’ of phenotypes. In our case,  $f$  has a number of contributions (to be explained below) but the most important one measures the degree to which a certain string FN model reproduces the measured values of fermion masses and mixing. Phenotypes are encoded as bit sequences, or genotypes, of

length  $N_{\text{bits}}$  and the map  $\psi$  is called the genotype-phenotype map. In general, this map does not need to be injective.

The genetic algorithm starts by selecting an initial population  $P_0 \subset \mathbb{Z}_2^{N_{\text{bits}}}$ , that is a genotype list of length  $N_{\text{pop}}$ , randomly selected using a probability distribution  $\pi_0$  (often, as it is also the case here, taken to be the flat distribution). The initial population is evolved in  $N_{\text{gen}}$  steps,

$$P_0 \longrightarrow P_1 \longrightarrow \dots \longrightarrow P_{N_{\text{gen}}} , \tag{4.2}$$

where each individual step,  $P_i \rightarrow P_{i+1}$ , involves carrying out three types of operations: (i) selection, (ii) cross-over, (iii) mutation. We briefly explain these operations in turn.

- (i) For selection, a probability distribution  $\pi_i : P_i \rightarrow [0, 1]$  is introduced on the  $i^{\text{th}}$  population. This is done such that the probability  $p_i(b)$  of a genotype  $b \in P_i$  increases with their fitness  $f(\psi(b))$ . There are several concrete realisations of selection, but here we adopt selection by ranking. Genotypes  $b \in P_i$  are first ranked by their fitness and the probability  $\pi_i$  depends linearly on the rank  $r(b)$ , specifically

$$\pi_i(b) = \frac{2}{(1 + \alpha)N_{\text{pop}}} \left( 1 + \frac{N_{\text{pop}} - r(b)}{N_{\text{pop}} - 1}(\alpha - 1) \right) . \tag{4.3}$$

Here,  $\alpha$  is a parameter, typically chosen in the range  $\alpha \in [2, 5]$ , which can be used to adjust the degree of bias towards fit individuals. Using the distribution  $\pi_i$ ,  $N_{\text{pop}}/2$  pairs of genotypes from  $P_i$  are selected.

- (ii) For each pair  $(b_1, b_2)$  selected in the previous step, a cross-over is performed. This amounts to randomly selecting a position along the bit string and swapping the ‘tails’ of  $b_1$  and  $b_2$  to the right of this position, thereby obtained a new pair  $(\tilde{b}_1, \tilde{b}_2)$ . Carrying this out for all pairs leads to a new population,  $\tilde{P}_{i+1}$  of the same size  $N_{\text{pop}}$ .
- (iii) Finally, a random fraction,  $p_{\text{mut}}$ , of bits in the population  $\tilde{P}_{i+1}$  is flipped and this leads to  $P_{i+1}$ .

When the evolution has been completed by iterating these steps, the ‘viable’ states  $b$  from all populations  $P_0, \dots, P_{N_{\text{gen}}}$  are collected. These are states with a sufficiently large fitness,  $f(b) \geq f_{\text{term}}$ , large enough to guarantee they satisfy all desired properties. For our applications, the main property of viable string FN models is that they reproduce the measured fermion masses and mixing to sufficient accuracy. By generating many evolutionary cycles starting from different random initial populations  $P_0$  and extracting viable states each time, it may be possible to obtain nearly complete lists of viable states, even for large environments [47, 48, 50].

## 4.2 The string FN environment

In order to apply GAs to string FN models, we have to provide a concrete realisation of the various ingredients in the diagram (4.1). We begin with the environment  $E$ .

First, we choose one of the partition vectors  $\mathbf{n}$  and charge patterns from table 3. We then fix the numbers  $N_\phi$  and  $N_\Phi$  of perturbative and non-perturbative singlet fields. Given these

choices, the associated string FN environment  $E$  consists of all possible charges not already fixed in table 3 and all choices of singlet VEVs. Concretely, this includes the charges of the three  $\bar{\mathbf{5}}_{a_I, b_I}^I$  families (or, rather, the directions in which the  $\bar{\mathbf{5}}$  multiplets are charged), the charges of the perturbative singlets  $\phi_{c_\alpha, d_\alpha}^{(\alpha)}$ , where  $\alpha = 1, \dots, N_\phi$ , and the charges  $\mathbf{k}^i$ , where  $i = 1, \dots, N_\Phi$ , of the non-perturbative singlets. To account for the VEVs in eq. (3.2), we also need to include the integers  $p_\alpha$  and  $\tilde{p}_i$ . Altogether, the environment is then given by

$$E = \{1, \dots, f\}^{(6+2N_\phi)} \times \{k_{\min}, \dots, k_{\max}\}^{N_\Phi} \times \{1, \dots, p_{\max}\}^{N_\phi} \times \{\tilde{p}_{\min}, \dots, \tilde{p}_{\max}\}^{N_\Phi} \quad (4.4)$$

In order to encode the elements of  $E$  (states) as bit strings, we proceed in several steps. Firstly, consider the directions  $(a_I, b_I)$  and  $(c_\alpha, d_\alpha)$  in which the  $\bar{\mathbf{5}}$  multiplets and the  $\phi$ -singlets are, respectively, charged. These take values in the range  $\{1, \dots, f\}$  and we collate them into a single number in base  $f$ , then translate this into base 2, obtaining a bit string of length  $\lceil \log_2(f^{6+2N_\phi}) - 1 \rceil$ , where  $\lceil \cdot \rceil$  is the ceiling function.

For the  $T$ -charges, we choose  $k_{\max} = k_{\min} + 2^{N_k} - 1$ , for some positive integer  $N_k$ , such that each integer  $k_a^i$  can be represented by  $N_k$  bits. Finally, for the VEV integers, we choose  $p_{\max} = 2^{N_p}$  and  $\tilde{p}_{\max} = \tilde{p}_{\min} + 2^{N_{\tilde{p}}} - 1$ , such that each  $p_\alpha$  and  $\tilde{p}_i$  is represented by  $N_p$  and  $N_{\tilde{p}}$  bits, respectively. By concatenating these various pieces we obtain a bit string with length

$$N_{\text{bits}} = \lceil \log_2(f^{6+2N_\phi}) - 1 \rceil + N_\Phi(f-1)N_k + N_\phi N_p + N_\Phi N_{\tilde{p}}. \quad (4.5)$$

The genotype-phenotype map  $\psi : \mathbb{Z}_2^{N_{\text{bits}}} \rightarrow E$  is defined by inverting this process. Due to the ceiling function in the first term in  $N_{\text{bits}}$ , the part of the bit string that encodes the charges of the  $\bar{\mathbf{5}}$  families and perturbative singlets may sometimes produce, under inversion, a number that is too large; whenever this happens, we drop the first digit in the base- $f$  number after performing the inversion.

### 4.3 The fitness function

Consider a bit string  $b \in \mathbb{Z}_2^{N_{\text{bits}}}$  and its associated phenotype  $\psi(b) \in E$ . In order to determine the physical viability of the model, we need to fix — at least approximately, in this first phase of GA search — the  $\epsilon$ -scale for the VEVs and the order-one coefficients.

The  $\epsilon$ -scale is set by first computing the  $\epsilon$ -order of magnitude for the eigenvalues of the up-type and down-type quark Yukawa matrices for generic values of the order-one coefficients. As in section 3.2, we define  $\mathcal{M}^u = M^u(M^u)^\dagger$  and  $\mathcal{M}^d = M^d(M^d)^\dagger$  where  $M^u$ ,  $M^d$  are the mass matrices for the up and, respectively, down quark sector. The eigenvalues of  $\mathcal{M}^u$  and  $\mathcal{M}^d$  satisfy

$$\begin{aligned} \lambda_1^u &= \mathcal{O}(\epsilon^{r_1^u - r_2^u - r_3^u}), & \lambda_2^u &= \mathcal{O}(\epsilon^{r_2^u - r_3^u}), & \lambda_3^u &= \mathcal{O}(\epsilon^{r_3^u}), \\ \lambda_1^d &= \mathcal{O}(\epsilon^{r_1^d - r_2^d - r_3^d}), & \lambda_2^d &= \mathcal{O}(\epsilon^{r_2^d - r_3^d}), & \lambda_3^d &= \mathcal{O}(\epsilon^{r_3^d}), \end{aligned} \quad (4.6)$$

where we have used the same conventions as in section 3.2. We then match the their ratios to the observed quark mass ratios measured at the electroweak scale  $M_Z$ . Concretely, we

define the four quantities  $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$  by setting

$$\frac{\lambda_1^u}{\lambda_2^u} \sim \epsilon_2^{r_1^u - 2r_2^u} = \frac{m_u^2}{m_c^2}, \quad \frac{\lambda_2^u}{\lambda_3^u} \sim \epsilon_1^{r_2^u - 2r_3^u} = \frac{m_c^2}{m_t^2} \quad (4.7)$$

$$\frac{\lambda_1^d}{\lambda_2^d} \sim \epsilon_4^{r_1^d - 2r_2^d} = \frac{m_d^2}{m_s^2}, \quad \frac{\lambda_2^d}{\lambda_3^d} \sim \epsilon_3^{r_2^d - 2r_3^d} = \frac{m_s^2}{m_b^2}. \quad (4.8)$$

Generically, the values  $\epsilon_i$  obtained from these four equations will be different. The  $\epsilon$ -scale is then chosen to minimise the logarithmic quadratic deviation, so it is given by the geometric mean of the  $\epsilon_i$ , that is,

$$\epsilon = (\epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4)^{1/4}. \quad (4.9)$$

The next step is to choose the 18 order-one coefficients for the up-type and down-type quark Yukawa matrices. Here we choose the order-one coefficients at  $M_Z$ -scale.<sup>8</sup> We would also need to specify the two Higgs VEVs,  $v^u$  and  $v^d$ . We start with a discussion of the latter, for which the following three relations are relevant:

$$m_t \approx v^u c_t \epsilon^{r_3^u}, \quad m_b \approx v^d c_b \epsilon^{r_3^d}, \quad v^2 = (v^u)^2 + (v^d)^2, \quad (4.10)$$

where  $v$  is the measured Higgs VEV and  $c_t, c_b$  are the order-one coefficients corresponding to the leading  $\epsilon$ -order entries along the diagonal in the two Yukawa matrices. To these equations, one can add the ratio  $\tan \beta = v^u/v^d$  of the vacuum expectation values of the Higgs fields in the MSSM. Phenomenologically,  $\tan \beta$  typically ranges from low (that is,  $\tan \beta \gtrsim 1$ ) to moderate values, implying a large SUSY breaking scale, which aligns with the non-observation of superpartners at the LHC [51, 52]. While  $\tan \beta$  is qualitatively significant, it does not serve as a strict quantitative discriminator between models, so its value does not directly contribute to the GA fitness. However, the electroweak breaking scale, manifested by the Higgs VEV  $v$ , must be matched to account for observed phenomenological effects. We hence proceed as follows.

The order-one coefficient  $c_b$  for the bottom Yukawa coupling can be set to 1 without loss of generality. This is so since the  $\epsilon$ -scale, as determined above, is only accurate up to a factor of order one. The order-one coefficient  $c_t$  can then be absorbed into the definition of the  $\epsilon$ -scale. The order-one coefficient  $c_t$  for the top Yukawa coupling is then fixed by matching the Higgs VEV,

$$v^2 = (v^u)^2 + (v^d)^2 \simeq \frac{m_t^2}{c_t^2 \epsilon^{2r_3^u}} + \frac{m_b^2}{\epsilon^{2r_3^d}}. \quad (4.11)$$

All other order-one coefficients are set to fixed random numbers (the same set of fixed random numbers for all models) in the range [0.5, 3.0] and ensuing quark masses and mixings are compared with their values at the electroweak scale  $M_Z$ .

Next, we should describe the fitness function  $f: E \rightarrow \mathbb{R}_{\leq 0}$ . The fitness  $f(b)$  of a state  $b \in E$  is given by the sum of several contributions,

$$f(b) = f_{\text{quark}}(b) + f_{\text{texture}}(b) + f_{\text{fine-tuning}}(b) + f_{\text{split type}}(b) + f_{\text{anomaly}}(b), \quad (4.12)$$

---

<sup>8</sup>As discussed in section 2.3, the coefficients in the Yukawa couplings should be of order-one at the GUT scale. However, during GA searches, we fix the coefficients to be of order one at the  $M_Z$  scale so as to reduce the computational cost associated with the numerical integration involved in the renormalisation.

described in detail in appendix C. Qualitatively, these contributions correspond to:

- $f_{\text{quark}}(b)$  measures how well the model reproduces the measured quark masses and mixings.
- $f_{\text{texture}}(b)$  checks whether the textures satisfy the naturalness conditions in eq. (3.10).
- $f_{\text{fine-tuning}}(b)$  measures how changes in the order-one coefficients will affect the measured quark masses and mixings. This is included so that the models found are not fine-tuned in the choice of order-one coefficients.
- $f_{\text{split type}}(b)$  accounts for the additional rule that the charge choice  $a = b$  for  $\bar{\mathbf{5}}_{a,b}$  is only allowed if the partition vector  $\mathbf{n}$  satisfies  $n_a > 1$ .
- $f_{\text{anomaly}}(b)$  measures how well the anomaly-cancellation conditions (2.10) can be satisfied.

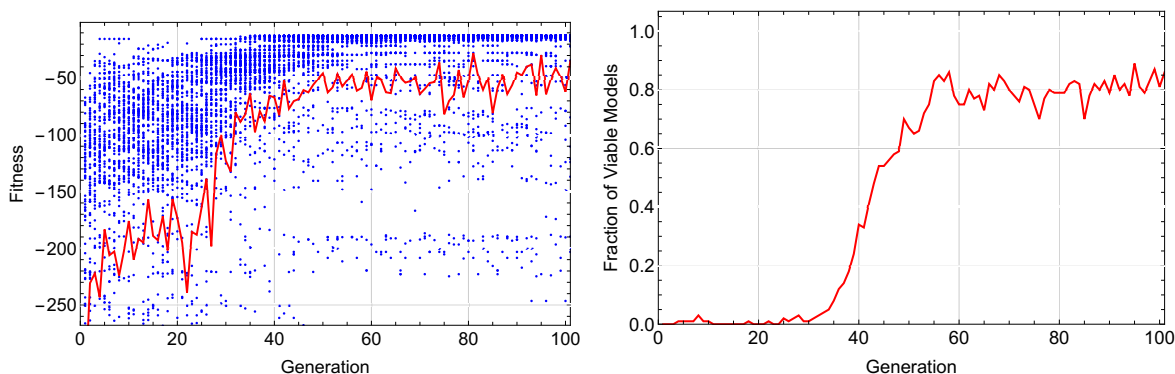
The  $f_{\text{quark}}(b)$  contribution is set such that two units of fitness correspond to an order of magnitude deviation away from any measured flavour parameter. A given model is deemed viable if the overall fitness is greater than  $-20$ , which roughly corresponds to allowing each of the 12 flavour parameters in the quark sector to deviate, on average, by no more than one order of magnitude from their measured values. Models that pass this criterion typically retain their viability when the order-one coefficients are changed to any other random, generic choice. The contributions  $f_{\text{split type}}(b)$  and  $f_{\text{anomaly}}(b)$  are set such that any model that violates the above charge rule or the anomaly cancellation conditions is not accepted as viable.

#### 4.4 Implementation of the search strategy

The size of the environment grows exponentially with the number of FN singlets. When the environment size is  $\lesssim 10^8$  after eliminating the permutation redundancies, it is computationally feasible to perform a complete scan over the entire space. In those cases, we have separately performed a complete scan over all possible charge assignments and VEV choices. For the larger environments, we use the environment described above, implemented both in Mathematica and in C, coupled to the GA packages [53] and [54], respectively. Since the charges of the perturbative singlets are much more constrained, we separately analyse the purely perturbative models and the models that also include non-perturbative insertions.

Figure 1 illustrates a typical GA run (evolutionary cycle), displaying a clear increase in the population fitness, for a randomly generated initial population. The first phase of the search (whether comprehensive or heuristic) is intended to produce a list of charge patterns and Yukawa textures that approximately account for the values of the flavour parameters in the quark sector.

In the second phase, we optimise the coefficients in the Yukawa couplings (assumed to be of order one at GUT scale) and the  $\epsilon$ -scale to obtain more accurate values for the Higgs VEV, the quark and charged lepton masses, and the CKM matrix. An ADAM optimiser realised in JAX is used for this latter purpose. The RG-effects on the Yukawa couplings, described in appendix D, are accounted for using the numerical integration module `diffrax` to obtain the Yukawa couplings at the  $M_Z$  scale.



**Figure 1.** The first plot shows the individual fitness (blue points) and the average population fitness (red curve) as a function of the generation, for a typical GA run in the string FN environment with  $\mathbf{n} = (1, 1, 3)$ ,  $(N_\phi, N_\Phi) = (5, 0)$  and the charges  $\mathbf{10}_1$ ,  $\mathbf{10}_2$ ,  $\mathbf{10}_3$  and  $\bar{\mathbf{5}}_{3,3}$  fixed. The second plot shows the fraction of viable models as a function of the generation.

More precisely, for the  $\epsilon$ -scale we search in the range  $[0.05, 0.9]$ , while for the coefficients in the Yukawa couplings at GUT scale we search in the range  $[0.5, 3]$ . The loss function is set to be the  $\ell^2$ -norm of the relative deviation vector including the Higgs VEV, quark masses and mixings and charged lepton masses, where deviations are computed with respect to the experimentally-measured Standard Model quantities:

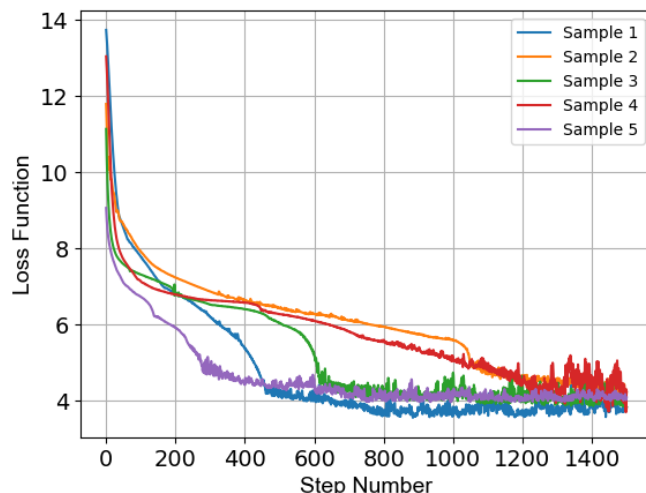
$$L(\delta) = \sqrt{\sum_\gamma \left( \frac{|\rho_\gamma - \rho_{\gamma,0}|}{\rho_{\gamma,0}(1 + 0.01\delta)} \right)^2} \tag{4.13}$$

where  $\rho_\gamma \in \{v, m^u, m^d, m^c, m^s, m^t, m^b, m^e, m^\mu, m^\tau, |(V_{\text{CKM}})_{ij}|\}$  are the parameters computed for the model, while  $\rho_{\gamma,0}$  are the corresponding experimental values. We use the stochastic variable  $\delta$  taken from a uniform distribution  $\delta \in [-1, 1]$  to encourage exploration beyond local optima. Given a large number of initial points in the search space, say, of order  $10^5$ , it is generally possible to optimise the  $\epsilon$ -scale and the set of order-one coefficients to obtain consistent masses and mixings. Figure 2 illustrates the evolution of the loss function during optimisation for five random initial points in a particular string FN model.

### 5 Models containing only perturbative singlets

We begin by searching for models that include only perturbative singlets since, for a given number of singlets, the resulting environment size in this configuration is substantially reduced compared to cases where both perturbative and non-perturbative singlets are considered, thus reducing the computational challenge. This setting also covers all the models containing non-perturbative singlets whose effects can be neglected.

The results of the scans and GA searches after reducing redundancies are shown in table 4. The naturalness conditions discussed in section 3 cannot be satisfied for a single FN scalar so we start with  $N_\phi = 2$  and cover all cases up to and including  $N_\phi = 5$ . We do not consider cases with  $N_\phi > 5$  as the computational cost becomes too large. Whenever the reduced size of the environment (as a consequence of the permutation symmetries listed table 3) is smaller or equal to  $10^8$ , we perform a complete scan.



**Figure 2.** Optimisation of the  $\epsilon$ -scale and order-one coefficients using the ADAM optimiser. The plot shows the evolution of the loss function during optimisation. The model corresponds to the partition vector  $\mathbf{n} = (1, 1, 1, 1, 1)$  and the spectrum  $\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_5, \bar{\mathbf{5}}_{1,2}, \bar{\mathbf{5}}_{1,2}, \bar{\mathbf{5}}_{1,2}, \bar{\mathbf{5}}_{4,5}^H, \mathbf{5}_{4,5}^H, \phi_{5,1}, \phi_{3,5}, \phi_{4,5}, \phi_{1,2}, \phi_{4,1}$ . The five curves correspond to different random initialisations. The loss function (4.13) is plotted on a log scale.

As expected, the number of viable models increases with the value of  $N_\phi$ . In particular, we were able to identify only one viable model with  $N_\phi = 2$ , while for larger values of  $N_\phi$ , the number of models grows roughly by one to two orders of magnitude for each additional perturbative singlet in the spectrum, irrespective of the partition vector  $\mathbf{n}$  and the charge assignments for the  $\mathbf{10}^I$  and  $\bar{\mathbf{5}}^H$  multiplets.

It is remarkable that the GA is capable of dealing with environments as large as  $10^{15}$  states. To demonstrate the performance, we have included in figure 3 a saturation plot for the number of inequivalent (with respect to the permutation symmetries listed in table 3) viable spectra (that is, charge patterns) identified as a function of the number of states visited. The plot suggests that almost all viable charge patterns in the given sector have been identified. From a high energy point of view, the set of viable charge patterns identified in the search may be translated into cohomological constraints for the compactification data. To this end, the viable charge patterns are more relevant than the full set of details describing viable models, which includes the VEV powers for the FN scalars. Below we exemplify the results by discussing in detail two viable models identified in the search.

### 5.1 Example model with $\mathbf{n} = (1, 1, 1, 1, 1)$

This model has already been mentioned in the Introduction, but we now present it in full detail. Its spectrum is given by

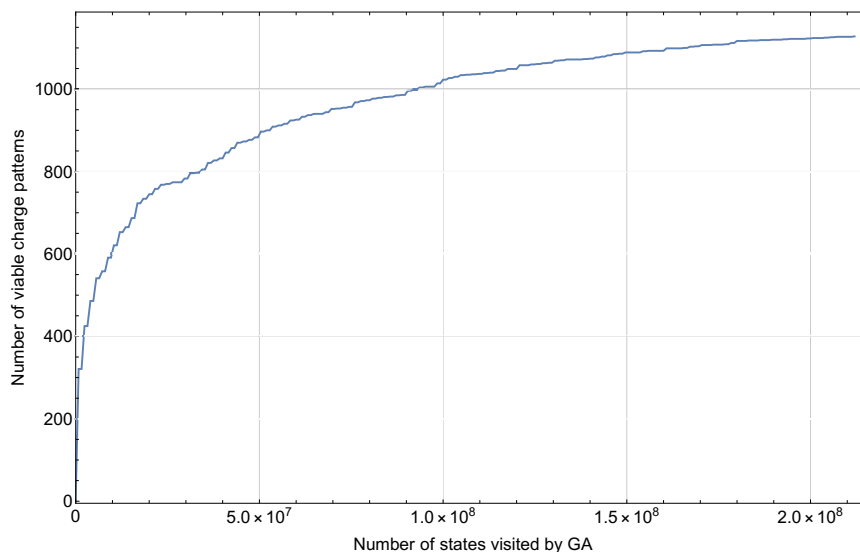
$$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_5, \bar{\mathbf{5}}_{1,2}, \bar{\mathbf{5}}_{1,2}, \bar{\mathbf{5}}_{1,2}, H_{4,5}^u, H_{4,5}^d, \phi_{5,1}, \phi_{3,5}, \phi_{4,5}, \phi_{1,2}, \phi_{4,1} \quad (5.1)$$

and the  $\epsilon$ -orders of magnitude for the singlet VEVs are

$$\langle \phi_{5,1} \rangle \sim \epsilon^3, \quad \langle \phi_{3,5} \rangle \sim \epsilon^4, \quad \langle \phi_{4,5} \rangle \sim \epsilon, \quad \langle \phi_{1,2} \rangle \sim \epsilon^5, \quad \langle \phi_{4,1} \rangle \sim \epsilon^4.$$

$\mathbf{n}$	Fixed charges	$N_\phi$	Env Size	States visited	Full scan	Models	Inequiv. spectra
(1, 1, 1, 1, 1)	$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_5, \bar{\mathbf{5}}_{4,5}^H$	2	$10^9$		Yes	0	0
		3	$10^{11}$	$10^8$		0	0
		4	$10^{13}$	$10^8$		301	4
		5	$10^{16}$	$10^9$		29213	289
(1, 1, 1, 2)	$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_4, \bar{\mathbf{5}}_{4,4}^H$	2	$10^8$		Yes	0	0
		3	$10^{10}$		Yes	98	1
		4	$10^{12}$	$10^8$		18825	55
		5	$10^{14}$	$10^8$		320557	449
(1, 1, 1, 2)	$\mathbf{10}_2, \mathbf{10}_3, \mathbf{10}_4, \bar{\mathbf{5}}_{1,4}^H$	2	$10^8$		Yes	0	0
		3	$10^{10}$		Yes	56	1
		4	$10^{12}$	$10^8$		11538	128
		5	$10^{14}$	$10^8$		259175	1128
(1, 1, 1, 2)	$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_3, \bar{\mathbf{5}}_{1,4}^H$	2	$10^8$		Yes	0	0
		3	$10^{10}$		Yes	70	3
		4	$10^{12}$	$10^8$		8110	63
		5	$10^{14}$	$10^8$		204148	500
(1, 1, 1, 2)	$\mathbf{10}_1, \mathbf{10}_3, \mathbf{10}_4, \bar{\mathbf{5}}_{1,2}^H$	2	$10^8$		Yes	0	0
		3	$10^{10}$		Yes	0	0
		4	$10^{12}$	$10^8$		0	0
		5	$10^{14}$	$10^8$		0	0
(1, 1, 3)	$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_3, \bar{\mathbf{5}}_{3,3}^H$	2	$10^6$		Yes	8	1
		3	$10^8$		Yes	1218	18
		4	$10^{10}$		Yes	22734	81
		5	$10^{12}$	$10^8$		154532	234
(1, 2, 2)	$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_3, \bar{\mathbf{5}}_{3,3}^H$	2	$10^6$		Yes	0	0
		3	$10^8$		Yes	0	0
		4	$10^{10}$		Yes	0	0
		5	$10^{12}$	$10^8$		0	0

**Table 4.** Search results for fully perturbative string FN models. The table includes the partition vector  $\mathbf{n}$ , the fixed charges for the  $\mathbf{10}$  and  $\bar{\mathbf{5}}^H$  multiplets, the number of singlets  $N_\phi$ , the size of the GA environment, the number of visited states in the GA search followed by an indication of whether a complete scan is feasible and was performed and, finally, the number of viable models and the number of inequivalent charge patterns found.



**Figure 3.** Saturation plot for the number of viable (terminal) charge patterns as a function of the number of states visited by the GA. The models correspond to a partition vector  $\mathbf{n} = (1, 1, 1, 2)$  and fixed charges  $\mathbf{10}_2, \mathbf{10}_3, \mathbf{10}_4, \bar{\mathbf{5}}_{1,4}^H$  with  $N_\phi = 5$ . The plot shows the total number of distinct charge patterns obtained in the GA search.

This implies the leading-order singlet insertions

$$\begin{aligned}
 \text{up sector:} & \quad \begin{pmatrix} \phi_{5,1}\phi_{4,1} & \phi_{5,1}\phi_{1,2}\phi_{4,1} & \phi_{4,1} \\ \phi_{5,1}\phi_{1,2}\phi_{4,1} & \phi_{5,1}\phi_{1,2}^2\phi_{4,1} & \phi_{1,2}\phi_{4,1} \\ \phi_{4,1} & \phi_{1,2}\phi_{4,1} & \phi_{4,5} \end{pmatrix}, \\
 \text{down sector:} & \quad \begin{pmatrix} \phi_{5,1}\phi_{3,5} & \phi_{5,1}\phi_{3,5} & \phi_{5,1}\phi_{3,5} \\ \phi_{5,1}\phi_{3,5}\phi_{1,2} & \phi_{5,1}\phi_{3,5}\phi_{1,2} & \phi_{5,1}\phi_{3,5}\phi_{1,2} \\ \phi_{3,5} & \phi_{3,5} & \phi_{3,5} \end{pmatrix},
 \end{aligned} \tag{5.2}$$

producing the Yukawa textures

$$\text{up sector} \sim \begin{pmatrix} \epsilon^7 & \epsilon^{12} & \epsilon^4 \\ \epsilon^{12} & \epsilon^{17} & \epsilon^9 \\ \epsilon^4 & \epsilon^9 & \epsilon \end{pmatrix}, \quad \text{down sector} \sim \begin{pmatrix} \epsilon^7 & \epsilon^7 & \epsilon^7 \\ \epsilon^{12} & \epsilon^{12} & \epsilon^{12} \\ \epsilon^4 & \epsilon^4 & \epsilon^4 \end{pmatrix}. \tag{5.3}$$

Diagonalising these implies the following patterns for the up-quark, down-quark and charged lepton masses:

$$(m_u, m_c, m_t) \sim (\epsilon^{17}, \epsilon^7, \epsilon), \quad (m_d, m_s, m_b), (m_e, m_\mu, m_\tau) \sim (\epsilon^{12}, \epsilon^7, \epsilon^4). \tag{5.4}$$

For the GA fitness analysis, we computed the parameters  $\epsilon = 0.44$ , and  $c_t = 3.0$ . The resulting Higgs VEVs were given by  $v^u = 130 \text{ GeV}$ ,  $v^d = 35.0 \text{ GeV}$ . With these choices, the total fitness value of the state was  $-16.4$ , hence the model passed as viable.

Higgs VEV	$\langle H \rangle = 174 \text{ GeV}$		
<b>Quark</b>	$m_u$ (MeV)	$m_c$ (GeV)	$m_t$ (GeV)
<b>Mass</b>	2.16	1.27	173
<b>Quark</b>	$m_d$ (MeV)	$m_s$ (MeV)	$m_b$ (GeV)
<b>Mass</b>	4.70	93.9	4.18
<b>Lepton</b>	$m_e$ (MeV)	$m_\mu$ (MeV)	$m_\tau$ (GeV)
<b>Mass</b>	0.511	106	1.78

**Table 5.** Values for the Higgs VEV, quark and lepton masses predicted by the model in section 5.1.

Finally, we carry out the optimisation of the 27 order-one coefficients in both the quark and charged lepton sectors and of the  $\epsilon$ -scale. The optimised  $\epsilon$ -scale is  $\epsilon = 0.554$  and the optimised order-one coefficients are:

$$\begin{array}{ccc}
 \begin{pmatrix} 1.090 & 2.282 & 1.896 \\ 0.961 & 2.027 & 1.979 \\ 1.966 & 2.978 & 2.648 \end{pmatrix} & 
 \begin{pmatrix} 1.379 & 1.843 & 0.947 \\ 2.708 & 1.726 & 2.063 \\ 1.530 & 2.526 & 0.680 \end{pmatrix} & 
 \begin{pmatrix} 1.064 & 2.051 & 1.707 \\ 1.183 & 2.628 & 2.262 \\ 0.514 & 1.623 & 0.706 \end{pmatrix} \\
 \text{up-quark sector} & \text{down-quark sector} & \text{lepton sector}
 \end{array} \tag{5.5}$$

With these coefficients, the resulting values for the Higgs VEV, quark and lepton masses are listed in table 5.

The CKM matrix is given by

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 0.970 & 0.242 & 0.00359 \\ 0.242 & 0.969 & 0.0447 \\ 0.00733 & 0.0443 & 0.999 \end{pmatrix}. \tag{5.6}$$

In this model, the relative deviations for the Higgs VEV, the quark and charged lepton masses are  $< 1\%$ . For the CKM matrix elements the relative deviations are slightly larger, going up to  $\simeq 15\%$ . This can be seen as a consequence of our two-step optimisation procedure, where the mass hierarchy is already addressed in the first step and further refined in the second.

It is likely that the model can be further optimised, however, in order to meaningfully do this one should also include a discussion of supersymmetry breaking into the RG analysis, something that goes beyond the scope of the present work.

### 5.2 Example model with $\mathbf{n} = (1, 1, 3)$

We now look at a viable model with a different number of U(1) symmetries (2 instead of 4), corresponding to the partition vector  $\mathbf{n} = (1, 1, 3)$ . The model corresponds to the following spectrum:

$$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_3, \bar{\mathbf{5}}_{1,2}, \bar{\mathbf{5}}_{3,3}, \bar{\mathbf{5}}_{1,2}, H_{3,3}^u, H_{3,3}^d, \phi_{3,1}, \phi_{2,3}, \phi_{1,1}, \tag{5.7}$$

with FN scalar VEVs of orders

$$\langle \phi_{3,1} \rangle \sim \epsilon^7, \quad \langle \phi_{2,3} \rangle \sim \epsilon^8, \quad \langle \phi_{1,2} \rangle \sim \epsilon^5.$$

The leading-order singlet insertions are:

$$\text{up sector: } \begin{pmatrix} \phi_{3,1}^2 & \phi_{3,1}^2 \phi_{1,2} & \phi_{3,1} \\ \phi_{3,1}^2 \phi_{1,2} & \phi_{3,1}^2 \phi_{1,2}^2 & \phi_{3,1} \phi_{1,2} \\ \phi_{3,1} & \phi_{3,1} \phi_{1,2} & 1 \end{pmatrix} \quad \text{down sector: } \begin{pmatrix} \phi_{3,1} & \phi_{2,3} & \phi_{3,1} \\ \phi_{3,1} \phi_{1,2} & \phi_{2,3} \phi_{1,2} & \phi_{3,1} \phi_{1,2} \\ 1 & \phi_{2,3}^2 \phi_{1,2} & 1 \end{pmatrix} \quad (5.8)$$

giving the following Yukawa textures

$$\text{up sector} \sim \begin{pmatrix} \epsilon^{14} & \epsilon^{19} & \epsilon^7 \\ \epsilon^{19} & \epsilon^{24} & \epsilon^{12} \\ \epsilon^7 & \epsilon^{12} & 1 \end{pmatrix}, \quad \text{down sector} \sim \begin{pmatrix} \epsilon^7 & \epsilon^8 & \epsilon^7 \\ \epsilon^{12} & \epsilon^{13} & \epsilon^{12} \\ 1 & \epsilon^{21} & 1 \end{pmatrix}, \quad (5.9)$$

and the following textures of quark and charged lepton masses:

$$(m_u, m_c, m_t) \sim (\epsilon^{24}, \epsilon^{14}, 1), \quad (m_d, m_s, m_b), \quad (m_e, m_\mu, m_\tau) \sim (\epsilon^{12}, \epsilon^7, 1). \quad (5.10)$$

For the GA fitness analysis, we find the optimised values  $\epsilon = 0.58$ ,  $c_t = 0.99$  and Higgs VEVs  $v^u = 174$  GeV,  $v^d = 2.46$  GeV. The model has a GA fitness value of  $-11.0$ .

In the final parameter optimisation phase, we find the optimal value  $\epsilon = 0.692$  and the following optimised order-one coefficients:

$$\begin{array}{ccc} \begin{pmatrix} 1.998 & 1.567 & 2.033 \\ 3.000 & 1.662 & 0.985 \\ 2.544 & 2.989 & 2.400 \end{pmatrix} & \begin{pmatrix} 2.340 & 0.501 & 1.292 \\ 2.362 & 0.500 & 0.925 \\ 2.576 & 3.000 & 0.509 \end{pmatrix} & \begin{pmatrix} 0.541 & 0.700 & 2.115 \\ 0.650 & 0.908 & 2.518 \\ 2.305 & 2.984 & 2.452 \end{pmatrix} \\ \text{up-quark sector} & \text{down-quark sector} & \text{lepton sector} \end{array} \quad (5.11)$$

With these coefficients the flavour parameters resulting from this model are given in table 6.

The CKM matrix is given by

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 0.970 & 0.242 & 0.00359 \\ 0.242 & 0.969 & 0.0447 \\ 0.00733 & 0.0442 & 0.999 \end{pmatrix}. \quad (5.12)$$

In this model, the relative deviations for the Higgs VEV, the quark and charged lepton masses are  $< 2\%$ , while for the CKM matrix elements they are  $< 10\%$ , except for the off-diagonal CKM element  $|V_{31}|$  whose relative deviation is  $\simeq 14.8\%$ .

Higgs VEV	$\langle H \rangle = 174 \text{ GeV}$		
<b>Quark</b>	$m_u$ (MeV)	$m_c$ (GeV)	$m_t$ (GeV)
Mass	2.18	1.25	173
<b>Quark</b>	$m_d$ (MeV)	$m_s$ (MeV)	$m_b$ (GeV)
Mass	4.68	94.0	4.18
<b>Lepton</b>	$m_e$ (MeV)	$m_\mu$ (MeV)	$m_\tau$ (GeV)
Mass	0.502	106	1.78

**Table 6.** Values for the Higgs VEV, quark and lepton masses predicted by the model in section 5.2.

## 6 Models with non-perturbative singlets

We now proceed to the discussion of models which include both perturbative and non-perturbative singlets. The charge patterns for the non-perturbative singlets are less constrained than their perturbative counterparts, being constrained only by eq. (2.5). For this reason, the environment size becomes significantly larger. Heuristic search methods such as GAs are, therefore, the only viable method of search.

Another consequence of the increased computational cost is that, in this setting, we will no longer aim to attain saturation in the number of viable models identified as a function of the number of GA runs. Instead, we will conduct approximately  $10^8$  GA runs in each sector, as outlined in table 7 and table 9. This number is comparable to the number of runs required to achieve saturation in the perturbative case. The resulting counts of inequivalent charge patterns are also similar, which, overall, represent only a tiny fraction of the total models explored.

To illustrate the results, we focus on two partition vectors  $\mathbf{n}$ , namely the completely split case  $\mathbf{n} = (1, 1, 1, 1, 1)$  and the case  $\mathbf{n} = (1, 2, 2)$ . The latter did not lead to any viable models in the fully perturbative setting but, as we will see, this situation changes with the inclusion of non-perturbative singlets.

### 6.1 Models with $\mathbf{n} = (1, 1, 1, 1, 1)$

Following the arguments in section 3, in this setting naturalness conditions allow us to fix the following multiplet charges:  $\mathbf{10}_1$ ,  $\mathbf{10}_2$ ,  $\mathbf{10}_5$ , and  $\bar{\mathbf{5}}_{4,5}$ .

The results of the search are summarised in table 7. Except for the first case with  $(N_\phi, N_\Phi) = (0, 1)$  where an exhaustive scan is possible, the other environments are searched using GA with a fixed total number of visited states of  $10^8$  states in each sector (which amounts to about 7000 core hours of computation). The results are presented in the same format as in the fully perturbative case.

To illustrate the results, it is useful to present an example. We do this for the sector  $(N_\phi, N_\Phi) = (2, 2)$ , where we have found a model with the following spectrum

$$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_5, \bar{\mathbf{5}}_{3,4}, \bar{\mathbf{5}}_{3,4}, \bar{\mathbf{5}}_{3,4}, H_{4,5}^u, H_{4,5}^d, \phi_{4,5}, \phi_{5,1}, \Phi^{(1)}, \Phi^{(2)} \quad (6.1)$$

$N_\phi$	$N_\Phi$	Env. Size	States Visited	Full Scan	Models	Inequiv. $q$ -patterns
0	1	$10^9$		Yes	0	0
0	2	$10^{13}$	$10^8$		0	0
0	3	$10^{18}$	$10^8$		0	0
0	4	$10^{22}$	$10^8$		22	2
1	1	$10^{11}$	$10^8$		0	0
1	2	$10^{16}$	$10^8$		0	0
1	3	$10^{20}$	$10^8$		648	107
1	4	$10^{25}$	$10^8$		7377	2154
2	1	$10^{14}$	$10^8$		0	0
2	2	$10^{18}$	$10^8$		328	23
2	3	$10^{23}$	$10^8$		6073	1490
2	4	$10^{27}$	$10^8$		7354	1708
3	1	$10^{16}$	$10^8$		76	5
3	2	$10^{20}$	$10^8$		720	159
3	3	$10^{25}$	$10^8$		892	149
4	1	$10^{18}$	$10^8$		263	9
4	2	$10^{23}$	$10^8$		2896	628

**Table 7.** Results of the search for non-perturbative models with  $\mathbf{n} = (1, 1, 1, 1, 1)$  and fixed multiplet charges  $\mathbf{10}_1$ ,  $\mathbf{10}_2$ ,  $\mathbf{10}_5$ , and  $\mathbf{\bar{5}}_{4,5}$ . The table includes the number of perturbative and non-perturbative fields  $N_\phi$  and, respectively,  $N_\Phi$ , the size of the GA environment, the number of visited states in the GA search followed by an indication of whether a complete scan was performed, and finally, the number of viable models and the number of inequivalent charge patterns found, where equivalence is understood in the sense of the permutation symmetries listed in table 3.

and charges for the two non-perturbative singlets  $\Phi^{(1)}$  and  $\Phi^{(2)}$  given by

$$(k_a^i) = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & -1 \end{pmatrix} \begin{matrix} \Phi^{(1)} \\ \Phi^{(2)} \end{matrix} \tag{6.2}$$

The leading-order singlet insertions in the Yukawa matrices are then given by

$$\begin{aligned}
 \text{up sector:} & \quad \begin{pmatrix} \phi_{4,5}\phi_{5,1}^2 & \phi_{4,5}^2\phi_{5,1}^2\Phi_1 & \phi_{4,5}\phi_{5,1} \\ \phi_{4,5}^2\phi_{5,1}^2\Phi_1 & \phi_{4,5}^3\phi_{5,1}^2\Phi_1^2 & \phi_{4,5}^2\phi_{5,1}\Phi_1 \\ \phi_{4,5}\phi_{5,1} & \phi_{4,5}^2\phi_{5,1}\Phi_1 & \phi_{4,5} \end{pmatrix}, \\
 \text{down sector:} & \quad \begin{pmatrix} \phi_{5,1}\Phi_2 & \phi_{5,1}\Phi_2 & \phi_{5,1}\Phi_2 \\ \phi_{4,5}\phi_{5,1}\Phi_1\Phi_2 & \phi_{4,5}\phi_{5,1}\Phi_1\Phi_2 & \phi_{4,5}\phi_{5,1}\Phi_1\Phi_2 \\ \Phi_2 & \Phi_2 & \Phi_2 \end{pmatrix}.
 \end{aligned} \tag{6.3}$$

Before proceeding further, let us first discuss how the anomaly cancellation condition in eq. (2.10) can be satisfied. A direct evaluation of  $\mathbf{A}^{(5)}$  gives  $\mathbf{A}^{(5)} = (3, 3, 3, 3, 3) \in \mathbb{Z}\mathbf{n}$ , which is equivalent to the zero vector. A solution to the anomaly cancellation condition for this case is simply  $(\beta_i) = (0, 0)$ , a result that imposes a strong constraint on the possible underlying string compactifications [11].

For the computation of Yukawa textures, we have identified the following  $\epsilon$ -powers for the singlet VEVs:

$$\langle \phi_{4,5} \rangle \sim \epsilon^3, \quad \langle \phi_{5,1} \rangle \sim \epsilon^8, \quad \langle \Phi_1 \rangle \sim \epsilon^8, \quad \langle \Phi_2 \rangle \sim \epsilon^7.$$

We note that for this particular model, the VEVs of the perturbative singlet  $\phi_{5,1}$  has to be stretched to a high power for the model to produce viable Yukawa textures. The resulting textures are given by:

$$\text{up sector} \sim \begin{pmatrix} \epsilon^{19} & \epsilon^{30} & \epsilon^{11} \\ \epsilon^{30} & \epsilon^{40} & \epsilon^{22} \\ \epsilon^{11} & \epsilon^{22} & \epsilon^3 \end{pmatrix}, \quad \text{down sector} \sim \begin{pmatrix} \epsilon^{15} & \epsilon^{15} & \epsilon^{15} \\ \epsilon^{26} & \epsilon^{26} & \epsilon^{26} \\ \epsilon^7 & \epsilon^7 & \epsilon^7 \end{pmatrix}. \tag{6.4}$$

which imply the following leading  $\epsilon$ -powers for the quark and charged lepton masses:

$$(m_u, m_c, m_t) \sim (\epsilon^{40}, \epsilon^{19}, \epsilon^3), \quad (m_d, m_s, m_b), (m_e, m_\mu, m_\tau) \sim (\epsilon^{26}, \epsilon^{15}, \epsilon^7)$$

For the GA fitness analysis, we have computed the optimised parameters  $\epsilon = 0.71$ ,  $o_t = 2.85$  and the Higgs VEVs  $v^u = 167.97 \text{ GeV}$ ,  $v^d = 13.48 \text{ GeV}$ . The fitness value was  $-10.5$ .

In the second phase of optimisation we have identified an optimal  $\epsilon$ -scale at  $\epsilon = 0.77$  and the following order-one coefficients:

$$\begin{array}{ccc}
 \begin{pmatrix} 1.403 & 1.701 & 2.414 \\ 2.465 & 2.805 & 1.150 \\ 2.999 & 2.999 & 2.914 \end{pmatrix} & 
 \begin{pmatrix} 1.389 & 1.737 & 1.554 \\ 0.564 & 0.500 & 2.995 \\ 2.997 & 2.962 & 0.500 \end{pmatrix} & 
 \begin{pmatrix} 0.608 & 0.708 & 2.567 \\ 0.658 & 0.675 & 2.658 \\ 1.354 & 0.577 & 0.500 \end{pmatrix} \\
 \text{up-quark sector} & 
 \text{down-quark sector} & 
 \text{lepton sector}
 \end{array} \tag{6.5}$$

The resulting masses and mixing for this model are given in table 8.

Higgs VEV	$\langle H \rangle = 175 \text{ GeV}$		
<b>Quark</b>	$m_u$ (MeV)	$m_c$ (GeV)	$m_t$ (GeV)
<b>Mass</b>	2.17	1.27	173
<b>Quark</b>	$m_d$ (MeV)	$m_s$ (MeV)	$m_b$ (GeV)
<b>Mass</b>	4.75	93.8	4.18
<b>Lepton</b>	$m_e$ (MeV)	$m_\mu$ (MeV)	$m_\tau$ (GeV)
<b>Mass</b>	0.533	106	1.79

**Table 8.** Values for the Higgs VEV, quark and lepton masses predicted by the model in section 6.1.

The CKM matrix is given by

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 0.970 & 0.242 & 0.00362 \\ 0.242 & 0.969 & 0.0449 \\ 0.00739 & 0.0445 & 0.999 \end{pmatrix}. \quad (6.6)$$

The flavour parameters obtained in this model are only a few percentages away from their measured values, the largest deviation corresponding to the off-diagonal CKM element  $|V_{31}|$  of  $\simeq 14.1\%$ . Again, this is consistent with the mass hierarchy being more accurately accounted for in our two-step optimisation compared to the CKM matrix elements.

## 6.2 Models with $\mathbf{n} = (1, 2, 2)$

It is interesting to consider the partition vector  $\mathbf{n} = (1, 2, 2)$  for which we have been unable to identify any viable perturbative models and investigate whether viable models with non-perturbative singlets can be found. To do this, we fix the multiplet charges  $\mathbf{10}_1$ ,  $\mathbf{10}_2$ ,  $\mathbf{10}_3$ , and  $\bar{\mathbf{5}}_{3,3}^H$ .

A summary of the search results is given in table 9. In particular, this setting differs from the completely-split  $\mathbf{n} = (1, 1, 1, 1, 1)$  setting in that no viable models could be found with  $N_\phi \leq 3$ , but this may be attributed to the rather limited search performed. In fact, looking at the viable models with  $N_\phi = 4$ , many of these feature spectator singlets whose effects can be neglected at the level of analysis performed here. Such models belong, effectively, to the case  $N_\phi \leq 3$ .

To illustrate this discussion, we analyse a model from the sector  $(N_\phi, N_\Phi) = (4, 2)$  with the following spectrum:

$$\mathbf{10}_1, \mathbf{10}_2, \mathbf{10}_3, \bar{\mathbf{5}}_{1,3}, \bar{\mathbf{5}}_{1,3}, \bar{\mathbf{5}}_{1,3}, H_{3,3}^u, H_{3,3}^d, 3\phi_{3,2}, \phi_{2,1}, \Phi^{(1)}, \Phi^{(2)}. \quad (6.7)$$

Although there are three singlets  $\phi_{3,2}$  in the spectrum, it is clear that they can be treated as the same singlet. The charges of the non-perturbative fields  $\Phi^{(1)}$  and  $\Phi^{(2)}$  are

$$(k_a^i) = \begin{pmatrix} 0 & 2 & -2 \\ 1 & -1 & 0 \end{pmatrix} \begin{matrix} \Phi^{(1)} \\ \Phi^{(2)} \end{matrix}$$

$N_\phi$	$N_\Phi$	Env Size	States Visited	Full Scan	Models	Q-Patterns
0	1	$10^6$		Yes	0	0
0	2	$10^8$		Yes	0	0
0	3	$10^{11}$	$10^8$		0	0
1	1	$10^8$		Yes	0	0
1	2	$10^{10}$	$10^8$		0	0
1	3	$10^{13}$	$10^8$		0	0*
2	1	$10^9$	$10^8$		0	0
2	2	$10^{12}$	$10^8$		0	0*
2	3	$10^{15}$	$10^8$		0	0*
3	1	$10^{11}$	$10^8$		0	0*
3	2	$10^{14}$	$10^8$		0	0*
3	3	$10^{17}$	$10^8$		0	0*
4	1	$10^{13}$	$10^8$		25165	77
4	2	$10^{16}$	$10^8$		317293	1338

**Table 9.** Results of the search for non-perturbative models with  $\mathbf{n} = (1, 2, 2)$  and fixed multiplet charges  $\mathbf{10}_1$ ,  $\mathbf{10}_2$ ,  $\mathbf{10}_3$ , and  $\bar{\mathbf{5}}_{3,3}$ . The table includes the number of perturbative and non-perturbative fields  $N_\phi$  and, respectively,  $N_\Phi$ , the size of the GA environment, the number of visited states in the GA search followed by an indication of whether a complete scan was performed, and finally, the number of viable models and the number of inequivalent charge patterns found, where equivalence is understood in the sense of the permutation symmetries listed in table 3.

With  $\mathbf{A}^{(5)} = (6, 3, 6) \sim (3, -3, 0)$ , the anomaly-cancellation condition (2.10) specialises to

$$\begin{pmatrix} 0 & 1 \\ 2 & -1 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix}. \quad (6.8)$$

This equation is solved for  $(\beta_1, \beta_2) = (0, 3)$  and therefore the anomaly can be indeed cancelled by the Green-Schwarz mechanism. The leading-order singlet insertions are:

$$\begin{aligned} \text{up sector:} & \begin{pmatrix} \phi_{2,1}^2 \phi_{3,2}^2 & \phi_{2,1} \phi_{3,2}^2 & \phi_{2,1} \phi_{3,2} \\ \phi_{2,1} \phi_{3,2}^2 & \phi_{3,2}^2 & \phi_{3,2} \\ \phi_{2,1} \phi_{3,2} & \phi_{3,2} & 1 \end{pmatrix}, \\ \text{down sector:} & \begin{pmatrix} \phi_{2,1} \phi_{3,2} \Phi_1 & \phi_{2,1} \phi_{3,2} \Phi_1 & \phi_{2,1} \phi_{3,2} \Phi_1 \\ \phi_{3,2} \Phi_1 & \phi_{3,2} \Phi_1 & \phi_{3,2} \Phi_1 \\ \Phi_1 & \Phi_1 & \Phi_1 \end{pmatrix}. \end{aligned} \quad (6.9)$$

The leading contributions in the up-sector are entirely perturbative. However, this is not the case in the down-sector where the singlet  $\Phi^{(1)}$  features, while  $\Phi^{(2)}$  is a spectator field and its main purpose is to facilitate anomaly cancellation. The model fails to appear in the list of perturbative models from section 5 despite the fact that  $\Phi^{(1)}$  can be replaced by the product of  $\phi_{2,3}^2$  of two perturbative singlets (assuming a spectrum that contains  $\phi_{3,2}$ ,  $\phi_{2,1}$ , and  $\phi_{2,3}$  instead of the above spectrum). This happens because the leading singlet insertion in the first line of the down Yukawa matrix would, in this case, be  $\phi_{2,1}\phi_{2,3}$  rather than  $\phi_{2,1}\phi_{3,2}\phi_{2,3}^2$ . This discussion illustrates the relationship between the classes of purely perturbative models and the class of models with both types of singlets, as discussed in section 3. While it is always possible to add non-perturbative singlets without significantly changing the phenomenology, it is not generally possible to replace every model with non-perturbative singlets by an equivalent purely perturbative one.

For the calculation of Yukawa textures, we have identified the following VEV patterns:

$$\langle \phi_{3,2} \rangle \sim \epsilon^7, \quad \langle \phi_{2,1} \rangle \sim \epsilon^6, \quad \langle \Phi_1 \rangle \sim \epsilon^5$$

which lead to the Yukawa textures

$$\text{up sector} \sim \begin{pmatrix} \epsilon^{26} & \epsilon^{20} & \epsilon^{13} \\ \epsilon^{20} & \epsilon^{14} & \epsilon^7 \\ \epsilon^{13} & \epsilon^7 & 1 \end{pmatrix}, \quad \text{down sector} \sim \begin{pmatrix} \epsilon^{18} & \epsilon^{18} & \epsilon^{18} \\ \epsilon^{12} & \epsilon^{12} & \epsilon^{12} \\ \epsilon^5 & \epsilon^5 & \epsilon^5 \end{pmatrix}, \quad (6.10)$$

and the masses

$$(m_u, m_c, m_t) \sim (\epsilon^{26}, \epsilon^{14}, 1), \quad (m_d, m_s, m_b), (m_e, m_\mu, m_\tau) \sim (\epsilon^{18}, \epsilon^{12}, \epsilon^5)$$

For the GA fitness calculation we found the optimal parameters  $\epsilon = 0.62$ ,  $c_t = 1.03$  and, furthermore, the Higgs VEVs  $v^u = 166$  GeV and  $v^d = 122$  GeV. The fitness value of the model was  $-5.1$ .

In the phase of optimisation, we find that  $\epsilon = 0.40$  and the optimised order-one coefficients

$$\begin{array}{ccc} \begin{pmatrix} 1.648 & 1.628 & 2.999 \\ 1.189 & 2.025 & 2.237 \\ 2.855 & 2.999 & 2.238 \end{pmatrix} & \begin{pmatrix} 2.698 & 2.254 & 1.096 \\ 1.255 & 1.248 & 0.501 \\ 3.000 & 2.699 & 0.500 \end{pmatrix} & \begin{pmatrix} 1.351 & 1.389 & 2.866 \\ 0.884 & 0.888 & 2.980 \\ 0.790 & 0.811 & 0.500 \end{pmatrix} \\ \text{up-quark sector} & \text{down-quark sector} & \text{lepton sector} \end{array} \quad (6.11)$$

lead to the masses and Higgs VEV listed in table 10.

The CKM matrix is:

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 0.970 & 0.241 & 0.00358 \\ 0.241 & 0.969 & 0.0448 \\ 0.00733 & 0.0443 & 0.999 \end{pmatrix}. \quad (6.12)$$

These flavour parameters are only a few percentages away from their measured values, the largest deviation being that of  $|V_{31}|$  of  $\simeq 14.7\%$ . This model highlights the fact that some

Higgs VEV	$\langle H \rangle = 175 \text{ GeV}$		
<b>Quark</b>	$m_u$ (MeV)	$m_c$ (GeV)	$m_t$ (GeV)
<b>Mass</b>	2.20	1.27	173
<b>Quark</b>	$m_d$ (MeV)	$m_s$ (MeV)	$m_b$ (GeV)
<b>Mass</b>	4.70	93.3	4.18
<b>Lepton</b>	$m_e$ (MeV)	$m_\mu$ (MeV)	$m_\tau$ (GeV)
<b>Mass</b>	0.521	107	1.84

**Table 10.** Values for the Higgs VEV, quark and lepton masses predicted by the model in section 6.2.

of the non-perturbative singlets may act as spectator fields (that is, they appear only in sub-leading contributions to the Yukawa couplings), however, their presence is needed to satisfy the anomaly cancellation conditions.

## 7 Conclusion

In this paper we have analysed a class of supersymmetric Froggatt-Nielsen (FN) models with multiple horizontal  $U(1)$  symmetries and multiple SM singlets whose structure was motivated by heterotic string compactifications on smooth Calabi-Yau threefolds with split rank-five holomorphic vector bundles. Our approach mirrors the FN mechanism, however, key differences arise due to the string-theoretic origin which, in particular, implies a certain pattern of charges for the SM multiplets and the FN singlets, dividing the latter into two types (perturbative and non-perturbative singlets). The first phase of exploration involved the identification of charge assignments for the SM multiplets and FN singlets, as well as singlet VEVs in terms of powers of a single scale  $\epsilon$ , such that the observed mass hierarchies in the quark sector were reproduced qualitatively. We have employed a combination of comprehensive and heuristic (GA) search methods with a dedicated search for purely perturbative models, followed by a larger search for models involving both perturbative and non-perturbative FN singlets. The second phase involved an optimisation of the singlet VEVs and the order-one coefficients in the Yukawa matrices such that all flavour parameters reproduced their measured values within a few percentages. The optimisation was successful for the large majority of models generated in the first phase.

In the fully perturbative sector we were able to identify several thousands of inequivalent spectra that can accommodate (with suitable choices of FN VEVs and order-one coefficients in the Yukawa matrices) the measured values for the Higgs VEV, the masses for the quark and charged leptons and the CKM matrix. While not exhaustive, we have run the GA until saturation, that is, up to the point where the number of new charge patterns becomes small in comparison with the number of charge patterns already identified. Remarkably, we were able to achieve saturation even for environments as large as  $10^{16}$  while visiting only a  $10^{-7}$  fraction of the available models.

The environment in the non-perturbative sector was significantly larger, going up to a size of  $10^{25}$  in certain sub-sectors and we have not attempted to reach any level of saturation during GA explorations. We have, however, identified several thousand of inequivalent charge patterns in this sector too. Some of the resulting models overlap with models from the perturbative sector, owing to the presence of spectator non-perturbative fields whose effects could be neglected. However, the examples discussed in section 6 illustrate the fact that many of the models identified in this sector are qualitatively distinct from those in the perturbative sector.

On the one hand, our results demonstrate that the class of string-inspired FN models is rich enough to accommodate all flavour physics. Although we have not studied the neutrino sector in this paper, the SM singlets present in these models can play the role of right handed neutrinos. We postpone the discussion of whether suitable masses can be generated in this sector to future work. On the other hand, compared to the size of the environments, the number of distinct charge patterns identified is very small, indicating that a careful engineering of U(1) symmetries, multiplet charges and FN singlet VEVs is needed. Our bottom-up approach complements previous top-down analyses of heterotic Standard Models [7, 11–13, 55–57], including the recent computation of quark masses performed in ref. [4]. Our results indicate certain limits on the viable charge patterns which may be translated into cohomological constraints, such as the requirement that the three **10** multiplets have distinct U(1) charges. Furthermore, assuming the existence of cohomology formulae such as those studied in refs. [58–61], these observations may be translated into topological constraints on the allowed bundles. We also note that our investigation of non-perturbative models opens up an interesting range of phenomenological possibilities which include non-perturbative effects. These have, so far, not been explored systematically in a top-down string context due to the inherent difficulties associated with the computation of non-perturbative effects and, in particular, worldsheet instantons.

Future work will attempt to bridge top-down and bottom-up approaches by identifying models from the databases in ref. [13, 57] that can accommodate the observed flavour parameters. We also hope to integrate the neutrino sector into the analysis, potentially applying more advanced search methods such as Reinforcement Learning (RL), to navigate the more complex landscape of models.

## Acknowledgments

AC is supported by a Royal Society Dorothy Hodgkin Fellowship, grant DHF/R1/231142. CSFT is supported by the Gould-Watson Scholarship. TRH is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>). LTYL is supported by the Oxford-Croucher Scholarship. AL acknowledges support by the STFC consolidated grant ST/X000761/1.

We would like to thank Jonathan Patterson for assistance with the Oxford Theoretical Physics computing cluster Hydra and Russell Jones for general technical support.

## A Naturalness conditions and charge-fixing constraints

In this appendix, we outline the main arguments, relying on naturalness considerations, that lead to the charge-fixing constraints presented in table 3 of section 3. We summarize the key findings as follows:

- Three distinct charges for the three  $\mathbf{10}$  multiplets are necessary.
- If  $\bar{\mathbf{5}}_{a,b}^H$  is the multiplet containing the down-Higgs, the spectrum cannot contain both a  $\mathbf{10}_a$  and a  $\mathbf{10}_b$  multiplet.

### A.1 Effects of order-one coefficients

In our string-inspired FN models, additional order-one coefficients (controlled, in the presumed underlying string compactification, by geometric complex structure and Kähler moduli) are required to resolve the rank-deficiency degeneracies present in the FN mass textures. More explicitly, the rank-deficiency degeneracies arise when the structure of the FN-generated mass matrix has insufficient rank, leading to either zero masses or degenerate masses. To fix these issues, extra order-one coefficients (non-zero numbers) can be introduced into the mass matrix to break the degeneracies and lift the rank, ensuring that all fermions acquire distinct, non-zero masses, and a realistic mass hierarchy is generated.

The order-one coefficients may introduce unwanted mass hierarchies via two mechanisms:

1. The coefficients have hierarchically different magnitudes. Such effects may arise in certain asymptotic regions of the string moduli space.
2. Small differences in order-one coefficients can also generate a hierarchy. We refer to this as the fine-tuning case.

In our analysis, we avoid the first mechanism by restricting the coefficients to be in the order-one range  $[0.5, 3.0]$ .

To illustrate the second mechanism, it is instructive to consider the matrix

$$B = \begin{pmatrix} 1 + \delta & 1 - \delta \\ 1 - \delta & 1 + \delta \end{pmatrix} \tag{A.1}$$

with eigenvalues

$$(\lambda_1, \lambda_2) = (2, 2\delta) \tag{A.2}$$

If  $\delta$  is a small parameter, a hierarchy between the two eigenvalues is generated solely through the small differences in the order-one coefficients. This second mechanism is particularly relevant for the up-type Yukawa matrices. Indeed, since the coupling  $\mathbf{5}^H \mathbf{10}_I \mathbf{10}_J$  is symmetric in the family indices, the up-type Yukawa matrix (including singlet VEV insertions) will be symmetric; as such, small differences in the order-one coefficients may introduce unnatural hierarchies, as in the illustrative example presented above. We prove the following result.

**Lemma A.1.** *To generate a natural FN structure, the three  $\mathbf{10}$  families must have distinct  $U(1)$  charges.*

*Proof.* The argument is straightforward. If two  $\mathbf{10}$  families have the same charge, the corresponding Yukawa submatrix would have four identical entries

$$\hat{\Lambda}_{[1,2]} \sim \begin{matrix} & \mathbf{10}_a^1 & \mathbf{10}_a^2 \\ \mathbf{10}_a^1 & \left( \begin{array}{cc} * & * \\ * & * \end{array} \right) \\ \mathbf{10}_a^2 & \end{matrix}. \tag{A.3}$$

The only way to generate a hierarchy in such a case is by tuning the order-one coefficients to lift a zero eigenvalue, which we aim to avoid.  $\square$

### A.2 Off-diagonal entries in the up-Yukawa matrix

In addition to the effects of order-one coefficients, we now discuss the effects of dominant off-diagonal entries in the up-Yukawa matrix.

**Lemma A.2.** *All off-diagonal terms in the up-type Yukawa matrix must be sub-dominant with respect to at least one diagonal entry in order to generate a natural hierarchy.*

*Proof.* This follows from the naturalness conditions of section 3.2. Without loss of generality, we can focus on the up-type Yukawa matrix structure

$$\hat{\Lambda}^u \sim \begin{pmatrix} \epsilon^a & \epsilon^b & \epsilon^c \\ \epsilon^b & \epsilon^d & 1 \\ \epsilon^c & 1 & \epsilon^e \end{pmatrix} \tag{A.4}$$

where  $a, b, c, d, e \in \mathbb{N}_{>0}$ .

If  $\lambda_1, \lambda_2, \lambda_3$  are the three eigenvalues of  $\hat{\Lambda}^u$ , it is straightforward to check that  $\text{tr}(\hat{\Lambda}^u) = \mathcal{O}(\epsilon^{\min(a,b,c)})$

$$\lambda_3 + \lambda_2 + \lambda_1 = \text{tr}(\hat{\Lambda}^u) = \mathcal{O}(\epsilon^{\min(a,d,e)}) \tag{A.5}$$

$$\lambda_3\lambda_2 + \lambda_3\lambda_1 + \lambda_2\lambda_1 = \frac{1}{2} \left( \text{tr}(\hat{\Lambda}^u)^2 - \text{tr}((\hat{\Lambda}^u)^2) \right) = \mathcal{O}(1), \tag{A.6}$$

which is clearly a contradiction, since  $a, d, e > 0$ .  $\square$

**Corollary A.1.** *If  $\mathbf{5}_{a,b}^H$  is the multiplet containing the up-Higgs, the spectrum cannot contain both the  $\mathbf{10}_a$  and  $\mathbf{10}_b$  multiplets.*

*Proof.* Suppose that the condition is not satisfied. Then the up-type Yukawa matrix contains the submatrix

$$\hat{\Lambda}_{a,b}^{(u)} \sim \begin{pmatrix} * & 1 \\ 1 & * \end{pmatrix}. \tag{A.7}$$

which is not allowed by Lemma A.2.  $\square$

## B Leading-order operators

In the first phase of exploration of the string FN landscape we identify viable Yukawa textures computed at the level of leading  $\mathcal{G}$ -invariant operators, where  $\mathcal{G} \cong \text{U}(1)^{f-1}$ .

For models containing only perturbative singlets, we use the following algorithm to compute the leading-order operators. Recall that the  $\mathcal{G}$ -charges of such singlets take the form

$$\mathbf{q}(\phi_{a,b}) = \mathbf{e}_a - \mathbf{e}_b, \tag{B.1}$$

where  $a = 1, \dots, f$ . The singlet  $\phi_{a,b}$  ‘moves’ charges like a flow, taking one unit of charge from the  $b$ -th charge entry into the  $a$ -th entry. This perspective leads to formulating the problem of finding leading-order operators as a Minimum Cost Maximum Flow (MCMF) problem.

**MCMF generalities.** The Minimum Cost Maximum Flow (MCMF) problem is a network optimization problem that seeks the least-cost distribution of a specified amount of flow in a network. The network is represented as a directed graph  $G = (V, E)$ , with:

1. A set of source vertices  $S \subset V$  and a set of sink vertices  $T \subset V$ .
2. A set of directed edges  $(u, v) \in E$  between vertices  $u, v \in V$ , each edge having a capacity  $c(u, v) > 0$  and a cost  $t(u, v) > 0$ .

Given a demand  $d_{s_i}$  for each source vertex  $s_i \in S$  and a demand  $d_{t_j}$  for each sink vertex  $t_j \in T$ , the objective is to determine a flow  $f(u, v)$  that minimizes the total cost of flow across all edges:

$$\sum_{(u,v) \in E} t(u, v) \cdot f(u, v) \tag{B.2}$$

subject to the following constraints:

- **Capacity constraint:** for each edge  $(u, v) \in E$ , the flow must not exceed the edge capacity:

$$f(u, v) \leq c(u, v).$$

- **Skew-symmetry constraint:** the flow must be skew-symmetric, meaning that for any pair of vertices  $u, v \in V$ :

$$f(u, v) = -f(v, u).$$

- **Flow conservation constraint:** for each vertex  $u \notin S \cup T$ , the total flow into the vertex must equal the total flow out, satisfying:

$$\sum_{v \in V} f(u, v) = 0 =: d_u.$$

For notational convenience, we write the net demand out of the vertex  $u$  as  $d_u$ . For any vertex that is not a source or a sink, we must therefore have  $d_u = 0$ .

- **Source consistency constraint:** for each source vertex  $s_i \in S$ , the net flow out must match the source demand:

$$\sum_{v \in V} f(s_i, v) = -d_{s_i}.$$

- **Sink consistency constraint:** for each sink vertex  $t_j \in T$ , the net flow into the sink must meet the sink demand:

$$\sum_{v \in V} f(v, t_j) = d_{t_j}.$$

**Implementation.** The problem of identifying leading-order operators can be translated into an MCMF problem as follows. First, we define the two charge vectors

$$\mathbf{q}^u = \mathbf{q}(H^u Q^I u^J) = \mathbf{q}(\mathbf{10}^I) + \mathbf{q}(\mathbf{10}^J) + \mathbf{q}(\mathbf{5}^H) \quad (\text{B.3})$$

$$\mathbf{q}^d = \mathbf{q}(H^d Q^I d^J) = \mathbf{q}(\mathbf{10}^I) + \mathbf{q}(\bar{\mathbf{5}}^J) + \mathbf{q}(\bar{\mathbf{5}}^H) \quad (\text{B.4})$$

and discuss them separately. Below we focus on  $\mathbf{q}^u$ , however, the other case is entirely analogous.

Since  $\mathcal{G}$ -charge vectors have  $f$  entries, the corresponding set  $V$  contains  $f$  vertices, labelled as  $1, \dots, f$ . The set  $S$  consists of the positive entries of  $\mathbf{q}^u$ , while the set of negative entries forms the set of sinks  $T$ . Vertices  $a$  and  $b$  in  $V$  are connected by the edge  $b \rightarrow a$  only if the spectrum contains the perturbative singlet  $\phi_{a,b}$ . The demand vector is  $\mathbf{d} = -\mathbf{q}^u$ . Furthermore, the cost associated with the edge  $b \rightarrow a$  is  $t(b, a) = k$  if  $\langle \phi_{a,b} \rangle = \epsilon^k$ . The capacity  $c(b, a)$  corresponds to the maximal number of allowed  $\phi_{a,b}$ -insertions. Finally, the aim is to determine the optimal flow function, with the understanding that  $f(b, a) = n$  corresponds to having  $n$ -insertions of  $\phi_{a,b}$  in the up Yukawa coupling. If the associated MCMF problem returns no solutions, this implies that there are no possible operators that match the SM operator charges up to the considered maximum number of insertions specified by edge capacity.

**Example.** We now illustrate the method with model corresponding to  $\mathbf{n} = (1, 1, 1, 1, 1)$ . Suppose we have the up-Yukawa term  $\mathbf{10}_1 \mathbf{10}_1 \mathbf{5}_{4,5}^H$  and the perturbative singlets  $\phi_{4,1}$ ,  $\phi_{5,1}$  and  $\phi_{5,4}$  with

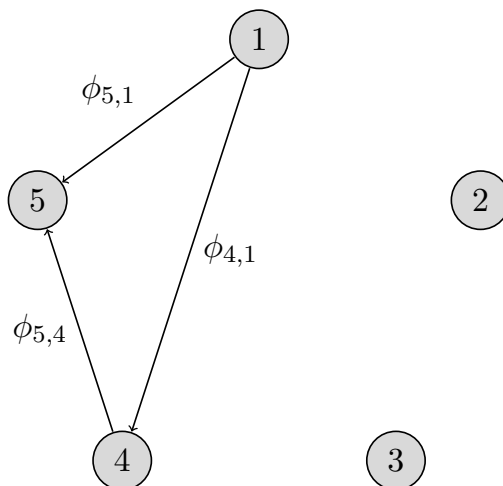
$$\langle \phi_{4,1} \rangle \sim \epsilon^1, \quad \langle \phi_{5,1} \rangle \sim \epsilon^4, \quad \langle \phi_{5,4} \rangle \sim \epsilon^2.$$

The associated graph in the MCMF problem has five vertices, with edges  $1 \rightarrow 4$ ,  $1 \rightarrow 5$  and  $4 \rightarrow 5$  having costs  $t(1, 4) = 1$ ,  $t(1, 5) = 4$  and  $t(4, 5) = 2$ , respectively. The demand vector is  $\mathbf{d} = -\mathbf{q}(\mathbf{10}_1 \mathbf{10}_1 \mathbf{5}_{4,5}^H) = (-2, 0, 0, 1, 1)$ , indicating that vertex 1 is the sink with  $d_1 = -2$  and vertices 4 and 5 are the sinks with  $d_4 = d_5 = 1$ . This is illustrated in figure 4.

The solution to the associated MCMF problem is  $f(1, 4) = 2$ ,  $f(1, 5) = 0$  and  $f(4, 5) = 1$ , such that

$$\mathbf{q}(\phi_{4,1}^2 \phi_{5,4}) = -\mathbf{q}(\mathbf{10}_1 \mathbf{10}_1 \mathbf{5}_{4,5}^H). \quad (\text{B.5})$$

The algorithm for solving a MCMF problem is readily realised in Mathematica. We have also implemented the algorithm in C, achieving a speedup by a factor of 100 in comparison



**Figure 4.** Graph for the MCMF problem associated with finding the leading-order operator for the up-Yukawa coupling  $\mathbf{10}_1 \mathbf{10}_1 \mathbf{5}_{4,5}^H$  with three perturbative singlets in the spectrum  $\phi_{4,1}, \phi_{5,1}, \phi_{5,4}$  and  $\mathbf{n} = (1, 1, 1, 1, 1)$ .

with the approach of listing all  $\mathcal{G}$ -invariant operators up to the given order and retaining the leading term. The algorithm complexity scales linearly with the number  $N_\phi$  of singlets.

For models including non-perturbative singlets we adapt the above algorithm as follows. If  $\{\Phi^{(i)}\}$  is the collection of non-perturbative singlets in the spectrum, we start by listing all possible combinations

$$(\Phi^{(1)})^{i_1} (\Phi^{(2)})^{i_2} \dots (\Phi^{(N_\Phi)})^{i_{N_\Phi}}, \quad 0 \leq i_1, \dots, i_{N_\Phi} \leq m, \quad (\text{B.6})$$

for a fixed positive integer  $m$ . We then apply the above algorithm to each of these combinations of non-perturbative fields multiplied into the Yukawa coupling in question. The leading-order operator is then picked from the available solutions. This is computationally-viable provided that  $m \leq 3$ , which is what we have done in our searches.

### C The fitness function

The fitness function (4.12) of a state  $b \in E$  in the GA environment consists of several contributions:

$$f(b) = f_{\text{quark}}(b) + f_{\text{texture}}(b) + f_{\text{fine-tuning}}(b) + f_{\text{split type}}(b) + f_{\text{anomaly}}(b), \quad (\text{C.1})$$

We discuss each of these contributions in turn.

**Masses and mixings.** The main purpose of the fitness function is to measure how accurately the quark masses and mixings are reproduced. In the GA search, all Yukawa couplings are multiplied by arbitrary  $\mathcal{O}(1)$  coefficients in order to break the rank-degeneracy of the Yukawa matrices. Since in this first phase the aim is to reproduce only qualitatively the observed hierarchies, we compute the logarithmic deviation of the computed quark masses and mixings from their corresponding experimental values at the  $M_Z$  scale, setting:

$$f_{\text{quark}} = -2 \sum_{\gamma} \min \left( \left| \log \left( \frac{\rho_{\gamma}}{\rho_{\gamma,0}} \right) \right|, 10 \right). \quad (\text{C.2})$$

Here,  $\rho_\gamma \in \{m^u, m^d, m^c, m^s, m^t, m^b, |(V_{CKM})_{ij}|\}$  are the quark masses and mixings computed for the model,  $\rho_{\gamma,0}$  are the corresponding experimental values and we have set a penalty bound of  $-10$ . The factor of 2 means that two units of fitness correspond to an order of magnitude deviation away from any measured flavour parameter.

**Texture contributions.** Following the discussion in section 3, the Yukawa textures must be non-fine-tuned and the Higgs VEV must be fixed to its measured value at  $M_Z$  scale. The texture contribution to the fitness function includes the following four terms:

$$f_{\text{texture}}(b) = -5 \left[ c_{\text{pow}} f_{\text{pow}} + c_\epsilon f_\epsilon + c_{O(1)} f_{O(1)} + c_{\text{range}} f_{\text{range}} \right], \quad (\text{C.3})$$

where the weights are set to  $(c_{\text{pow}}, c_\epsilon, c_{O(1)}, c_{\text{range}}) = (1, 0.25, 1, 1)$ . The overall factor of 5 is to make this contribution comparable with the other contributions in the fitness function in eq. (C.1). The various contributions are:

- (i)  $f_{\text{pow}}$  accounts for the naturalness conditions described by eq. (3.10).  $f_{\text{pow}}$  is zero when the naturalness conditions are satisfied, otherwise we take  $f_{\text{pow}}$  to be the difference of the powers:

$$f_{\text{pow}} = \sum_{i=1,2} \left[ \min(2r_{i+1}^{(u)} - r_i^{(u)}, 0) + \min(2r_{i+1}^{(d)} - r_i^{(d)}, 0) \right] \quad (\text{C.4})$$

- (ii)  $f_\epsilon$  is the geometric standard deviation of the  $(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)$ -distribution from which the optimised scale  $\epsilon$  is obtained via eq. (4.9),

$$f_\epsilon = \exp \left[ \frac{1}{4} \sum_{i=1}^4 \log \left( \frac{\epsilon_i^2}{\epsilon^2} \right) \right]^{1/2}. \quad (\text{C.5})$$

- (iii)  $f_{O(1)}$  measures whether the coefficient  $c_t$  calculated from eq. (4.11) is of order one and is real. If  $c_t$  is real but outside of the range  $[0.5, 3]$ , we set  $f_{O(1)}$  to be the logarithm of the ratio between the bound and  $c_t$  as follows:

$$f_{O(1)} = \begin{cases} 0 & \text{if } 0.5 < c_t < 3 \\ \log(c_t/3) & \text{if } 3 < c_t \\ \log(0.5/c_t) & \text{if } c_t < 0.5 \end{cases}. \quad (\text{C.6})$$

If  $c_t$  is imaginary, the correct Higgs VEV cannot be obtained, in which case  $f_{O(1)} = -c_t^2$ .

- (iv)  $f_{\text{range}}$  accounts for the ranges of the singlet VEVs. We typically expect the singlets to lie between the following ranges:  $\langle \phi \rangle \in [0.005, 0.5]$  and  $\langle \Phi \rangle = [0.0005, 0.05]$ . However, since we calculate the  $\epsilon$ -scale after specifying the VEV powers, the actual VEV of the singlets may be outside the range as specified. We therefore penalise models that give a singlet VEV outside the range by the following fitness contribution,

$$f_{O(1)} = \sum_\alpha f_{O(1),\phi^{(\alpha)}} + \sum_i f_{O(1),\Phi^{(i)}} \quad (\text{C.7})$$

where the individual contributions are,

$$f_{O(1),\phi^{(\alpha)}} = \begin{cases} 0 & \text{if } 0.005 < \langle \phi^{(\alpha)} \rangle < 0.5 \\ \log(\langle \phi^{(\alpha)} \rangle / 0.5) & \text{if } 0.5 < \langle \phi^{(\alpha)} \rangle \\ \log(0.005 / \langle \phi^{(\alpha)} \rangle) & \text{if } \langle \phi^{(\alpha)} \rangle < 0.005 \end{cases}, \quad (\text{C.8})$$

and

$$f_{O(1),\Phi^{(i)}} = \begin{cases} 0 & \text{if } 0.0005 < \langle \Phi^{(i)} \rangle < 0.05 \\ \log(\langle \Phi^{(i)} \rangle / 0.05) & \text{if } 0.05 < \langle \Phi^{(i)} \rangle \\ \log(0.0005 / \langle \Phi^{(i)} \rangle) & \text{if } \langle \Phi^{(i)} \rangle < 0.0005 \end{cases}. \quad (\text{C.9})$$

**Order-one tuning contributions.** The order-one coefficient fine-tuning contributions to quark mixing are not considered in the fine-tuning analysis of Yukawa textures. This however should be accounted for, as by the fitness contribution  $f_{\text{fine-tuning}}$ . Following ref. [62], we have taken the fine-tuning fitness contribution  $f_{\text{fine-tuning}}$  to be the logarithmic numerical differentiation of the SM quantities with respect to the order-one coefficients:

$$f_{\text{fine-tuning}} = -\frac{1}{10} \sum_i \max_j \left| \frac{\delta \log \rho_i}{\delta \log o_j} \right| \quad (\text{C.10})$$

where  $\rho_i \in \{\langle H \rangle, m_i, |V_{ij}|\}$  are the set of Standard Model quantities. The overall factor of 1/10 is to allow this contribution to be comparable to the other contributions in eq. (C.1).

**Contributions from the charge constraint.** Whenever the spectrum contains charge patterns that are forbidden, such as having  $e_a = e_b$  for the multiplets  $\bar{\mathbf{5}}$  in cases where the partition vector  $\mathbf{n} = (1, 1, 1, 1, 1)$ , the contribution  $f_{\text{split type}}$  is evaluated as:

$$f_{\text{split type}} = -20 \times N_{\text{violations}} \quad (\text{C.11})$$

where  $N_{\text{violations}}$  is the number of forbidden charge patterns in the spectrum. This contribution choice is shown to have better convergence effects in the GA search than simply setting  $f_{\text{split type}} = -20$  whenever such charge constraints are violated.

**Anomaly cancellation contributions.** The term  $f_{\text{anomaly}}$  accounts for the failure to satisfy the anomaly cancellation condition in eq. (2.10). To eliminate all models which fail to satisfy the anomalous cancellation condition from the viable model list,  $f_{\text{anomaly}} = 0$  if an integer solution of  $\beta_i$  can be found and  $f_{\text{anomaly}} \leq -20$  for models with no integer solutions of  $\beta_i$ . Similar to the charge constraint contribution, we introduce continuous-varying contribution for the cases with no integer solutions of  $\beta_i$  to achieve better convergence in the GA search. In particular, when  $\beta_i$  has no integer solutions but a solution  $\beta_i \in \mathbb{R}^{N_T}$  could be found, we set

$$f_{\text{anomaly}} = -20 \times \left[ \min \left( \sum_i [\text{round}(\beta_i) - \beta_i] - 1, 0 \right) + 1 \right] \quad (\text{C.12})$$

where round stands for rounding to the nearest integer. If no solutions can be found for eq. (2.10), then

$$f_{\text{anomaly}} = -20 \times \left[ \min_{\beta_i} \left( \|\mathcal{A}_a - \beta_i k_a^i\| - 1, 0 \right) + 1 \right] \quad (\text{C.13})$$

where the minimal  $\beta_i \in \mathbb{R}^{n_T}$  is the approximate solution to the original linear system found by the normal equation (using the notation  $k^i_a$  to indicate the matrix elements),

$$\beta_i = (k^T k)^j_i (k^T)^a_j \mathcal{A}_a \quad (\text{C.14})$$

in the method of ordinary least squares.

## D Crude renormalisation scheme

The MSSM one-loop RG equations [63] are given by:

$$\frac{d\Lambda^{u,d,e}}{dt} = \frac{1}{16\pi^2} \beta^{u,d,e}, \quad \frac{dg_a}{dt} = \frac{g_a^3}{16\pi^2} \beta_a \quad (\text{D.1})$$

where  $t = \log(M/M_0)$ ,  $M_0$  is the reference scale,  $(\beta_a) = (33/5, 1, -3)$  for U(1), SU(2) and SU(3), respectively, and

$$\beta^u = \left( 3 \text{tr}(\Lambda^u \Lambda^{u\dagger}) + 3\Lambda^u \Lambda^{u\dagger} + \Lambda^d \Lambda^{d\dagger} - \frac{16}{3}g_3^2 - 3g_2^2 - \frac{13}{15}g_1^2 \right) \Lambda^u \quad (\text{D.2})$$

$$\beta^d = \left( \text{tr}(3\Lambda^d \Lambda^{d\dagger} + \Lambda^e \Lambda^{e\dagger}) + 3\Lambda^d \Lambda^{d\dagger} + \Lambda_u \Lambda_u^\dagger - \frac{16}{3}g_3^2 - 3g_2^2 - \frac{7}{15}g_1^2 \right) \Lambda^d \quad (\text{D.3})$$

$$\beta^e = \left( \text{tr}(3\Lambda^d \Lambda^{d\dagger} + \Lambda^e \Lambda^{e\dagger}) + 3\Lambda^e \Lambda^{e\dagger} - 3g_2^2 - \frac{9}{5}g_1^2 \right) \Lambda^e \quad (\text{D.4})$$

We take the following radical approach. We neglect the effects of  $g_1$  and  $g_2$  (only retaining strong effects) and the effects of the (3,3)-entries of the Yukawa matrices, i.e.

$$\Lambda^u \simeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & y_t \end{pmatrix}, \quad \Lambda^d \simeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & y_b \end{pmatrix}, \quad \Lambda^e \simeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & y_\tau \end{pmatrix}. \quad (\text{D.5})$$

We insert these in the above beta functions. Integrating the  $g_3$  one-loop RG equation and redefining the variable  $t \mapsto t/16\pi^2$  gives

$$g_3^2 = \frac{g_{3,0}^2}{1 - 2g_{3,0}^2 \beta_3 t}, \quad (\text{D.6})$$

where  $t = 0$  corresponds to  $M = M_Z$ ,  $t \simeq 0.201$  corresponds to  $M = M_{\text{GUT}}$ ,  $\beta_3 = -3$ ,  $g_{3,0}^2 = 4\pi\alpha_3$  and  $\alpha(M_Z) \simeq 0.1155$ . The differential equations for the Yukawa matrices then become:

$$\frac{dy_t}{dt} = (6y_t^2 + y_b^2 - \frac{16}{3}g_3^2)y_t, \quad \frac{dy_b}{dt} = (6y_b^2 + y_t^2 + y_\tau^2 - \frac{16}{3}g_3^2)y_b, \quad \frac{dy_\tau}{dt} = (3y_b^2 + 4y_\tau^2)y_\tau \quad (\text{D.7})$$

$$\frac{d\Lambda_{3,i}^u}{dt} = (6y_t^2 + y_b^2 - \frac{16}{3}g_3^2)\Lambda_{3,i}^u, \quad \frac{d\Lambda_{3,i}^d}{dt} = (6y_b^2 + y_t^2 + y_\tau^2 - \frac{16}{3}g_3^2)\Lambda_{3,i}^d, \quad \frac{d\Lambda_{3,i}^e}{dt} = (3y_b^2 + 4y_\tau^2)\Lambda_{3,i}^e \quad (\text{D.8})$$

$$\frac{d\Lambda_{i,j}^u}{dt} = (3y_t^2 - \frac{16}{3}g_3^2)\Lambda_{i,j}^u, \quad \frac{d\Lambda_{i,j}^d}{dt} = (3y_b^2 + y_\tau^2 - \frac{16}{3}g_3^2)\Lambda_{i,j}^d, \quad \frac{d\Lambda_{i,j}^e}{dt} = y_\tau^2 \Lambda_{i,j}^e \quad (\text{D.9})$$

where we have denoted  $\Lambda_{3,3}^u = y_t$ ,  $\Lambda_{3,3}^d = y_b$  and  $\Lambda_{3,3}^e$ . The coupled differential equations in eq. (D.7) have no known closed solutions (a recursive solution exists in [64, 65]) but can be solved numerically. The other Yukawa couplings  $x = u, d, e$  can be solved using:

$$\frac{\Lambda_{m,n}^x}{(\Lambda_0)_{m,n}^x} = \exp \left[ \int_0^t dt \left( k_t y_t^2 + k_b y_b^2 + k_\tau y_\tau^2 - \frac{16}{3} g_3^2 \right) \right] \quad (\text{D.10})$$

where  $i, j \in \{1, 2\}$ ,  $x$  is the family index  $x = u, d, e$  for the up-, down- and electron-Yukawa couplings, respectively, and

$$(k_t, k_b, k_\tau) = \begin{cases} (6, 1, 0) & \text{if } x = u, (n = 3 \text{ and } m = i) \text{ or } (n = i \text{ and } m = 3) \\ (3, 0, 0) & \text{if } x = u, (m = i \text{ and } n = j) \\ (1, 6, 1) & \text{if } x = d, (n = 3 \text{ and } m = i) \text{ or } (n = i \text{ and } m = 3) \\ (0, 3, 1) & \text{if } x = d, (m = i \text{ and } n = j) \\ (0, 3, 4) & \text{if } x = e, (n = 3 \text{ and } m = i) \text{ or } (n = i \text{ and } m = 3) \\ (0, 0, 1) & \text{if } x = e, (m = i \text{ and } n = j) \end{cases} . \quad (\text{D.11})$$

**Data Availability Statement.** This article has associated data in a data repository.

**Code Availability Statement.** This article has associated code in a code repository.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] L.E. Ibanez and A.M. Uranga, *String theory and particle physics: an introduction to string phenomenology*, Cambridge University Press (2012) [[INSPIRE](#)].
- [2] F. Marchesano, G. Shiu and T. Weigand, *The Standard Model from String Theory: What Have We Learned?*, *Ann. Rev. Nucl. Part. Sci.* **74** (2024) 113 [[arXiv:2401.01939](#)] [[INSPIRE](#)].
- [3] M. Cvetič, J. Halverson, G. Shiu and W. Taylor, *Snowmass White Paper: String Theory and Particle Physics*, [arXiv:2204.01742](#) [[INSPIRE](#)].
- [4] A. Constantin et al., *Computation of quark masses from string theory*, *Nucl. Phys. B* **1010** (2025) 116778 [[arXiv:2402.01615](#)] [[INSPIRE](#)].
- [5] G. Butbaia et al., *Physical Yukawa Couplings in Heterotic String Compactifications*, [arXiv:2401.15078](#) [[DOI:10.4310/ATMP.241119041341](#)] [[INSPIRE](#)].
- [6] P. Berglund et al., *Precision string phenomenology*, *Phys. Rev. D* **111** (2025) 086007 [[arXiv:2407.13836](#)] [[INSPIRE](#)].
- [7] E.I. Buchbinder, A. Constantin and A. Lukas, *Non-generic Couplings in Supersymmetric Standard Models*, *Phys. Lett. B* **748** (2015) 251 [[arXiv:1409.2412](#)] [[INSPIRE](#)].
- [8] J. Distler and B.R. Greene, *Aspects of (2,0) String Compactifications*, *Nucl. Phys. B* **304** (1988) 1 [[INSPIRE](#)].

- [9] R. Blumenhagen, S. Moster and T. Weigand, *Heterotic GUT and standard model vacua from simply connected Calabi-Yau manifolds*, *Nucl. Phys. B* **751** (2006) 186 [[hep-th/0603015](#)] [[INSPIRE](#)].
- [10] R. Blumenhagen, S. Moster, R. Reinbacher and T. Weigand, *Massless Spectra of Three Generation  $U(N)$  Heterotic String Vacua*, *JHEP* **05** (2007) 041 [[hep-th/0612039](#)] [[INSPIRE](#)].
- [11] L.B. Anderson, J. Gray, A. Lukas and E. Palti, *Heterotic Line Bundle Standard Models*, *JHEP* **06** (2012) 113 [[arXiv:1202.1757](#)] [[INSPIRE](#)].
- [12] L.B. Anderson, J. Gray, A. Lukas and E. Palti, *Two Hundred Heterotic Standard Models on Smooth Calabi-Yau Threefolds*, *Phys. Rev. D* **84** (2011) 106005 [[arXiv:1106.4804](#)] [[INSPIRE](#)].
- [13] L.B. Anderson et al., *A Comprehensive Scan for Heterotic  $SU(5)$  GUT models*, *JHEP* **01** (2014) 047 [[arXiv:1307.4787](#)] [[INSPIRE](#)].
- [14] A. Font and L.E. Ibanez, *Yukawa Structure from  $U(1)$  Fluxes in F-theory Grand Unification*, *JHEP* **02** (2009) 016 [[arXiv:0811.2157](#)] [[INSPIRE](#)].
- [15] E. Dudas and E. Palti, *Froggatt-Nielsen models from  $E(8)$  in F-theory GUTs*, *JHEP* **01** (2010) 127 [[arXiv:0912.0853](#)] [[INSPIRE](#)].
- [16] S. Krippendorff, S. Schafer-Nameki and J.-M. Wong, *Froggatt-Nielsen meets Mordell-Weil: a Phenomenological Survey of Global F-theory GUTs with  $U(1)$ s*, *JHEP* **11** (2015) 008 [[arXiv:1507.05961](#)] [[INSPIRE](#)].
- [17] C.D. Froggatt and H.B. Nielsen, *Hierarchy of Quark Masses, Cabibbo Angles and CP Violation*, *Nucl. Phys. B* **147** (1979) 277 [[INSPIRE](#)].
- [18] M. Leurer, Y. Nir and N. Seiberg, *Mass matrix models*, *Nucl. Phys. B* **398** (1993) 319 [[hep-ph/9212278](#)] [[INSPIRE](#)].
- [19] M. Leurer, Y. Nir and N. Seiberg, *Mass matrix models: the Sequel*, *Nucl. Phys. B* **420** (1994) 468 [[hep-ph/9310320](#)] [[INSPIRE](#)].
- [20] A. Davidson, M. Koca and K.C. Wali,  *$U(1)$  as the Minimal Horizontal Gauge Symmetry*, *Phys. Rev. Lett.* **43** (1979) 92 [[INSPIRE](#)].
- [21] A. Davidson and K.C. Wali, *Minimal flavor unification via multigenerational Peccei-Quinn symmetry*, *Phys. Rev. Lett.* **48** (1982) 11 [[INSPIRE](#)].
- [22] A. Davidson, V.P. Nair and K.C. Wali, *Peccei-Quinn Symmetry as Flavor Symmetry and Grand Unification*, *Phys. Rev. D* **29** (1984) 1504 [[INSPIRE](#)].
- [23] E. Dudas, S. Pokorski and C.A. Savoy, *Yukawa matrices from a spontaneously broken Abelian symmetry*, *Phys. Lett. B* **356** (1995) 45 [[hep-ph/9504292](#)] [[INSPIRE](#)].
- [24] E. Dudas, C. Grojean, S. Pokorski and C.A. Savoy, *Abelian flavor symmetries in supersymmetric models*, *Nucl. Phys. B* **481** (1996) 85 [[hep-ph/9606383](#)] [[INSPIRE](#)].
- [25] L.E. Ibanez and G.G. Ross, *Fermion masses and mixing angles from gauge symmetries*, *Phys. Lett. B* **332** (1994) 100 [[hep-ph/9403338](#)] [[INSPIRE](#)].
- [26] M.B. Green, J.H. Schwarz and E. Witten, *Superstring theory. Vol. 2 Loop amplitudes, anomalies and phenomenology*, Cambridge University Press (2012).
- [27] T.R. Harvey and A. Lukas, *Quark Mass Models and Reinforcement Learning*, *JHEP* **08** (2021) 161 [[arXiv:2103.04759](#)] [[INSPIRE](#)].

- [28] F. Feruglio, *Are neutrino masses modular forms?*, in *From My Vast Repertoire...: Guido Altarelli's Legacy*, A. Levy et al. eds., World Scientific (2019), p. 227–266 [[DOI:10.1142/9789813238053\\_0012](#)] [[arXiv:1706.08749](#)] [[INSPIRE](#)].
- [29] F.J. de Anda, S.F. King and E. Perdomo, *SU(5) grand unified theory with  $A_4$  modular symmetry*, *Phys. Rev. D* **101** (2020) 015028 [[arXiv:1812.05620](#)] [[INSPIRE](#)].
- [30] F. Feruglio and A. Romanino, *Lepton flavor symmetries*, *Rev. Mod. Phys.* **93** (2021) 015007 [[arXiv:1912.06028](#)] [[INSPIRE](#)].
- [31] T. Kobayashi and M. Tanimoto, *Modular flavor symmetric models*, *Int. J. Mod. Phys. A* **39** (2024) 2441012 [[arXiv:2307.03384](#)] [[INSPIRE](#)].
- [32] A. Baur et al., *Top-down anatomy of flavor symmetry breakdown*, *Phys. Rev. D* **105** (2022) 055018 [[arXiv:2112.06940](#)] [[INSPIRE](#)].
- [33] H.P. Nilles and S. Ramos-Sanchez, *The flavor puzzle: textures and symmetries*, *Int. J. Mod. Phys. A* **39** (2024) 2441003 [[arXiv:2308.14810](#)] [[INSPIRE](#)].
- [34] K. Ishiguro, T. Kobayashi, S. Nishimura and H. Otsuka, *Modular forms and hierarchical Yukawa couplings in heterotic Calabi-Yau compactifications*, *JHEP* **08** (2024) 088 [[arXiv:2402.13563](#)] [[INSPIRE](#)].
- [35] E.I. Buchbinder, A. Constantin, J. Gray and A. Lukas, *Yukawa Unification in Heterotic String Theory*, *Phys. Rev. D* **94** (2016) 046005 [[arXiv:1606.04032](#)] [[INSPIRE](#)].
- [36] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001 [[INSPIRE](#)].
- [37] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed., Addison-Wesley Longman Publishing Co., Inc., U.S.A. (1989).
- [38] F. Ruehle, *Data science applications to string theory*, *Phys. Rept.* **839** (2020) 1 [[INSPIRE](#)].
- [39] S. Katoch, S.S. Chauhan and V. Kumar, *A review on genetic algorithm: past, present, and future*, *Multimedia Tools and Applications* **80** (2020) 8091.
- [40] S.A. Abel, A. Constantin, T.R. Harvey and A. Lukas, *Cosmic Inflation and Genetic Algorithms*, *Fortsch. Phys.* **71** (2023) 2200161 [[arXiv:2208.13804](#)] [[INSPIRE](#)].
- [41] A. Kamberkar, S. Nesseris and L. Pinol, *Machine learning cosmic inflation*, *Phys. Rev. D* **108** (2023) 043509 [[arXiv:2211.14142](#)] [[INSPIRE](#)].
- [42] B.C. Allanach, D. Grellscheid and F. Quevedo, *Genetic algorithms and experimental discrimination of SUSY models*, *JHEP* **07** (2004) 069 [[hep-ph/0406277](#)] [[INSPIRE](#)].
- [43] Y. Akrami et al., *A Profile Likelihood Analysis of the Constrained MSSM with Genetic Algorithms*, *JHEP* **04** (2010) 057 [[arXiv:0910.3950](#)] [[INSPIRE](#)].
- [44] S. Abel, D.G. Cerdeño and S. Robles, *The Power of Genetic Algorithms: what remains of the  $p$ MSSM?*, [arXiv:1805.03615](#) [[INSPIRE](#)].
- [45] S. Abel and J. Rizos, *Genetic Algorithms and the Search for Viable String Vacua*, *JHEP* **08** (2014) 010 [[arXiv:1404.7359](#)] [[INSPIRE](#)].
- [46] J. Halverson, B. Nelson and F. Ruehle, *Branes with Brains: Exploring String Vacua with Deep Reinforcement Learning*, *JHEP* **06** (2019) 003 [[arXiv:1903.11616](#)] [[INSPIRE](#)].
- [47] S. Abel, A. Constantin, T.R. Harvey and A. Lukas, *Evolving Heterotic Gauge Backgrounds: Genetic Algorithms versus Reinforcement Learning*, *Fortsch. Phys.* **70** (2022) 2200034 [[arXiv:2110.14029](#)] [[INSPIRE](#)].

- [48] S. Abel, A. Constantin, T.R. Harvey and A. Lukas, *String Model Building, Reinforcement Learning and Genetic Algorithms*, in the proceedings of the *Nankai Symposium on Mathematical Dialogues: In celebration of S.S.Chern's 110th anniversary*, , China, August 02–13 (2021) [[arXiv:2111.07333](#)] [[INSPIRE](#)].
- [49] A. Cole, S. Krippendorff, A. Schachner and G. Shiu, *Probing the Structure of String Theory Vacua with Genetic Algorithms and Reinforcement Learning*, in the proceedings of the *35th Conference on Neural Information Processing Systems*, Online Conference, Canada, December 06–14 (2021) [[arXiv:2111.11466](#)] [[INSPIRE](#)].
- [50] S.A. Abel et al., *Decoding Nature with Nature's Tools: Heterotic Line Bundle Models of Particle Physics with Genetic Algorithms and Quantum Annealing*, *Fortsch. Phys.* **72** (2024) 2300260 [[arXiv:2306.03147](#)] [[INSPIRE](#)].
- [51] CMS collaboration, *Searches for additional Higgs bosons and for vector leptokuarks in  $\tau\tau$  final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **07** (2023) 073 [[arXiv:2208.02717](#)] [[INSPIRE](#)].
- [52] CMS collaboration, *Searches for Higgs boson production through decays of heavy resonances*, *Phys. Rept.* **1115** (2025) 368 [[arXiv:2403.16926](#)] [[INSPIRE](#)].
- [53] S. Abel, A. Constantin, T.R. Harvey and A. Lukas, *GAMathematica: Mathematica package for Genetic Algorithms*, <https://github.com/harveyThomas4692/GAMathematica>.
- [54] S. Abel, A. Constantin, T.R. Harvey, A. Lukas and L.A. Nutricati, *GA-C: a realisation of genetic algorithms in C*, <https://github.com/harveyThomas4692/GA-C>.
- [55] E.I. Buchbinder, A. Constantin and A. Lukas, *The Moduli Space of Heterotic Line Bundle Models: a Case Study for the Tetra-Quadric*, *JHEP* **03** (2014) 025 [[arXiv:1311.1941](#)] [[INSPIRE](#)].
- [56] E.I. Buchbinder, A. Constantin and A. Lukas, *A heterotic standard model with  $B - L$  symmetry and a stable proton*, *JHEP* **06** (2014) 100 [[arXiv:1404.2767](#)] [[INSPIRE](#)].
- [57] A. Constantin, Y.-H. He and A. Lukas, *Counting String Theory Standard Models*, *Phys. Lett. B* **792** (2019) 258 [[arXiv:1810.00444](#)] [[INSPIRE](#)].
- [58] A. Constantin and A. Lukas, *Formulae for Line Bundle Cohomology on Calabi-Yau Threefolds*, *Fortsch. Phys.* **67** (2019) 1900084 [[arXiv:1808.09992](#)] [[INSPIRE](#)].
- [59] D. Klaewer and L. Schlechter, *Machine Learning Line Bundle Cohomologies of Hypersurfaces in Toric Varieties*, *Phys. Lett. B* **789** (2019) 438 [[arXiv:1809.02547](#)] [[INSPIRE](#)].
- [60] C.R. Brodie, A. Constantin, R. Deen and A. Lukas, *Machine Learning Line Bundle Cohomology*, *Fortsch. Phys.* **68** (2020) 1900087 [[arXiv:1906.08730](#)] [[INSPIRE](#)].
- [61] M. Larfors and R. Schneider, *Line bundle cohomologies on CICYs with Picard number two*, *Fortsch. Phys.* **67** (2019) 1900083 [[arXiv:1906.00392](#)] [[INSPIRE](#)].
- [62] G.G. Ross, K. Schmidt-Hoberg and F. Staub, *Revisiting fine-tuning in the MSSM*, *JHEP* **03** (2017) 021 [[arXiv:1701.03480](#)] [[INSPIRE](#)].
- [63] D.J. Castano, E.J. Piard and P. Ramond, *Renormalization group study of the Standard Model and its extensions. 2. The Minimal supersymmetric Standard Model*, *Phys. Rev. D* **49** (1994) 4882 [[hep-ph/9308335](#)] [[INSPIRE](#)].
- [64] G. Auberson and G. Moutaka, *A convergent scheme for one loop evolutions of the Yukawa couplings in the MSSM*, *Eur. Phys. J. C* **12** (2000) 331 [[hep-ph/9907204](#)] [[INSPIRE](#)].
- [65] MSSM WORKING GROUP collaboration, *The minimal supersymmetric standard model: group summary report*, in the proceedings of the *GDR (Groupement De Recherche) - Supersymetrie*, Montpellier, France, April 15–17 (1998) [[hep-ph/9901246](#)] [[INSPIRE](#)].