Navigating the String Landscape with Machine Learning Techniques

Thomas Rhys Harvey

Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2024

Abstract

This thesis explores the possibility of using heuristic search algorithms from data science, namely genetic algorithms and reinforcement learning, to navigate the string landscape to uncover phenomenologically interesting constructions. Specifically, we apply these algorithms to construct holomorphic slope-stable vector bundles over Calabi-Yau three-folds. These vector bundles lead to the particle spectrum of the minimally supersymmetric standard model (plus uncharged moduli), via compactifications of the $E_8 \times E_8$ heterotic string. We explore two types of vector bundles: sums of line bundles, which have been extensively explored in existing literature and thus serve as a benchmark for the effectiveness of the algorithms, and monad bundles, where only one quasi-realistic model was previously known. For both environments, these search algorithms were able to discover many models, while exploring as little as 10^{-19} of the total space. As an example of these methods in a simpler context, we also explored Froggatt-Nielsen models of quark masses.

Navigating the String Landscape with Machine Learning Techniques

Thomas Rhys Harvey
Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2024

This thesis is dedicated to

Ronald Harvey

(1928-2023)

Acknowledgements

I extend my deepest gratitude to Andre Lukas, whose unwavering support and guidance have been invaluable throughout my DPhil. I consider myself exceptionally fortunate to have had Andre as a supervisor. I am also indebted the many brilliant collaborators I have had the pleasure of working with. Namely: Steve Abel, Aditi Chandra, Andrei Constantin, Kit Fraser-Taliente, Luca Nutricati, Burt Ovrut and Sebastian von Hausegger.

Over the last four years, I have had the delightful opportunity to discuss physics with the many friends I have made in the department. Among the faculty, I am particularly grateful for the time and patience that both Prateek Agrawal and John Wheater have had for myself and the other graduate students. While among fellow students, I have especially enjoyed my conversations with Kit Fraser-Taliente, Giulio Gambuti, Liam Gladden, Lucas Leung and Arthur Platschorre. Of these, I especially thank Kit (once again) for frequently humouring my half-formed thoughts.

Beyond the theoretical physics department, Oxford has introduced me to many wonderful people whom I'm grateful to count as friends. In particular I would like to mention Tobias Géron, Gustavo Pedroso and Jakob Wenninger for making my time at Oxford as enjoyable as it has been.

Finally, I would like to thank my family and fiancée, Sarah, for their unwavering love and support. Not just over the course of my DPhil, but throughout the long journey up to this point.

My graduate studies were supported by the Science and Technology Facilities Council, and the donations to the Rudolf Peierls Centre for Theoretical Physics from Philip Hudson.

Abstract

This thesis explores the possibility of using heuristic search algorithms from data science, namely genetic algorithms and reinforcement learning, to navigate the string landscape to uncover phenomenologically interesting constructions. Specifically, we apply these algorithms to construct holomorphic slope-stable vector bundles over Calabi-Yau three-folds. These vector bundles lead to the particle spectrum of the minimally supersymmetric standard model (plus uncharged moduli), via compactifications of the $E_8 \times E_8$ heterotic string. We explore two types of vector bundles: sums of line bundles, which have been extensively explored in existing literature and thus serve as a benchmark for the effectiveness of the algorithms, and monad bundles, where only one quasi-realistic model was previously known. For both environments, these search algorithms were able to discover many models, while exploring as little as 10^{-19} of the total space. As an example of these methods in a simpler context, we also explored Froggatt-Nielsen models of quark masses.

Contents

Appendices

What I've said so far doesn't guarantee anything about the existence of a limit where Newton's constant is actually infinity, but nevertheless that limit may exist. If the limit exists the limiting theory is a theory in 11 dimensions. It'll be a supersymmetric relativistic theory in 11 dimensions, with 11-dimensional supergravity as its low-energy limit, and depending on no dimensionless parameters.

Does it exist? I haven't the faintest idea...

All I'm saying is, while in the past I would have bet against it I shall place no such bet today.

— Edward Witten [1]

1

Introduction and Background

Contents

1.1 Motivation

String theory is a theory of quantum gravity which replaces elementary particles with fundamental strings. It is beyond any doubt that string theory has had a phenomenal impact on the study of theoretical high energy physics – from the discovery of holography [2, 3], to improving our understanding of the quantum nature of black holes [4], alongside discoveries in mathematics such as mirror symmetry [5]. Despite this success, one major question has remained unanswered: Does string theory describe our world?

At a first glance, string theory appears to have all the necessary components for a "theory of everything". Namely it contains quantum gravity, and non-abelian gauge theories. At the same time, there are at least two immediate problems. The first of these problems is that string theory requires the space-time dimension of our universe to be ten, while the second is that string theory usually requires a yet unobserved symmetry called supersymmetry with 16 or 32 supercharges (the generators of supersymmetry).

If one were to try and search for the standard model of particle physics within superstring theory, one would begin by looking at its features in broad strokes: it is a four dimensional, chiral gauge theory with three generations of matter and with gauge group $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. Immediately, the requirement of chirality enforces us to have no more than 4 unbroken real supercharges, which we refer to as $\mathcal{N} = 1$ supersymmetry. Similarly, to match with the observed four dimensions, we need to choose a particular compact manifold for the extra six dimensions. Different choices for this compact manifold will lead to different four dimensional physics.

We focus on the case of $\mathcal{N} = 1$ supersymmetry, which will need to be spontaneously broken above the electroweak scale. We make this simplification as the relevant equations for supersymmetric solutions to string theory are only first order in derivatives, unlike the full second order equations of motion. At the same time, supersymmetric solutions offer us the best controlled and understood regions of the string landscape.

To engineer such a four dimensional gauge theory from string theory, at least in the geometric regime, there are three typical routes:

- $E_8 \times E_8$ heterotic string theory on smooth Calabi-Yau manifolds with holomorphic and slope-stable vector bundles [6], or the strongly coupled version via the Horava-Witten construction [7–9]. This will be the focus of this thesis, and will be elaborated on in Sec. 1.3.

- Type IIB string theory on Calabi-Yau orientifolds with D-branes [10, 11], or at strong coupling F-theory on singular and elliptically-fibred Calabi-Yau 4-folds [12].
- Type IIA string theory on Calabi-Yau orientifolds with D-branes [10, 11], or at strong coupling 11 Dimensional supergravity on singular G2 manifolds [13].

All of these options result in vast landscapes. Focusing on perturbative IIB sector, early estimates suggest $\mathcal{O}(10^{500})$ vacua by enumerating the possible choices of flux [14]. Later work extended the previous number to $\mathcal{O}(10^{272,000})$ by considering the strong-coupling regime with F-Theory [15]. Alternatively, focusing on just the number of Calabi-Yau threefolds, recent estimates indicate there could be as many as $\mathcal{O}(10^{400})$ in the largest database of examples [16–18].

One may worry, that with such a gigantic "string landscape", identifying those few constructions of phenomenological interest may seem impossible. However, in recent years, developments in computer science have allowed for the navigation of other large environments. For example, the AlphaGo project was able to train a neural network to play the game of Go (a game with $\mathcal{O}(10^{172})$ board configurations) at the same level as the world's best players [19]. This thesis explores the possibility of using similar heuristic algorithms to explore the string landscape to find instances of the standard model.

This thesis is based on the following research papers:

- Chapter 2:

  T. R. Harvey and A. Lukas. "Quark Mass Models and Reinforcement Learning". In: JHEP 08 (2021), p. 161. [20]

- Chapter 3:

  Steve A. Abel, Andrei Constantin, Thomas R. Harvey, Andre Lukas and Luca A. Nutricati. "Decoding Nature with Nature's Tools: Heterotic Line Bundle Models of Particle Physics with Genetic Algorithms and Quantum Annealing". In: Fortsch. Phys. 72.2 (2024), p. 2300260. [21]

Field	Names	Representation under $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U_Y}(1)$
ψ^i_L	Left handed leptons	$(\mathbf{1},\mathbf{2})_{-1/2}$
ψ^i_E	Right handed electrons	$(\mathbf{1},\mathbf{1})_{1}$
ψ^i_Q	Left handed quarks	$(\mathbf{3},\mathbf{2})_{1/6}$
ψ^i_U	Right handed up quarks	$(\bar{\mathbf{3}},\mathbf{1})_{-2/3}$
ψ^i_D	Right handed down quarks	$(\bar{\mathbf{3}},\mathbf{1})_{1/3}$
H	The Higgs	$(\mathbf{1},\mathbf{2})_{-1/2}$

Table 1.1: The representation content of the standard model, where the index $i = 1, 2, 3$ indicates the three different families.

- Chapter 4:

  Steve A. Abel, Andrei Constantin, Thomas R. Harvey and Andre Lukas. “Heterotic String Model Building with Monad Bundles and Reinforcement Learning”. In: Fortsch. Phys. 70.2-3 (2022), p. 2100186. [22]

  Steve A. Abel, Andrei Constantin, Thomas R. Harvey and Andre Lukas. “Evolving Heterotic Gauge Backgrounds: Genetic Algorithms versus Reinforcement Learning”. In: Fortsch. Phys. 70.5 (2022), p. 2200034. [23]

  Steve A. Abel, Andrei Constantin, Thomas R. Harvey and Andre Lukas. “String Model Building, Reinforcement Learning and Genetic Algorithms”. In: Nankai Symposium on Mathematical Dialogues: In celebration of S.S.Chern’s 110th anniversary. Nov. 2021. [24]

The rest of this chapter presents the relevant background material for the thesis. We start with a review of the standard model of particle physics (SM), along with suggested physics for what can come beyond the standard model (BSM), in Sec. 1.2. The relevant parts of string theory are discussed in Sec. 1.3, followed by the required mathematical and computational tools and techniques in Sec. 1.4 and Sec. 1.5.

1.2 The Standard Model

As mentioned in the previous section, the standard model is a chiral gauge theory, with gauge group $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U_Y}(1)$, coupled to 3 generations of fermionic matter and a complex scalar [25]. In schematic notation, the Lagrangian is of the form

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F^2 + \frac{\theta}{32\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \\ &+ \sum_I \bar{\psi}_I \not{\mathcal{D}}_I \psi_I \\ &+ \sum_{I\,J} Y_{IJ}\bar{\psi}_I H \psi_J + \text{h.c.} \\ &- |\mathcal{D}H|^2 + \mu^2|H|^2 - \frac{\lambda}{4}|H|^4.\end{aligned} \tag{1.1}$$

The representation content is listed in Tab. 1.1, and the index I is taken over all fields in that table except for the Higgs H. The Higgs field is tachyonic when expanded around $H = 0$. Expanding around the true vacuum $|H| = \sqrt{\frac{2\mu^2}{\lambda}}$ results in spontaneous symmetry breaking of the electroweak sector $\mathrm{SU}(2) \times \mathrm{U}_\mathrm{Y}(1)$ to the gauge group of electromagnetism $\mathrm{U}_\mathrm{E}(1)$.

The free parameters of the renormalizable standard model are the three gauge couplings, the theta parameters θ (one for each of the group factors), the Yukawa couplings Y_{ij}, and the parameters of the Higgs potential μ and λ. If interpreted as an effective field theory, one should also include an infinite number of higher order couplings, including those that lead to neutrino masses.

While the Yukwawa couplings appear in the action, they are not directly observed. Instead, after electroweak symmetry breaking we generate terms quadratic in the fermion fields

$$\sqrt{\frac{2\mu^2}{\lambda}}\sum_{I\,J} Y_{IJ}\bar{\psi}_I\psi_J + \text{h.c.} = \sum_{I\,J} M_{IJ}\bar{\psi}_I\psi_J + \text{h.c.}\,, \tag{1.2}$$

which lead to the masses and mixings of the quarks and leptons. This results in hierarchical masses between the different generations of quarks and leptons (except the neutrinos), along with large mixing between the different generations of leptons and relatively small mixing between the quarks. Further, as the right handed rotations from the singular value decomposition drop out of all observable quantities, the Yukawa couplings are degenerate with respect to the observables. The theta term is often dropped from (1.1), as the experimental bounds require it to be very small ($\theta \leq 10^{-10}$ [26]), and it is also a total derivative so only contributes non-perturbatively.

The Standard Model has shown remarkable success in explaining observed collider data. Despite this, we know it is incomplete, and some parts of it remain mysterious. For one, there is no explanation as to why the group theoretic structure, observed in Tab. 1.1, is so complicated and repeated across three generations. There also appears to be no natural explanation for the detailed structure of Yukawa couplings required to generate the observed masses and mixings – with these instead chosen "by hand" to match with experiment.

Also, as presented in (1.1), the standard model does not contain gravity. For the purposes of early universe cosmology one could formulate the standard model in a curved background, and treat the gravitational effects semi-classically. However, one expects such methodology to break down if the curvature becomes large, or the energy scales approach $\mathcal{O}(M_{pl})$. Really, a complete microscopic and quantum theory of gravity is required to consider these regimes.

1.2.1 Beyond the Standard Model

The problems listed at the end of the last subsection have motivated several extensions to the standard model. Of particular interest for this thesis are grand unified theories (GUTs), supersymmetry (SUSY) and string theory. The necessary information for the first two will be discussed in this subsection, while string theory will be covered in much greater detail in Sec. 1.3.

Technically, the string models forming the major focus of this thesis are not traditional GUTs, but have a GUT-like structure. The differences between our string models and traditional GUTs will be discussed when necessary.

Grand Unified Theories (GUTs)

Aesthetically, it would be preferable to unify the electroweak and strong forces. That is to say, we wish to embed $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U_Y}(1)$ into a simple Lie group, potentially simplifying the group theoretic structure of the standard model in the

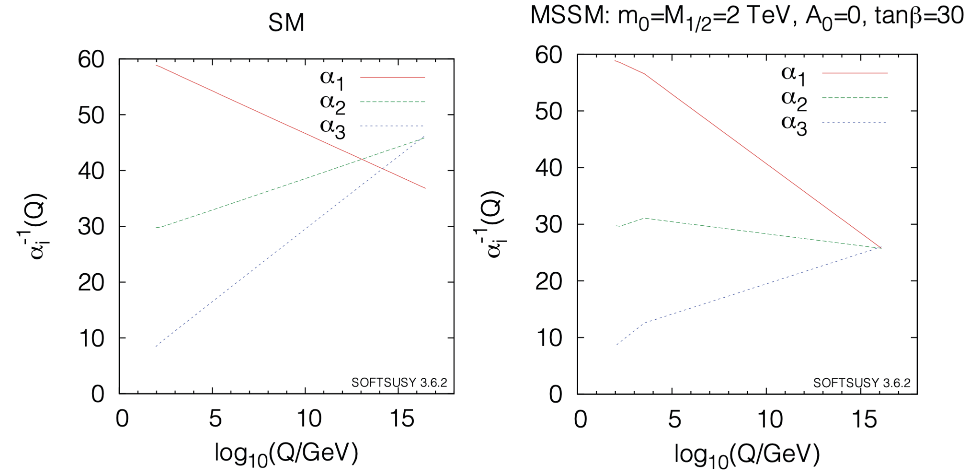


Figure 1.1: The two-loop running for the gauge couplings in the standard model (left) and its supersymmetric extension (right) with example parameters for a SUSY threshold at 2TeV [26].

process. There is also phenomenological evidence for such a scenario. Using the one-loop beta functions of the standard model to extrapolate the gauge couplings, we find that they would appear to nearly intersect at high energies. As can be seen in Fig. 1.1, in models which also contain low energy supersymmetry the accuracy of this intersection dramatically improves and appears to be at $\Lambda_{GUT} \sim 10^{16}$GeV [27–29]. However, one may also find an increase in accuracy for a unification scale by introducing other heavy fields, not necessarily related to supersymmetry. As a result, this should not necessarily be seen as evidence of supersymmetry.

Typically, in such unified models, the GUT group is broken to the standard model by spontaneous symmetry breaking at $E \sim \Lambda_{GUT}$. In our string models however, we will instead break the gauge group via discrete Wilson lines, which require the extra dimensions we have available in string theory. Since we will not need a GUT-Higgs, we do not discuss them.

Since the standard model is chiral, we are limited to GUT groups that admit complex representations. This, alongside considerations of the representation content, essentially limits us to three possible GUT groups: SU(5), SO(10) and E_6. These form a chain of maximal subgroups

$$\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U_Y}(1) \subset \mathrm{SU}(5) \subset \mathrm{SO}(10) \subset \mathrm{E}_6, \tag{1.3}$$

which we will discuss in order of increasing dimension.

Beginning with SU(5), we consider the branching of

$$\mathrm{SU}(5) \rightarrow \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}_{\mathrm{Y}}(1), \tag{1.4}$$

where the $\overline{\mathbf{5}}$ and $\mathbf{10}$ irreps of SU(5) become

$$\begin{aligned} \overline{\mathbf{5}} &\rightarrow (\overline{\mathbf{3}}, \mathbf{1})_{1/3} \oplus (\mathbf{1}, \mathbf{2})_{-1/2}, \\ \mathbf{10} &\rightarrow (\overline{\mathbf{3}}, \mathbf{1})_{-2/3} \oplus (\mathbf{3}, \mathbf{2})_{1/6} \oplus (\mathbf{1}, \mathbf{1})_{1}, \end{aligned} \tag{1.5}$$

which we recognise as being the same as the matter spectrum of a single standard model family (c.f. Tab. 1.1). This is, in some sense, the minimal possible simple GUT group – no new particles are introduced in the matter spectrum. One can also obtain a Higgs doublet from a $\overline{\mathbf{5}}$ representation, but this results in an extra, and unobserved, triplet on breaking to the standard model. The difficulty in generating a sufficently large mass for the extra triplet is referred to as the doublet-triplet problem.

For SU(5) the gauge fields are in adjoint representation $\mathbf{24}$. We can branch this to the standard model, leading to

$$\mathbf{24} \rightarrow (\mathbf{8}, \mathbf{1})_{0} \oplus (\mathbf{1}, \mathbf{3})_{0} \oplus (\mathbf{1}, \mathbf{1})_{0} \oplus (\mathbf{3}, \mathbf{2})_{-5/6} \oplus (\overline{\mathbf{3}}, \mathbf{2})_{5/6}. \tag{1.6}$$

The first three terms are identified with the gauge bosons of the standard model. The last two terms are referred to as X and Y bosons, which after spontaneous symmetry breaking have a mass of $\mathcal{O}(\Lambda_{\mathrm{GUT}})$.

Both the Higgs-triplet and the X and Y bosons can generate higher dimensional operators that cause fast proton decay when integrated out. As mentioned above, in our string models, we will break the GUT group to the standard model via discrete Wilson lines rather than with a GUT-Higgs. This discrete Wilson line can be chosen to project out the Higgs-triplet, but proton decay via X and Y bosons is still a potential concern. We will not comment on the gauge fields for the other GUT groups, as their qualitative behaviour is the same.

Continuing along the chain of GUT groups from (1.3), we turn to SO(10). Branching into SU(5), we find we only need a single irrep for each standard model family – the **16** irrep of SO(10).

$$\mathbf{16} \to \mathbf{10} \oplus \overline{\mathbf{5}} \oplus \mathbf{1}. \tag{1.7}$$

We already know, from earlier considerations, that the first two irreps are able to form a standard model family. The new extra singlet has the correct quantum numbers to be a right handed neutrino in the low energy theory. Although we had to introduce a new particle, the SO(10) GUT has the benefit of only requiring a single irrep of the GUT group per family of the standard model. It is also worth mentioning that right handed neutrinos are well motivated by the indirect observation of neutrino masses from neutrino oscillations [28].

The standard model Higgs comes from a $\mathbf{10} \to \mathbf{5} \oplus \overline{\mathbf{5}}$ rep of SO(10). However, this leads to two Higgs doublets usually labelled as H_u and H_d. While this may appear unusual, having a pair of Higgs doublets will be a requirement when we consider supersymmetric extensions to the standard model.

Lastly we turn to E_6, where the **27** irrep branches into $SO(10)$ as

$$\mathbf{27} \to \mathbf{16} \oplus \mathbf{10} \oplus \mathbf{1}. \tag{1.8}$$

On further branching to the standard model, the **16** irrep provides us with one standard model family (and a right handed neutrino). In this process, we are left with three **10** multiplets of which, for a supersymmetric GUT, one single linear combination can be identified with a pair of Higgs doublets. The two extra **10** multiplets, and the other singlets in (1.8), are extra, and unobserved states, which would need to obtain a large mass.

We will find in Sec. 1.3.3 that GUTs, and models with a GUT-like structure, naturally arise in compactifications of the $\mathrm{E}_8 \times \mathrm{E}_8$ heterotic string. We will construct string models with GUT-like structures with GUT groups SU(5) and SO(10) in Chapter 3 and Chapter 4 respectively.

Supersymmetry (SUSY) and Supergravity (SUGRA)

The Coleman-Mandula Theorem states that, under assumptions that hold for most traditional relativistic field theories, the most general symmetry group G of a quantum field theory is of the form [30]

$$G = G_{\text{Poincaré}} \times G_{\text{Internal}}. \tag{1.9}$$

That is, the spacetime and internal symmetries cannot be combined in a non-trivial way.

However, one of the assumptions of the Coleman-Mandula theorem, alluded to above, is that G is a Lie group. One can extend the Poincaré group to a supergroup by also including fermionic generators Q^I_α, where $I = 1, \ldots \mathcal{N}$, called supercharges which satisfy anti-commutation relations[1]. The symmetry generated by such a supercharge is called supersymmetry (SUSY) [32–35]. When transforming an operator with respect to a SUSY transformation, fermions must be taken into bosons and vice-versa. As a result, in the same way that the Wigner classification states that particle multiplets are projective representations of the Poincaré group, multiple particles are combined into projective representations of the new supergroup called supermultiplets [35–39].

For $\mathcal{N} = 1$ in 4D, we introduce a single Weyl (or Majorana) spinor Q_α into the Poincaré algebra which satisfies

$$\{Q_\alpha, \overline{Q}_{\dot{\alpha}}\} = 2\sigma^\mu_{\alpha\dot{\alpha}} P_\mu, \tag{1.10}$$

where P_μ is the generator for translations in the Poincaré Lie algebra. In this case, there are three types of supermultiplets. Indicating spinor indices with $\alpha, \beta, \ldots$, spacetime indices with $\mu, \nu, \ldots$ and gauge indices with $a, b, \ldots$, the supermultiplets are:

---

[1] Counting the amount of symmetry depends on dimension, as it is closely tied to the spinor representations in that dimension. This is only further complicated by the varying conventions in the literature. In this thesis $\mathcal{N} = 1$ supersymmetry will mean that SUSY is generated by a single and minimal spinor for that space time dimension. That is $\mathcal{N} = 1$ in 4 dimensions implies there is a single Weyl (or Majorana) spinor added to the Poincaré algebra, or equivalently 4 real supercharges [31].

- **Chiral Multiplet** (χ_α, ϕ) - Consisting of a chiral fermion χ, and complex scalar ϕ. These can be charged under any of the internal symmetries, including gauge symmetries, of the field theory.
- **Vector Multiplet** $(\lambda^a_\alpha, A^a_\mu)$ - Consisting of a Weyl fermion λ, and a gauge field A. Both of these transform in the adjoint of some gauge group.
- **Gravity Multiplet** $(\psi^\mu_\alpha, g_{\mu\nu})$ - Consisting of a spin-3/2 gravitino ψ and a graviton g. Strictly speaking this requires supergravity (SUGRA), which we will discuss at the end of this section.

When important to stress the difference between SUSY and SUGRA, we shall refer to SUSY theories as having global-SUSY, as opposed to SUGRA theories which have local-SUSY.

In addition to requiring the fields to appear in such supermultiplets, SUSY also heavily constrains the form of the Lagrangian. For example, the most generic 2-derivative, 4D $\mathcal{N} = 1$ global-SUSY theory has a bosonic Lagrangian of the form

$$\begin{aligned}\mathcal{L} = &- \mathrm{Re}(f_{ab}(\phi))\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu\, b} + \frac{1}{8}\mathrm{Im}(f_{ab}(\phi))F^a_{\mu\nu}F^b_{\rho\sigma}\epsilon^{\mu\nu\rho\sigma} \\ &- K_{\bar{i}j}(\phi,\overline{\phi})\overline{D^\mu\phi^i}D_\mu\phi^j - V_F(\phi,\overline{\phi}) - V_D(\phi,\overline{\phi}) \\ &+ [\text{fermion terms}](W(\phi), K(\phi,\overline{\phi})),\end{aligned} \tag{1.11}$$

where the index $i, j, ...$ run over the chiral multiplets of the theory.

Due to this restricted form, once the gauge group and representations of the supermultiplets are specified, the rest of the theory is defined by three components, each of which depend on the scalar fields: $f_{ab}(\phi^i), W(\phi^i)$ and $K(\phi^i, \overline{\phi^i})$. The first is the gauge-kinetic function, and is holomorphic in the fields ϕ^i. Similarly, W is a function called the superpotential, which is also holomorphic in ϕ^i. Lastly, K is a real function of ϕ^i and $\overline{\phi^i}$ called the Kähler potential, and is used to generate the Kähler metric $K_{\bar{i}j} = \partial_{\bar{i}}\partial_j K$. The two contributions to the scalar potential, V_F and V_D, are called F-terms and D-terms respectively. The F-term scalar potential can appear in any theory with chiral multiplets, and is given by

$$V_F = K^{\bar{i}j}(\phi,\overline{\phi})\overline{\partial_i W(\phi)}\partial_j W(\phi). \tag{1.12}$$

D-terms, however, only occur in SUSY theories with chiral multiplets charged under gauge groups, which are isometries of the Kähler metric $K_{\bar{i}j}(\phi, \overline{\phi})$. The D-terms are introduced to restore gauge invariance of the theory.

Looking at (1.12), we observe that the scalar potential is positive definite. As a result, the vacuum energy is guaranteed to be $\Lambda \geq 0$ for a theory with global-SUSY. In fact, on considering the variation of the fermionic fields with respect to a SUSY transformation, we observe that solutions that preserve supersymmetry require the stronger condition that $\Lambda = 0$.

We now wish to include gravity in our supersymmetric models. In some sense, gravity can be viewed as a gauge theory of the Poincaré group [40]. As a result, since we have extended the Poincaré algebra to include fermionic generators, we also need to gauge supersymmetry. This gauged supersymmetry is called supergravity[2] (SUGRA). The bosonic action is schematically of the same form as (1.11), but the theory is now coupled to gravity and additional terms suppressed by $1/M_{\text{pl}}$ appear in the scalar potential. Specifically the F-term scalar potential is now given by

$$V_F^{(SUGRA)} = e^{K/M_{\text{pl}}^2} \left[K^{\bar{i}j}(\phi, \overline{\phi}) \overline{D_i W(\phi)} D_j W(\phi) - \frac{3|W|^2}{M_{\text{pl}}^2} \right], \tag{1.13}$$

where $D_i = \partial_i + (\partial_i K)/M_{\text{pl}}$. The new, Planck-suppressed, second term of this potential removes the positive definiteness property we observed in the global-SUSY case. Further, if we again consider the variation of the fermion fields with respect to a SUSY transformation, it emerges that the solutions that preserve supersymmetry require the vacuum energy to be $\Lambda \leq 0$.

To embed the standard model into string theory, we need to assign supersymmetric partners for each of the standard model fields in Tab. 1.1, such that they fit into supermultiplets. The simplest way of achieving this is called the Minimal Supersymmetric Standard Model (MMSM) [41]. The gauge bosons are lifted to vector multiplets, while the Higgs and the standard model families are lifted to chiral multiplets – introducing their appropriate supersymmetric partners in the process. SUSY is incompatible with a single Higgs boson, and so we consider a

---

[2] This is distinct to gauged supergravity, which refers to a supergravity theory where the gravitini are charged under the vector multiplet gauge symmetries [40].

Multiplet	Multiplet Type	Standard Model Component	Superpartner
L^i	Chiral	Left handed leptons ψ^i_L	Left sleptons ϕ^i_L
E^i	Chiral	Right handed electrons ψ^i_E	Right sleptons ϕ^i_E
Q^i	Chiral	Left handed quarks ψ^i_Q	Left squarks ϕ^i_L
U^i	Chiral	Left handed up quarks ψ^i_U	Up squarks ϕ^i_U
D^i	Chiral	Left handed down quarks ψ^i_D	Down squarks ϕ^i_D
W	Vector	Gauge bosons A^μ	Gauginos λ
H_u	Chiral	Up-Higgs H	Up Higgsino ψ_{H_u}
H_d	Chiral	Down-Higgs	Down Higgsino ψ_{H_d}

Table 1.2: The particle content of the MSSM, where $i = 1, 2, 3$ indicates the three different families. The standard model components are in the same notation as the in Tab. 1.1, except we now have a pair of Higgs doublets – which is required to avoid an anomaly as explained in the main text. The representation multiplets are there same as their respective standard model counterparts. However H_d transforms in the conjugate representation to the standard model Higgs H. The vector multiplet in this list is for all the gauge bosons in the standard model.

pair of Higgs doublets as described in the context of SU(5) GUTs above. The extra Higgs is needed to avoid introducing an anomaly from the chiral fermion superpartner of the original Higgs. This is summarised in Tab. 1.2.

From this point onwards, unless otherwise specified, any use of the phrase "the standard model" should be understood as meaning the MSSM.

As SUSY remains unobserved, for a fully realistic model it will need to be spontaneously broken at some scale above the electroweak scale probed by the LHC. For the string models constructed in this thesis, the resulting 4D theory will have $\mathcal{N} = 1$ SUSY, and SUSY breaking would be closely related to moduli stabilisation. While a detailed account of moduli stabilisation for our models will not be considered, some generalities will be discussed in Sec. 1.3.3.

1.3 String Theory

Compared to GUTs and SUSY, string/M-theory is a far more ambitious and bold extension to known physics. We will begin with some qualitative statements about string/M-theory, before moving onto some more concrete and quantitative statements.

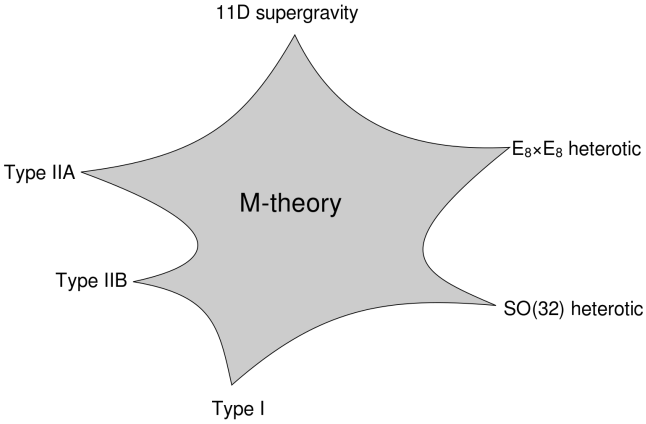


Figure 1.2: A cartoon of M-theory, with the different superstring theories at special limits along with 11-dimensional supergravity [42].

M-theory has a long, and somewhat complicated history [43–45]. Yet, despite this long history, in its current state, it is impossible to give a completely satisfactory description of what M-theory is. M-theory describes a world of not just particles, but other extended objects called branes. In certain limits, the perturbative sectors are described by one-dimensional branes called strings, which at low energies produce, and propagate in, ten-dimensional supergravity theories. In other limits, the universe is best described by eleven-dimensional supergravity.

In certain regions of parameter space, two different descriptions may both be accurate descriptions of the world, but may rely on two very different descriptions of the degrees of freedom – such overlaps are called dualities [46, 47]. The current state of M-theory is analogous to an incomplete description of a manifold, where the open sets we currently understand are the different superstring theories and the transition functions are dualities between them. A cartoon representation of this is indicated by Fig. 1.2, where M-Theory is some continuous "parameter space" that is well described by the different string theories in different limits. Crucially this "parameter space" should not be considered as encoding free parameters of the theory, but instead the background values of dynamical fields. The only non-dynamical parameter of string theory is $\alpha' \sim \left(M_{pl}^{(10)}\right)^{-2}$ which sets the overall scale of the theory.

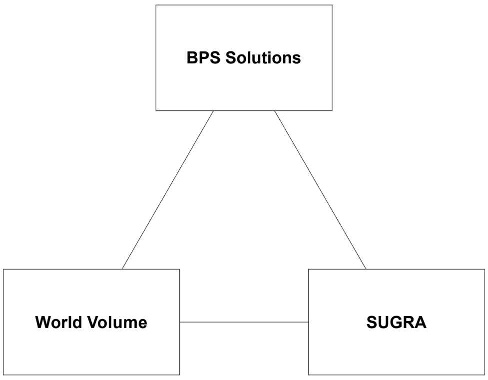


Figure 1.3: Most of the understanding of string theory has come from studying the relationship between the three parts of this image. Namely the world volume theories, the space time SUGRA theories and the realisation of extended objects as BPS solutions to supergravity theories.

String theory offers a framework in which all of the fundamental interactions of nature can operate in a unified and quantum mechanical framework. Specifically, it contains both non-abelian gauge theories and Einstein gravity.

From now on, we will be more precise and concrete in our discussions of string theory. One of the salient aspects of string theory is the relationship between spacetime and world volume theories (see Fig. 1.3). Focusing on the string for the moment, rather than the higher-dimensional analogues, one can consider a field theory living on its worldsheet – the higher-dimensional analogue of the world line. The action of such a field theory, called the Nambu-Goto action, is proportional to the area of the worldsheet WS, and is given by

$$S_{NG}[X^\mu(\sigma^a)] = \frac{1}{2\pi\alpha'}\int_{WS} d^2\sigma \sqrt{-\det_{(\mathrm{ab})}\left(g_{\mu\nu}(X)\frac{dX^\mu}{d\sigma^a}\frac{dX^\nu}{d\sigma^b}\right)}, \tag{1.14}$$

where $g_{\mu\nu}$ is the metric for the spacetime the string is propagating in. Quantising this action in flat space, first performed by Dirac in a rather different context [43], is the typical focus for most introductions to string theory [45, 48]. We do not have the time to go through the quantisation in great detail, but we state the key results for closed strings (that is, those with periodic boundary conditions):

- The dimension of spacetime is required to be $D = 26$.

- Each state on the string is, when viewed from spacetime, identified with a particle. There are no fermionic excitations and the lightest state is a tachyon. There are three massless states: the graviton (as a perturbation around flat space) $h_{\mu\nu}$, a 2-form field $B_{\mu\nu}$ and a scalar ϕ called the dilaton. The dynamics of these massless states can be described by a 26-dimensional effective field theory. For higher states, there exists an infinite tower of increasingly massive states, the first of which has a mass of $\mathcal{O}\left(\frac{1}{\sqrt{\alpha'}}\right)$.

- The vev of the dilaton sets the value for all couplings in the theory, called the string-coupling, $g_{st} = e^{-\langle\phi\rangle}$.

To resolve the lack of fermionic states, we can supersymmetrise the worldsheet theory by introducing extra fermionic modes on the worldsheet. This is called superstring theory. Continuing down this path, one finds that there are precisely five different possible tachyon free superstring theories with spacetime SUSY. These have a number of similar properties to the bosonic case, but we now predict ten spacetime dimensions and the ten-dimensional effective field theories that describe the massless states are all SUGRA theories.

While the above may lead one to the idea that the supergravity description is merely an approximation to the full worldsheet theory, this is not quite true. In fact, SUGRA theories allow one to probe some of the non-perturbative properties of the full string theory, via their BPS (SUSY preserving) solutions. Amongst a plethora of other extended objects, all of the 10D SUGRA theories contain a string as a BPS solution called the F1-string. This indicates that the SUGRA theory already “knows” something about the full string theory. One can also consider the world volume theories of the other extended objects from the SUGRA theories.

We will begin by describing one of the five superstring theories, the $E_8 \times E_8$ heterotic string, in Sec. 1.3.1 before moving onto relating it to four dimensional physics and the standard model in Sec. 1.3.2 and Sec. 1.3.3 respectively.

1.3.1 The Heterotic String

As mentioned above, excitations of quantum strings form infinite towers of increasing mass where the first massive states have Planckian masses [31, 45, 48–50]. As we are predominately interested in the low energy phenomenology of string theory, we will only concern ourselves with the massless states, and the low energy SUGRA description. The observed masses of quarks and leptons will instead be generated dynamically with the Higgs mechanism.

At low energies $E < M_{pl}^{(10)}$, and weak coupling[3] $g_{st} < 1$, the massless states of the $E_8 \times E_8$ Heterotic string form a 10 dimensional $\mathcal{N} = 1$ (i.e. 16 real super charges) SUGRA theory coupled to $\mathcal{N} = 1$ super-Yang-Mills with gauge group $E_8 \times E_8$. Indicating spacetime indices with $\mu, \nu, ...$, spinor indicies with $\alpha, \beta, ...$ and gauge indices $a, b...$, the field content of such a theory is

- **Gravity Sector**: metric $g_{\mu\nu}$, 2-form field $B_{\mu\nu}$, the dilaton ϕ, dilatino λ_α and gravitino ψ^μ_α.
- **Gauge Sector**: gauge bosons A^a_μ with field strength $F^a_{\mu\nu}$ and adjoint valued gauginos χ^a_α.

Of particular importance to us, is the ten-dimensional bosonic action

$$S = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-g} e^{-2\phi} \left(R + 4(\partial\phi)^2 - \frac{1}{2} H^2 - \frac{\alpha'}{4} \operatorname{Tr} F^2 \right) + \text{(fermions)}, \tag{1.15}$$

where $\kappa_{10} \propto (\alpha')^4$ is the ten-dimensional Newton's constant and H is defined by

$$\begin{aligned} H &= dB - \frac{\alpha'}{4}(\Omega_{\text{YM}} - \Omega_{\text{Gr}}), \\ \Omega_{\text{YM}} &= \text{Tr}_{\text{E}_8 \times \text{E}_8} \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right), \\ \Omega_{\text{Gr}} &= \text{Tr}_{\text{TX}} \left(\omega \wedge d\omega + \frac{2}{3} \omega \wedge \omega \wedge \omega \right), \end{aligned} \tag{1.16}$$

---

[3] It should be noted that, to match the the relevant 4D scales, the models we construct will be in the regime where $g_{st} > 1$. One may worry about our ability to calculate outside of the perturbative regime. However as discussed in Sec. 1.3, there exist dualities between different string theories and 11 dimensional SUGRA. It turns out there exists one such duality between 11 dimensional SUGRA on a line interval, and the strongly coupled 10-dimensional heterotic string called the Horava-Witten construction [7, 8, 51, 52]. For practical purposes, we will be working with the weakly coupled heterotic string, but on the understanding that this needs to be lifted to the strongly coupled version. Luckily, this uplifting results in no changes to the eventual 4D theories we will be constructing [52].

where ω is the spin connection, and the second trace is over space-time indices (the subscript TX refers to the trace being associated to the tangent bundle).

Under a SUSY variation with fermionic parameter ϵ, the fermion fields transform as

$$\begin{aligned}
\delta_\epsilon \psi_M &= \frac{1}{\kappa_{10}} D_M \epsilon + \frac{1}{8\sqrt{3}\kappa_{10}} e^{-\phi} \left(\Gamma_M{}^{NPQ} - 9\delta_M^N \Gamma^{PQ} \right) \epsilon H_{NPQ} + (\text{fermion})^2, \\
\delta_\epsilon \chi^a &= -\frac{\sqrt{\alpha'}}{4\sqrt{2}\kappa_{10}} e^{-\phi/2} \Gamma^{MN} F^a_{MN} \epsilon + (\text{fermion})^2, \qquad (1.17) \\
\delta_\epsilon \lambda &= -\frac{1}{\sqrt{2}} (\Gamma \cdot \partial\phi)\epsilon + \frac{1}{4\sqrt{6}\kappa_{10}} e^{-\phi} \Gamma^{MNP} \epsilon H_{MNP} + (\text{fermion})^2.
\end{aligned}$$

We will not need to consider the variation of the bosonic fields when looking for supersymmetry preserving solutions, as these are all of the form $\delta_\epsilon \langle B \rangle \sim \langle F \rangle \epsilon = 0$, where $\langle B \rangle$ and $\langle F \rangle = 0$ represent generic bosonic and fermionic background values respectively.

Why does this appear to be a promising starting point for phenomenology? For one, this theory is already chiral, and contains non-Abelian gauge fields – both being key components of the standard model. Further to that, E_8 contains $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$ as a subgroup. On the other hand, the immediate difficulty in relating any of the superstring theories to our world is the six extra spacial dimensions. We take the compactification route, where the geometry of the universe is locally a product space

$$\mathcal{M} = \mathcal{M}_{1,3} \times X, \qquad (1.18)$$

where the volume of X is significantly smaller than the volume of $\mathcal{M}_{1,3}$, yet significantly larger than α'^3 so that the supergravity limit is valid. In Sec. 1.3.2 we will explore this in further detail, and in particular impose that the effective four dimensional theory is supersymmetric. This amounts to setting (1.17) to zero, with the modified Bianchi identitty (1.16) as a constraint, and is sufficient to solve the full second-order equations of motion. The details of the resulting 4-dimensional theories are then given in Sec. 1.3.3.

1.3.2 Calabi-Yau (CY) Compactifications

We now turn to relating heterotic string theory to four dimensional phenomenology, by considering the requirements for the compactification to lead to a four dimensional $\mathcal{N} = 1$ theory. This section will necessarily closely follow Refs. [6, 49, 53, 54].

We wish to take our ten-dimensional theory to a four-dimensional theory with $\mathcal{N} = 1$ supersymmetry. We begin by assuming a product manifold $\mathcal{M}_{1,9} = \mathcal{M}_{1,3} \times X$, where $\mathcal{M}_{1,3}$ is assumed to be maximally symmetric. We label our 10D indices as $M, N, ... = 0, 1, ..., 9$, 4D indices as $\mu, \nu, ... = 0, 1, 2, 3$, and the indices for the compact dimensions are $i, j, ... = 4, 5, ..., 9$. We also label 4D coordinates as x^μ and compact coordinates as y^i.

For the moment we will ignore the gauge sector, and our remaining fields split as

$$\begin{aligned} g_{MN} &\quad \rightarrow \quad g_{\mu\nu} \quad g_{\mu i} \quad g_{ij}, && (1.19) \\ H_{MNP} &\quad \rightarrow \quad H_{\mu\nu\rho} \quad H_{\mu\nu i} \quad H_{\mu ij} \quad H_{ijk}, && (1.20) \\ \phi &\quad \rightarrow \quad \phi. && (1.21) \end{aligned}$$

By assuming $g_{\mu\nu}$ is maximmally symmetric (that is $g_{\mu\nu}$ is the metric on Minkowski space or AdS space[4]), we are forced to set $g_{\mu i} = H_{\mu\nu i} = H_{\mu ij} = H_{\mu\nu\rho} = 0$. These would otherwise correspond to some background electric field throughout space time, breaking isotropy.

To preserve some amount of sypersymmetry, we take this ansatz and set the variations of the gravitino and dilatino with respect to the preserved supercharges, with corresponding SUSY parameter η, from equations (1.17) to zero:

$$\delta\psi_\mu = \partial^{(4)}_\mu \eta = 0, \tag{1.22}$$

$$\delta\psi_i = \left(\nabla^{(6)}_i + \frac{1}{8} H_{ijk}\Gamma^{jk}\right)\eta = 0, \tag{1.23}$$

$$\delta\lambda = \left(\nabla\!\!\!\!/^{(10)}\phi + \frac{1}{12} H_{ijk}\Gamma^{ijk}\right)\eta = 0. \tag{1.24}$$

---

[4] dS is ruled out by our later requirement of preserving 4D $\mathcal{N} = 1$ SUSY, as described in Sec. 1.2.1.

As written, the above equations describe a compactification with NS-flux. For now, we will set the flux to zero ($H_{ijk} = 0$), but we will return to this in the context of moduli stabilisation in Sec. 1.3.3.

In this specific ansatz, we are left with 3 equations to satisfy:

$$\partial^{(4)}_\mu \eta = 0, \tag{1.25}$$

$$\nabla^{(6)}_i \eta = 0, \tag{1.26}$$

$$(\not{\nabla}^{(10)} \phi)\eta = 0. \tag{1.27}$$

The first and third of these are satisfied by making ϵ independent of the 4D coordinates and setting ϕ to a constant.

The second condition states that, in the absence of flux, if we preserve some amount of SUSY in 4D we require covariantly constant spinors over the compact space. This requires some further analysis as it is incredibly restrictive. For one, by taking a second derivative we find

$$\nabla^{(6)}_i \nabla^{(6)}_j \eta = 0 \quad \Rightarrow \quad [\nabla^{(6)}_i, \nabla^{(6)}_j]\eta = 0 \quad \Rightarrow \quad R^{(6)}_{ij} = 0. \tag{1.28}$$

That is to say, the compact 6D manifold is Ricci-flat.

How many covariantly constant spinors do we need? Consider the branching of the real Majorana-Weyl 10D spinor into $SO(1,3) \times SO(6)$

$$\mathbf{16} \to (\mathbf{2}, \mathbf{4}) \oplus (\mathbf{2}', \overline{\mathbf{4}}). \tag{1.29}$$

If the $\mathbf{4}$ and $\overline{\mathbf{4}}$ both contain a single covariantly constant spinor, then we are left with a $\mathbf{2} \oplus \mathbf{2}'$ spinor in the 4 dimensional theory. From the reality condition of 10 dimensional spinor, $\mathbf{2} \oplus \mathbf{2}'$ is a single Majorana spinor in 4 dimensions, which is sufficient to generate $\mathcal{N} = 1$ SUSY. We require this singled-out direction of the $\mathbf{4}$ to be preserved under parallel transport through the compact space. The $\mathbf{4}$ and $\overline{\mathbf{4}}$ spinor representations of $SO(6) \approx SU(4)$ are the fundamental and anti-fundamental representations of $SU(4)$, and the subgroup of $SU(4)$ that preserves this single spinor is $SU(3)$. Therefore, we are interested in manifolds with $SU(3)$ holonomy – that is, the group action under parallel transport is $SU(3) \subset SO(6)$.

Mathematicians have studied manifolds with these properties in great detail, and a manifold with $SU(3)$ holonomy is called a Calabi-Yau (CY) manifold. There are several definitions of Calabi-Yau manifolds, not all of which are equivalent [55]. One that is particularly convenient for us is that it is a Kähler manifold with vanishing first Chern class. Yau's theorem guarantees that a unique Ricci-flat metric exists in each Kähler class [56].

For now, we shall postpone any further details of these manifolds, and their construction, to Sec. 1.4.1. However, we state a few properties of these manifolds that we will require before then. As these manifolds are Kähler they, by definition, admit a closed and non-exact (1,1)-form J. From the Calabi-Yau condition there also exists a closed, non-exact and holomorphic (3,0)-form Ω. CYs are also complex manifolds, and as such the complex structure splits the tangent bundle into holomorphic and anti-holomorphic parts. Consequently, from this point onwards we split the indices of the compact manifold into holomorphic and anti-holomorphic indices

$$\{i, j, k, ... = 4, 5, ...9\} \to \left\{i, j, k, ... = 1, 2, 3;\ \bar{i}, \bar{j}, \bar{k}, ... = \bar{1}, \bar{2}, \bar{3}\right\}.$$

Once we have a solution for the metric, we must also solve for the gauge fields. These can take a topologically non-trivial field profile over the CY, which we should expect to describe in the language of vector bundles. In fact, the gauge fields are required to be topologically non-trivial due to a consistency condition for our compactification. Specifically, we have, due to the non-trivial Bianchi identity in (1.16), the anomaly cancellation condition condition [53]

$$dH \sim \mathrm{tr}(F \wedge F) - \mathrm{tr}(R \wedge R). \tag{1.30}$$

As we will not be constructing explicit solutions for the metric or connection, we instead impose this constraint at the level of cohomology. Namely, the above constraint can be written in terms of the second Chern classes of the tangent bundle TX and the vector bundle V corresponding to the gauge fields

$$\mathrm{c}_2(V) - \mathrm{c}_2(TX) = 0. \tag{1.31}$$

Generically $c_2(TX) \neq 0$ for CY manifolds and so V cannot be the trivial bundle.

With the above constraint in mind, to actually find a solution of the gauge fields we return to the gaugino variation. Again requiring maximal symmetry of the 4D theory, equations (1.17) imply

$$g^{i\bar{j}}F_{i\bar{j}} = 0 \qquad F_{ij} = 0 \qquad F_{\bar{i}\bar{j}} = 0. \tag{1.32}$$

These are known as the Hermetian Yang-Mills (HYM) equations. From the Donaldson-Uhlenbeck-Yau theorem, they are guaranteed to have a solution given the underlying bundle is a holomorphic and poly-stable vector bundle V over the CY with vanishing slope [57, 58]. That is, the bundle has holomorphic transition functions, and the slope, defined by

$$\mu(V) = \frac{1}{\mathrm{rk}(V)} \int_X c_1(V) \wedge J \wedge J, \tag{1.33}$$

is equal to zero, where $\mathrm{rk}(V)$ and $c_1(V)$ are the rank and first Chern class of the bundle respectively. A bundle is stable if the slope of any sub-sheaf of V with non-zero rank is strictly less than $\mu(V)$. Poly-stability of a bundle means that it is a direct sum of stable bundles where all elements of the sum have equal slope.

The trivial solution to (1.32), subject to the constraint (1.31), now referred to as the "standard embedding", is to set $V = TX$. This was the original proposal for heterotic phenomenology proposed in [6]. Since then, it has become clear that this is very limited and it is difficult to reproduce the spectrum of the standard model in this context. Therefore, there has been a move towards more complicated constructions for V [53]. We should also note that the condition (1.31) can be relaxed by including NS5-branes, which magnetically couple to H. This generalised version of the anomaly condition is given by

$$\mathrm{c}_2(V) - \mathrm{c}_2(TX) = [\mathcal{C}]. \tag{1.34}$$

where $[\mathcal{C}] \in H^4(X)$ is the class dual to an effective holomorphic curve wrapped by five-branes which also span across the four un-compactified dimensions.

To summarise this subsection, given a Calabi-Yau threefold X and holomorphic and slope poly-stable vector bundle V with vanishing slope over X, we can construct a 4D $\mathcal{N} = 1$ chiral gauge theory from heterotic string theory – the first step in constructing the MSSM from string theory. Further to this, due to the existence theorems of Calabi, Donaldson, Uhlenbeck and Yau, these conditions can be checked at the level of topology, rather than needing to construct the explicit gauge field solution or Ricci-flat metric. In Sec. 1.3.3, we will find that we can go even further. Specifically, by diligent choice of CY manifold and holomorphic slope-stable vector bundle we can engineer the 4D theory to have the same charged spectrum as the MSSM. This comes at the cost of a number of extra chiral multiplets, uncharged under the SM gauge group, known as moduli. In principle, the vevs of the scalar components of the multiplets become the couplings of the low energy theory. The process of generating a potential for these fields is know as moduli stabilisation, and is one several topics in the following subsection. As with the gauge group, the correct matter spectrum can be achieved at the level of topological conditions[5], but finer details such as the values of couplings in the 4D theory also require knowledge of the geometry.

As the resulting four-dimensional theory is an $\mathcal{N} = 1$ SUGRA theory, we know from our discussions in Sec. 1.2.1 that the resulting theory has a very constrained form. In particular, once the spectrum is obtained, the action is completely specified by the superpotential W, Kähler potential K and gauge kinetic function f_{ab}.

1.3.3 The Four Dimensional Action and The Model Builder's Checklist

Setting the ambitious goal of embedding the standard model into string theory, there are three key milestones[6]:

- A compactification that results in the massless particle spectrum of the standard model. We include the quarks and leptons as massless at this stage. By massless, we mean those that are massless above the electroweak phase transition.

---

[5] There is a slight caveat here, in that the number of massless particles can increase for special values of moduli. So these quantities are often referred to as quasi-topological, as there values are fixed for generic values of moduli.

[6] In no particular order.

- Once the field content of the 4D theory is determined, one can turn to the form of the action. In particular, we would like to generate Yukawa couplings that lead to the observed masses and mixing.
- Generically, string compactifications lead to a number of massless scalar fields, called moduli. The potential for these scalars have no minima. We need to generate extra terms in the potential that stabilises these moduli.

In this section we will discuss each of the above in turn in the context of the heterotic string on a smooth CY X with holomorphic slope-stable vector bundle V. While this thesis is focused on obtaining the correct particle spectrum, the other two goals will be included for completeness.

Particle Spectrum

Once we have constructed a given vector bundle V, the low energy spectrum is given by the dimensions of particular cohomology classes of that bundle. To see why, it is useful to begin by considering the the compactification of a D dimensional field theory containing a scalar on some $d = D - 4$ dimensional compact manifold. Given a massless scalar in the high dimensional theory $\varphi(x^M)$, the equation of motion is given by

$$\Delta_{(D)}\varphi = (\Delta_{(d)} + \Delta_{(4)})\varphi = 0, \tag{1.35}$$

where $\Delta_{(D)}$, $\Delta_{(d)}$ and $\Delta_{(4)}$ are the Laplacian on the D, d and 4 dimensional spaces respectively.

Applying separation of variables between the coordinates of the external (x^μ) and internal (y^i) spaces, we have

$$\varphi(x, y) = \sum_n \phi_n(x)\chi_n(y), \tag{1.36}$$

where χ_n are the eigen-functions of the laplacian $\Delta_{(d)}\chi_n = \lambda_n\chi_n$. Plugging this into the equation of motion, we have a family of scalar fields in the low energy theory which satisfy

$$(\Delta_{(4)} + \lambda_n)\phi_n(x) = 0 \ \forall n. \tag{1.37}$$

In particular, the number of massless fields is given by the number of zero modes of the Laplacian of the compact space[7]. This, by Hodge's theorem, is a topological quantity given by the dimension of the zeroth cohomology of the tangent bundle $b^0 = \dim(H^0(X))$ [49, 53]. By considering the compactification of different p-forms, we also need to consider the dimensions of higher cohomologies of the tangent bundle.

As the eventual four dimensional theory has $\mathcal{N} = 1$ SUSY, we will need only consider the reduction of the bosonic fields. The fermionic content can be implied by noting that all four dimensional scalar and vector fields are required to parts of chiral and vector supermultiplets respectively. We first consider the reduction of the gauge fields, before moving onto the 10 dimensional degrees of freedom.

For gauge fields, the story is a bit more complicated than the scalar example. We need to specify which subgroup of $E_8 \times E_8$ is used as the structure group of V before looking at the resulting massless spectrum. The low energy gauge group will be given by the commutant of this structure group within $E_8 \times E_8$. As the standard model gauge group is a subgroup of a single E_8, we will attempt to construct models of particle physics by only making use of one of the two E_8 factors. The second E_8 factor will lead to a dark sector in the low energy theory, coupled to the standard model by gravitational interactions only. Within this context, the simplest way of satisfying the vanishing slope condition is to set $c_1(V) \propto \mathrm{tr}(F)$ to zero. This is true in the case of bundles with $SU(N)$ structure-group. Given structure-group $H = SU(3), SU(4), SU(5)$ and $SU(6)$, the low energy gauge group is, up to discrete global factors, $G = E_6, SO(10), SU(5)$ and $SU(3) \times SU(2)$ respectively. We recognise the first three of these as the GUT groups introduced in Sec. 1.2.1. For the fourth option, there have been attempts to split the $SU(6)$ bundle into a maximal subgroup, such that an extra $U(1)$ is present in G. This leads to exactly the standard model gauge group in 4D. However, these are extremely restrictive and so we will not consider them [59].

---

[7] The heavy modes would have masses of $\mathcal{O}(\mathcal{V}^{-2/d})$, where $\mathcal{V}$ is the volume of the compact space. We can choose $\mathcal{V}$ such that these heavy modes are sufficiently heavy to be ignored in the low energy theory.

Taking the ten dimensional gauge field, we perturb the gauge field in the compact directions

$$\delta A_M(x,y)dx^M = \sum_I \varphi^I(x)\nu_I(y), \tag{1.38}$$

where the $\nu_I(y)$ are some bundle-valued forms which satisfy conditions such that φ^I will be massless in the 4D theory. From the equations of motion (1.32), one can show the massless condition is

$$D_i\nu_I(y) = 0, \tag{1.39}$$

where D_i is the gauge covariant derivative of the bundle. As a result, we are looking for zero-modes of a different connection, and the low energy spectrum is given by the cohomologies of the bundle V. Specifically, on branching under the three choices of H leading to the 4D GUTs, we have [53]

$$\begin{aligned}
&\mathbf{248} \xrightarrow{E_6\times SU(3)} \underbrace{(\mathbf{78},\mathbf{1})}_{1} \oplus \underbrace{(\mathbf{27},\mathbf{3})}_{h^1(V)} \oplus \underbrace{(\overline{\mathbf{27}},\overline{\mathbf{3}})}_{h^2(V)} \oplus \underbrace{(\mathbf{1},\mathbf{8})}_{h^1(V\otimes V^*)}\ , \\
&\mathbf{248} \xrightarrow{SO(10)\times SU(4)} \underbrace{(\mathbf{45},\mathbf{1})}_{1} \oplus \underbrace{(\mathbf{16},\mathbf{4})}_{h^1(V)} \oplus \underbrace{(\overline{\mathbf{16}},\overline{\mathbf{4}})}_{h^2(V)} \oplus \underbrace{(\mathbf{10},\mathbf{6})}_{h^1(\wedge^2 V)} \oplus \underbrace{(\mathbf{1},\mathbf{15})}_{h^1(V\otimes V^*)}\ , \\
&\mathbf{248} \xrightarrow{SU(5)\times SU(5)} \underbrace{(\mathbf{24},\mathbf{1})}_{1} \oplus \underbrace{(\mathbf{5},\overline{\mathbf{10}})}_{h^1(\wedge^2 V^*)} \oplus \underbrace{(\overline{\mathbf{5}},\mathbf{10})}_{h^1(\wedge^2 V)} \oplus \underbrace{(\mathbf{10},\mathbf{5})}_{h^1(V)} \oplus \underbrace{(\overline{\mathbf{10}},\overline{\mathbf{5}})}_{h^2(V)} \oplus \underbrace{(\mathbf{1},\mathbf{24})}_{h^1(V\otimes V^*)}\ .
\end{aligned} \tag{1.40}$$

Therefore, if we engineer a vector bundle V on a CY with structure group $SU(5)$, such that $h^1(V) = 3$, $h^2(V) = 0$ and $h^1(\wedge^2 V) = 1$, we obtain the charged spectrum of an $SU(5)$ GUT from string theory with $h^1(V \otimes V^*)$ extra uncharged chiral multiplets. All of these resulting chiral multiplets are referred to as matter fields and we will label the scalar components of these with φ^I.

We still need to break this GUT group to the standard model – we achieve this with discrete Willson lines. If our Calabi-Yau manifold is not simply connected (i.e. $\pi(X) \neq 0$), then there exist field configurations over the CY with no moduli and vanishing field strength [49, 53]. Mathematically, we say that these live in the torsion component of the cohomology [60, 61]. In the low energy theory, this can be engineered to break to the standard model [62–64], as only the part of the gauge group that commutes with the Wilson line remains.

It is important to realise that using discrete Wilson lines to break a gauge group is distinct to traditional spontaneous breaking. In particular, the modes that would become heavy under spontaneous breaking are instead projected out of the theory entirely. Taking our previous example, underneath (1.40), if an appropriate discrete Wilson line is introduced to break to the standard model gauge group then there is no $SU(5)$ phase of the low energy theory. Additionally, the states we would usually interpret as X and Y bosons are removed, but massive higher modes with the same quantum numbers remain and could lead to fast proton decay [49].

The gauge fields are not the only bosonic fields in the 10 dimensional theory. We also need to consider the fluctuations of the Calabi-Yau metric $g_{i\bar{j}}$, the B field and the dilaton ϕ. These appear as scalar components of various chiral supermultiplets of the 4D theory. Specifically, they form three distinct types of chiral multiplets: Kähler moduli T^i, complex structure moduli Z^α and the S modulus. The geometric origin of the first two is outlined in Sec. 1.4.1, but they can be intuitively thought of as the sizes of various cycles in the CY, while the universal S modulus is defined by

$$S(x) = e^{-2\phi(x)} + ib(x), \tag{1.41}$$

where $b(x)$ is the magnetic dual of $B_{\mu\nu}(x)$. That is the 4D field strengths are Hodge dual $\star_4(d_4 B) = d_4 b$.

Low Energy Couplings and Interactions

Once the correct spectrum is obtained, one can turn to the finer details of the low energy theory – namely, the couplings and interactions. These will be given in terms of the functions $W(\varphi^I, Z^\alpha, T^i, S)$, $K(\varphi^I, \overline{\varphi^I}, Z^\alpha, \overline{Z^\alpha}, T^i, \overline{T^i}, S, \overline{S})$ and $f_{ab}(\varphi^I, Z^\alpha, T^i, S)$ – the functional dependence will be dropped from now on for brevity. As was the case in the previous section, due to the constaining supersymmetric structure, we only need consider the reduction of the bosonic fields.

For the S, T, and Z moduli one finds, from the dimensional reduction, the following contributions to the Kähler potential

$$K \ni \frac{K_{mod}}{M_{pl,4}^2} = -\underbrace{\log\left(\int_X J \wedge J \wedge J\right)}_{K_T(T,\overline{T})} - \underbrace{\log\left(-i\int_X \Omega \wedge \overline{\Omega}\right)}_{K_Z(Z,\overline{Z})} - \underbrace{\log(S+\overline{S})}_{K_S(S,\overline{S})}, \qquad (1.42)$$

where J and Ω are the forms introduced in the previous subsection, $M_{pl,4}$ is the 4 dimensional Planck mass, and only the complex structure moduli appear in the superpotential perturbatively [65, 66]. However, non-perturbative contributions can introduce S and T into the superpotential. For example, terms of the form

$$\begin{aligned} W &\sim e^{-T} g(\varphi, Z), \\ W &\sim e^{-S} g(\varphi, Z), \end{aligned} \qquad (1.43)$$

where $g(\phi, Z)$ is a function of complex structure and of the scalar components of the matter multiplets, can be introduced by Euclidean string instantons [67–70] and gaugino condensation of the hidden E_8[71–73] respectively.

The gauge kinetic function is, as expected from our GUT-like structure, universal across all the gauge bosons $f_{ab} = (S + f^{\text{th.}}(T))\delta_{ab}$, where $f^{\text{th.}}(T)$ is model dependant. The universal proportionality constant being the S modulus is not unexpected, as $\text{Re}(S) = g_{st}^2$.

The interactions between the matter fields can be somewhat more complicated. Focusing on the perturbative Yukawa couplings, as these are of particular interest to particle physics, we require information about the cubic term in the superpotential

$$W = \lambda_{IJK} \varphi^I \varphi^J \varphi^K + \cdots , \qquad (1.44)$$

the moduli Kähler potential (1.42), and the matter field contribution to the Kähler potential K_{mat}.

After dimensional reduction, one finds the following expression for the cubic terms appearing in the superpotential

$$\lambda_{IJK}(Z) \sim \int_X \nu_I \wedge \nu_J \wedge \nu_K \wedge \Omega, \qquad (1.45)$$

where ν_I are the harmonic bundle-valued (0,1)-forms on the Calabi-Yau corresponding to the low energy fields φ^I, as described in Sec. 1.3.3. While numerical methods are frequently required to compute these integrals the terms are quasi-topological in the sense that they do not require the Ricci-flat metric to compute [74–76]. These are the perturbative contributions to the superpotential, and therefore only depend on the complex structure moduli. However, non-perturbative effects, such as those in (1.43), bring in dependence on the S and T moduli and also contribute to the Yukawa couplings if $g(\varphi, Z)$ has a term cubic in φ^I.

Also from dimensional reduction of the 10 dimensional gauge fields, one finds the following expression for the Kähler potential for the matter fields

$$K \ni K_{mat} \sim \varphi^I \overline{\varphi}^{\overline{J}} \int_X \nu_I \wedge \star_V \nu_{\overline{J}}. \tag{1.46}$$

The Hodge star in this expression brings in an explicit dependence on the Ricci-flat metric. Recent work involving numerical metrics [77–84] has been used to compute these integrals in order to find the physical Yukawa couplings [85]. At the present time these calculations are computationally expensive, so thorough explorations of the moduli space have not taken place. However, they are now in principle possible.

Moduli Stabilisation

Moduli stabilisation has proven to be a contentious and complicated issue for string theory, especially in the context of the heterotic string. For a recent review see [86].

The main difficulty in moduli stabilisation is that generally, except possibly in IIA compactifications [87], non-perturbative and perturbative contributions need to compete with each other. For example, in type-IIB, the KKLT mechanism relies on finding a choice of fluxes such that the perturbative contributions are unusually small, to compete with non-perturbative effects [88]. The Large Volume Scenario (LVS) instead relies on certain non-perturbative contributions being unusually large, to compete with perturbative effects [89].

Turning to the heterotic string, things are far more difficult. One can typically stabilise the complex structure and bundle moduli with NS-flux (that is, turning on the H_{ijk} that we set to zero in Sec. 1.3.2), but this moves us away from the CY geometry and sometimes leads to domain wall solutions in 4D rather than those with maximal symmetry [90][8]. One then uses non-perturbative effects, like (1.43), to stabilise the Kähler moduli and the dilaton [29].

Another approach is to use the bundle, and supersymmetry conditions, to stabilise moduli [91, 92]. To be more specific, one of the terms from the 10D action (1.15) that contributes to the scalar potential, after some unpleasant manipulations, is of the form

$$S = -\frac{\alpha'}{2\kappa_{10}^2} \int d^{10}x \sqrt{-g} \left\{ -\frac{1}{2} \operatorname{tr} \left(g^{i\bar{j}} F_{i\bar{j}} \right)^2 + \operatorname{tr} \left(g^{i\bar{i}} g^{j\bar{j}} F_{ij} F_{\bar{i}\bar{j}} \right) \right\} + \dots , \tag{1.47}$$

which we notice is made up from the same objects from the supersymmetry conditions (1.32). Clearly at the supersymmetric locus these terms vanish, but any perturbations away from the supersymmetric locus will give rise to a positive contribution to the scalar potential. From this point of view, these two terms are associated with D-term and F-term contributions to the scalar potential respectively. Crucially, the first SUSY constraint in (1.32) often only holds for certain choices of Kähler moduli (and the dilaton when one considers strong coupling), while the second and third only hold for certain values of complex structure. This has the effect of stabilising many of the geometric moduli and the dilaton. At this stage the resulting moduli space is supersymmetric and Minkowski, without the use of flux. The remaining flat directions, including the bundle moduli, of the potential can potentially be stabilised with non-perturbative corrections. At the end of this construction, one is typically has a supersymmetric AdS minimum.

1.4 Some Algebraic Geometry

In the previous two sections, we demonstrated that compactification of heterotic string theory on CYs with holomorphic and slope-stable vector bundles leads to an $\mathcal{N} = 1$ gauged SUGRA theory. In such a theory, at least in principle, we can engineer the MSSM charged spectrum with appropriate choice of bundle and manifold.

---

[8] This is not necessary a problem. The hope of the authors of Ref. [90] is that maximal symmetry is again a solution, once non-perturbative effects are introduced to stabilise the other moduli.

In this section, we present the mathematics of how one can construct both CY manifolds and the simplest of bundles over them – line bundles. These line bundles will serve as the basis for constructing more complicated holomorphic slope-stable bundles in Chapters 3 and 4.

1.4.1 Calabi-Yau Manifolds and Complete Intersection Calabi-Yau Manifolds (CICYs)

As mentioned in Sec. 1.3.2, CY 3-folds have several definitions, but for our purposes they are Kähler manifolds with vanishing first Chern class. We are guaranteed, by Yau's theorem, to have a unique Ricci flat metric in every Kähler class. This unique metric also has, at most, $SU(3)$ holonomy. We also stated that such manifolds admit the closed and non-exact (1,1)-form J and (3,0)-form Ω. By definition, these are representatives of elements of the cohomology classes $H^{1,1}(X)$ and $H^{3,0}(X)$ respectively, and it turns they contain all the necessary information to construct the Ricci-Flat metric.

One particular property of CYs, with $SU(3)$ holonomy, is that their cohomology structure is highly restricted. In particular, there are only two independent Hodge numbers (the dimensions of the Dolbeault cohomology classes): $h^{1,1}$ and $h^{2,1}$. Written as a "Hodge diamond", the Dolbeault cohomology of for a CY manifold X is

$$\begin{matrix} & & & h^{0,0} & & & \\ & & h^{1,0} & & h^{0,1} & & \\ & h^{2,0} & & h^{1,1} & & h^{0,2} & \\ h^{3,0} & & h^{2,1} & & h^{1,2} & & h^{0,3} \\ & h^{3,1} & & h^{2,2} & & h^{1,3} & \\ & & h^{3,2} & & h^{2,3} & & \\ & & & h^{3,3} & & & \end{matrix} \;=\; \begin{matrix} & & & 1 & & & \\ & & 0 & & 0 & & \\ & 0 & & h^{1,1} & & 0 & \\ 1 & & h^{2,1} & & h^{2,1} & & 1 \\ & 0 & & h^{1,1} & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{matrix} \;, \tag{1.48}$$

which also implies the Euler character of the CY is

$$\chi(X) = \sum_{n=0}^{3}(-1)^n b^n = \sum_{n=0}^{3}(-1)^n \left(\sum_{p+q=n} h^{p,q}\right) = 2(h^{1,1} - h^{2,1}), \tag{1.49}$$

where b^n are the Betti-numbers (the dimensions of the de Rahm cohomology classes).

From the discussions in Sec. 1.3.3, it may be unsurprising that the dimensions of these cohomology classes are related to fluctuations of the metric. Clearly $h^{1,1}$ counts the deformations of J, which correspond to Kähler moduli, and similarly $h^{2,1}$ counts the deformations of Ω, which are the complex structure moduli[9].

Antother topological quantity we make frequent use of is the set of triple intersection numbers of the CY. Given a basis $\{J_i : i = 1, .., h^{1,1}\}$ for $H^{1,1}(X)$ the triple intersection numbers are given by

$$d_{ijk} = \int J_i \wedge J_j \wedge J_k. \tag{1.50}$$

We now move onto the realisation of explicit CYs. There are many different constructions in the literature [53, 93–95], but the largest database of examples comes from toric geometry and results in approximately $\mathcal{O}(10^{400})$ different examples [16, 18, 96]. However, it turns out that these are not the most practical examples for heterotic model building as we can only identify discrete symmetries (for introducing appropriate Wilson lines) in a handful of examples [97]. As a result, we instead consider the smaller database of Complete Intersection Calabi-Yau Manifolds (CICYs) [98, 99]. There are 7890 manifolds in this database[10] and their discrete symmetries have been classified [100, 101].

Three simple example CICYs are called the quintic X_Q, the tetraquadric X_{TQ} and the bicubic X_{BC}. The qunitic is constructed by the vanishing of a 5th order polynomial in the homogeneous coordinates of $\mathbb{P}^4$, the tetraquadric from the vanishing of a polynomial of degree $(2, 2, 2, 2)$ in the coordinates of 4 $\mathbb{P}^1$s and the bicubic from the vanishing of a degree $(3, 3)$ polynomial in the coordinates of 3 $\mathbb{P}^2$s. All of these are called "favourable", meaning that the second cohomology descends directly from the ambient space. Since $h^{1,1}(\mathbb{P}^n) = 1\,\forall n$, we have $h^{1,1} = 1$ for the quintic, $h^{1,1} = 4$ for the tetraquadric, and $h^{1,1} = 2$ for the bicbuic. The complex-structure is encoded by the coefficients of the defining polynomials.

---

[9] In the low energy theory, both of these become complex scalars by combining with appropriate reductions of the B field.

[10] Though some entries of the database are degenerate descriptions of the same manifold.

We will frequently refer to the CICYs by the defining matrices. For the above three examples, these are:

$$X_{BC} = \begin{array}{c} \mathbb{P}^2 \\ \mathbb{P}^2 \end{array} \left[\begin{array}{c} 3 \\ 3 \end{array} \right], \quad X_{TQ} = \begin{array}{c} \mathbb{P}^1 \\ \mathbb{P}^1 \\ \mathbb{P}^1 \\ \mathbb{P}^1 \end{array} \left[\begin{array}{c} 2 \\ 2 \\ 2 \\ 2 \end{array} \right], \quad X_Q = \mathbb{P}^4 \left[\, 5 \, \right], \tag{1.51}$$

where each row refers to the degree that the polynomial is in with respect to the homogeneous coordinates of each projective space.

As an example we focus on the quintic X_Q. The simplest smooth defining polynomial (i.e. the simplest choice of complex structure) is

$$P = X_0^5 + X_1^5 + X_2^5 + X_3^5 + X_4^5, \tag{1.52}$$

where X_n are the homogenous coordinates of $\mathbb{P}^4$. The quintic with this choice of complex structure is called the Fermat-Quintic. Taking the Fubini-Study metric on $\mathbb{P}^4$

$$g^{(FS)}_{a\bar{b}} = \partial_a \partial_{\bar{b}} \log \left(|X_0|^2 + |X_1|^2 + |X_2|^2 + |X_3|^2 + |X_4|^2 \right), \tag{1.53}$$

and pulling back to the CY manifold ($P = 0$), we find the Ricci 2-form $\mathcal{R}$ is exact. As a result, the first Chern class $c_1(X) \propto [\mathcal{R}] = 0$ and so we know, by Yau's theorem, that the resulting space is Calabi-Yau. However, the Ricci-flat metric is not $g^{(FS)}|_{P=0}$, but both metrics live in the same Kähler class.

Given a generic CICY defining matrix

$$X = \begin{array}{c} \mathbb{P}^{n_1} \\ \mathbb{P}^{n_2} \\ \vdots \\ \mathbb{P}^{n_p} \end{array} \left[\begin{array}{cccc} m_1^1 & m_1^2 & ... & m_1^q \\ m_2^1 & m_2^2 & ... & m_2^q \\ \vdots & \vdots & ... & \vdots \\ m_p^1 & m_p^2 & ... & m_p^q \end{array} \right], \tag{1.54}$$

each row corresponds to a separate polynomial, and we consider the complete intersection of all q polynomials. In this general context, the CY condition ($c_1(X) = 0$), is equivalent to

$$n_i = 1 + \sum_{j=1}^{q} m_i^j \quad \forall i = 1, 2...p. \tag{1.55}$$

These Calabi-Yau manifolds have a number of computational advantages. For one, the ambient spaces are particularly simple and have been extensively studied by mathematicians [102, 103]. These CYs also have an algebraic definition, which opens the possibility of using the powerful techniques and tools of algebraic geometry.

At this stage, all the CICYs are simply connected, obstructing us from introducing discrete Wilson lines. However, there are a large number of discrete freely acting symmetries on these manifolds [100, 101]. Given a CICY X, with a freely-acting symmetry Γ, we can construct another smooth Calabi-Yau manifold $\hat{X} = X/\Gamma$ which is no longer simply connected. In practice, we will mostly work with the "upstairs manifolds" X, then take the quotient at the end of the calculation. As an example of such a freely-acting symmetry, consider the Fermat-Quintic and the $\mathbb{Z}_5$ transformation

$$X_n \to e^{2\pi i n/5} X_n \ \forall n. \tag{1.56}$$

Under this transformation, the defining equation (1.52) remains smooth for all point on the CY. As a result, this transformation generates a freely acting $\mathbb{Z}_5$ symmetry.

1.4.2 Line bundles and Cohomology

Constructing bundles over Calabi-Yau manifolds is a complicated task. Our approach will be to build bundles out of simpler objects – namely line bundles. These are Abelian bundles, and have been studied by mathematicians in great detail. In particular, line bundles are classified by their first Chern class $c_1(L) = k^i J_i \in H^2(X, \mathbb{Z})$, where J_i is a basis of $H^2(X)$. As a result, we often denote line bundles by this integer vector $L = \mathcal{O}_X(\mathbf{k})$. In this subsection, we introduce some basic properties of line bundles, but leave the details of using them to construct larger bundles to Chapters 3 and 4.

Given a Calabi-Yau 3-fold X, with second chern class $c_2(X) = c_{2i}(X)J^i$, and intersection numbers d_{ijk}, the index of a line bundle $L = \mathcal{O}_X(\mathbf{k})$ on X, is given by

$$\chi(L) = \sum_{q=0}^{3}(-1)^q h^q(L) = \frac{1}{2}\int_X c_3(X) = \frac{1}{6}c_{2i}(X)k^i + \frac{1}{12}d_{ijk}k^i k^j k^k, \tag{1.57}$$

where $c_3(X)$ is the third Chern class of X.

The index of a line bundle is a simple quantity to calculate, but the individual cohomolgies $h^q(L)$ are much more complicated. In practice, this is done by making use of the ambient space [54, 93]. In particular, the Bott formula tells us the dimension of line bundles cohomologies over projective space:

$$h^q(\mathbb{P}^n, \mathcal{O}(\mathbf{k})) = \begin{cases} \begin{pmatrix} k+n+1 \\ 0 \end{pmatrix}\begin{pmatrix} k+n \\ n \end{pmatrix} & q=0 \qquad k>-1, \\ 1 & q=n \qquad k=-n-1, \\ \begin{pmatrix} -k-1 \\ -k-n-1 \end{pmatrix}\begin{pmatrix} -k-n-2 \\ 0 \end{pmatrix} & q=n \qquad k<-n-1, \\ 0 & \text{otherwise} \quad . \end{cases} \tag{1.58}$$

Using this as a starting point, the cohomology of line bundles on products of projective spaces can be derived from the Künneth formula

$$H^n\left(\mathbb{P}^{n_1} \times \ldots \mathbb{P}^{n_m}, \mathcal{O}\left(q_1, \ldots q_m\right)\right) = \bigoplus_{k_1+\ldots k_m=n} H^{k_1}\left(\mathbb{P}^{n_1}, \mathcal{O}\left(q_i\right)\right) \times \ldots H^{k_m}\left(\mathbb{P}^{n_m}, \mathcal{O}\left(q_m\right)\right). \tag{1.59}$$

Given a line bundle over the projective ambient space, we can pull it back to define a line bundle on the CY. As we know the cohomology on the ambient space, we can use techniques from algebraic geometry to calculate the cohomology of the line bundle on the CY. Let us quickly go over how this done for the simplest case. For the case of a codimension one CICY X in ambient space $\mathcal{A}$, the following is an exact sequence (that is, the image of each map in the sequence is in the kernel of the following map) called the Koszul sequence [11]

$$0 \to \mathcal{N}^* \otimes \mathcal{O}_{\mathcal{A}}(\mathbf{k}) \to \mathcal{O}_{\mathcal{A}}(\mathbf{k}) \to \mathcal{O}_X(\mathbf{k}) \to 0, \tag{1.60}$$

where $\mathcal{N}^*$ is dual to the normal bundle $\mathcal{N}$ of X, which given a general CICY X with defining matrix (1.54) is given by

$$\mathcal{N} = \bigoplus_{n=1}^{q} \mathcal{O}(m_1^n, m_2^n, \ldots m_p^n). \tag{1.61}$$

---

[11] We note that line bundle entries are additive under a tensor power. For example $\mathcal{O}_X(\mathbf{k}) \otimes \mathcal{O}_X(\mathbf{p}) = \mathcal{O}_X(\mathbf{k}+\mathbf{p})$, and the trivial bundle is given by $\mathcal{O}_X = \mathcal{O}_X(\mathbf{0})$. As a result, the dual of a line bundle is another line bundle $\mathcal{O}_X(\mathbf{k})^* = \mathcal{O}_X(-\mathbf{k})$.

The Koszul sequence implies a corresponding long exact sequence in cohomology [102, 103]:

$$\begin{aligned} 0 \to & H^0(\mathcal{N}^* \otimes O_{\mathcal{A}}(\mathbf{k})) \to H^0(O_{\mathcal{A}}(\mathbf{k})) \to H^0(O_X(\mathbf{k})) \to \\ & H^1(\mathcal{N}^* \otimes O_{\mathcal{A}}(\mathbf{k})) \to H^1(O_{\mathcal{A}}(\mathbf{k})) \to H^1(O_X(\mathbf{k})) \to \\ & H^2(\mathcal{N}^* \otimes O_{\mathcal{A}}(\mathbf{k})) \to H^2(O_{\mathcal{A}}(\mathbf{k})) \to H^2(O_X(\mathbf{k})) \to \\ & H^3(\mathcal{N}^* \otimes O_{\mathcal{A}}(\mathbf{k})) \to H^3(O_{\mathcal{A}}(\mathbf{k})) \to H^3(O_X(\mathbf{k})) \to 0 \end{aligned} , \tag{1.62}$$

which is the tool used to calculate the cohomology of line bundles on the CY.

Consider the line bundle $\mathcal{O}_{X_Q}(1)$ on the quintic, which has normal bundle $\mathcal{N} = \mathcal{O}_{X_Q}(5)$. Using the Bott-formula, we can fill in the dimensions of the first two columns of the long exact sequence above and find

$$\begin{array}{c|ccccc} & \mathcal{O}_{\mathbb{P}^4}(-4) & \to & \mathcal{O}_{\mathbb{P}^4}(1) & \to & \mathcal{O}_{X_Q}(1) \\ h^0 & 0 & \to & 5 & \to & ? \\ h^1 & 0 & \to & 0 & \to & ? \\ h^2 & 0 & \to & 0 & \to & ? \\ h^3 & 0 & \to & 0 & \to & ? \end{array} . \tag{1.63}$$

Therefore, from the exactness of the sequence, we conclude $h^0(\mathcal{O}_{X_Q}(1)) = 5$ while the other cohomolgies of $\mathcal{O}_{X_Q}(1)$ vanish. This was clearly a particularly simple example as we can easily read off the dimensions of the cohomology. If, however, the exact sequence had some subsequence of the form

$$0 \to ? \to 5 \to 3 \to ? \to 0, \tag{1.64}$$

we could only conclude, from the above arguments, that the difference between the two unknown entries is $2 = 5-3$. In these cases, one needs to explicitly construct the cohomologies (as opposed to just their dimensions), the maps in the Koszul sequence and the maps in the corresponding long exact sequence. Such calculations use methods from commutative algebra and computational algebraic geometry [104–106].

One major hurdle is that these methods from computational algebraic geometry are computationally expensive, and frequently have exponential complexity or worse. However, recently, through some explorations with machine learning techniques, explicit formulas for the line bundle cohomolgies on Calabi-Yaus have been found. These formulae involve a partition of the space of line bundles on a CY into disjoint regions, and in each region the cohomology function is polynomial [21, 22, 64, 107–114]. For example, on the bicubic manifold X_{BC} the zeroth cohomology of line bundles have dimensions given by:

region	$h^0\left(X, \mathcal{O}_{X_{BC}}\left(k_1, k_2\right)\right)$
$k_1 > 0, k_2 > 0$	$\frac{3}{2}\left(k_1 + k_2\right)\left(2 + k_1 k_2\right)$
$k_1 > 0, k_2 = 0$	$\frac{1}{2}\left(1 + k_1\right)\left(2 + k_1\right)$
$k_1 = 0, k_2 > 0$	$\frac{1}{2}\left(1 + k_2\right)\left(2 + k_2\right)$
$k_1 = k_2 = 0$	1
otherwise	0

(1.65)

1.5 Computational Methods

Up to this point, we have discussed the standard model and laid the path for how to realise it in string theory. However, as research in this area has progressed, the vastness of the string landscape has become abundantly clear. Focusing on perturbative IIB flux vacua, early estimates suggest $\mathcal{O}(10^{500})$ different possibilities, while considerations of the strong coupling regime has increased this to $\mathcal{O}(10^{272{,}000})$ [14, 15]. With the plethora of string dualities, it should be expected that similar numbers would be present for the heterotic string.

To resolve this issue, we now turn to the modern computational methods that have shown remarkable ability in navigating other large landscapes: namely, genetic algorithms (GA) and reinforcement learning (RL) in Sec. 1.5.1 and Sec. 1.5.2 respectively. The remainder of this thesis will be on the application of these methods to regions of heterotic landscape, in search of the standard model. Such computational methods have proven incredibly effective at exploring the kind of parameter spaces that occur in physics [20–24, 115–121]

1.5.1 Genetic Algorithms (GA)

Genetic algorithms are an incredibly simple heuristic method of solving discrete optimisation problems, inspired by the Darwinian theory of evolution and survival of the fittest [122, 123].

To introduce GA we first need to define a 'creature' or 'individual'. All the information of interest regarding a single creature can be encoded by its 'genotype', which one should intuitively think of as the genetic code of the creature. There is a large amount of choice in how to express the genotype, and one just needs to be able to define the ideas of mutation of a single genotype and breeding of two genotypes. For our purposes, we will always take the genotype to be a string of binary digits.

The mutation of a single binary string will be represented by the random flipping of a bit in the string. The probability p of a single bit mutating is a parameter of the algorithm. Given two binary strings, breeding will be defined by crossover, where both strings are cut at a random point and exchange bits around that point. For example:

$$\begin{aligned}
&\{01000111101010101\} \to \{01000111101010101\} \to \{01000111101101111\} \\
&\{10011101011101111\} \to \{10011101011101111\} \to \{10011101011010101\}.
\end{aligned}$$

We also have a map from the genotype to the 'phenotype', which carries the properties of interest of each creature, including a real number called the 'fitness'. The aim of GA is to find the creatures with the maximum fitness. For our purposes, each creature will correspond to a string model and the phenotype will be the properties of that model. The fitness will be a measure of how far away a particular model is from a realisation of the MSSM.

With the phenotype, genotype, fitness, mutation and breeding defined we can proceed with GA, which is summarised by the following pseudocode:

```
pop = RandomPop(N) #Initialise random population of size N
pop = SortPop(pop) #Sort pop according to fitness

#Loop over M generations
FOR i = 1 to M DO
        newPop = []
        #Loop over new population
        FOR j = 1 to N/2 DO
                #Draw two random creatures from the population
                creature1 = RandomDraw(pop)
                creature2 = RandomDraw(pop)
```

```
                #Breed the two creatures with bit crossover
                [creature3, creature4] = Breed(creature1, creature2)

                #Apply mutation to both creatures via random
                #bit-flips
                creature3 = Mutate(creature3)
                creature4 = Mutate(creature4)

                #Append the new creatures to the population
                newPop.Append(creature3)
                newPop.Append(creature4)
        END LOOP

        #If elitism is true copy over best from last pop
        IF elitism THEN
                newPop[0] = pop[0]
        END IF

        pop = SortPop(newPop)
END LOOP
```

We begin with a random set of N of creatures $\{c_k : k = 1, 2...N\}$, each drawn from a flat distribution and rank them by their fitness $f(c_k) \leq f(c_{k-1}) \, \forall k$. We now draw pairs of individuals from this population, for breeding, via bit cross-over. The probability of selecting the k'th individual is given by

$$P_k = \frac{2}{(1+\alpha)N} \left(1 + \frac{N-k}{N-1}(\alpha - 1) \right), \tag{1.66}$$

where α is a parameter of the algorithm. The key point of the above formula is that the fittest creature will be selected for breeding multiple times, while the least fit individual will be selected few times, if at all. The two new creatures formed from breeding then mutate by random bit-flips.

We continue drawing pairs for breeding from the population, followed by mutation, until we are left with a new population of size N. One may also allow for "elitism", where the fittest creature from the previous population is copied over to the new population.

This is repeated over many generations and one should find, given diligent choices of parameters, that the average fitness of population and the fitness of the fittest individual increases over the generations.

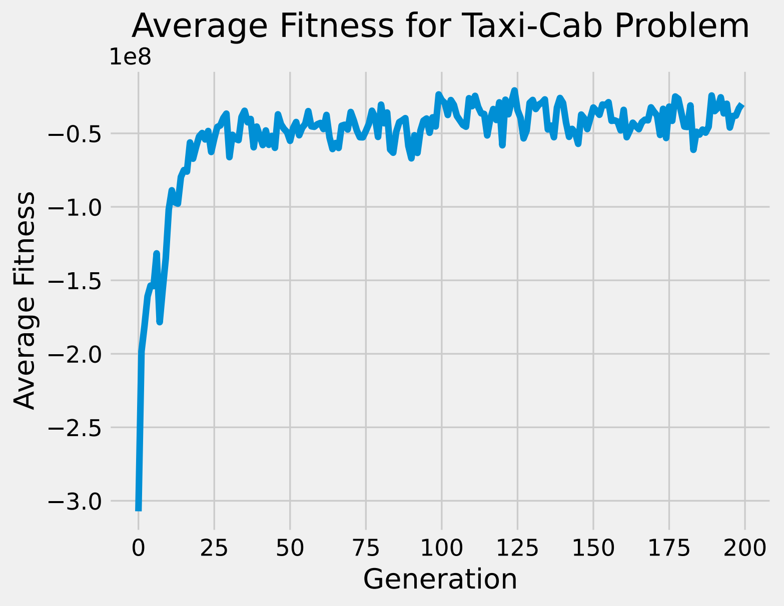


Figure 1.4: The average fitness of creatures during a GA run of the taxicab problem.

Example - Taxicab numbers

As an example, let us consider using GA to find taxi-cab numbers. That is, we want to find positive integer solutions to

$$a^3 + b^3 = c^3 + d^3, \tag{1.67}$$

where $a \neq c, d$.

If we limit ourselves to $1 \leq a, b, c, d \leq 2^5$, then each of the variables can be bijectively mapped into binary in the natural way. Our phenotype is therefore the set $\{a, b, c, d\}$, and our genotype is the concatenation of their binary representations. We want the fitness to be maximised for a solution to the above equation, and so choose the following fitness function:

$$f(a, b, c, d) = -\beta_1 |a^3 + b^3 - c^3 - d^3| - \beta_2(\delta_{a,c} + \delta_{a,d}), \tag{1.68}$$

where β_1 and β_2 are parameters to be tuned.

After some experimentation, we choose $N = 200$, $\alpha = 5$, $p = 1/25$, $\beta_1 = 1$ and $\beta_2 = 500$. We also allow elitism and run for 200 generations. After a single instance of GA, we find the average fitness increases with generation, as can be seen in Fig. 1.4.

In addition, we also find only one state (up to permutations) with fitness zero:

$$a = 1 \quad b = 12 \quad c = 9 \quad d = 3. \tag{1.69}$$

It turns out, up to permutations, this is the only taxi-cab number in the range we specified. This is distinct to the applications we will consider the reminder of the thesis, where there will be many creatures with maximal fitness but GA will only find a few for each genetic run. GA is fundamentally a heuristic method, and for the applications in this thesis we rerun GA multiple times, slowly building up a database of creatures with maximal fitness.

Throughout this thesis, we use two realisations of GA. One written in `C`, and the other in `Mathematica`. The code of both can be found on GitHub [124, 125].

1.5.2 Reinforcement Learning (RL)

Machine Learning (ML) and Neural Networks (NNs)

Before talking about RL, we first discuss Machine Learning (ML) and Neural Networks (NNs) more broadly. For a comprehensive introduction to the subject, see [123].

Machine learning is a general class of algorithms involving functions with learnable parameters $\vec{\theta}$ – usually denoted as $f_{\vec{\theta}}(\mathbf{x})$. These parameters can be updated such that the function satisfies desired properties. This is usually encoded by defining some loss function $\mathcal{L}[f_{\vec{\theta}}]$ which is to be minimised with respect to θ. This minimisation is achieved by first initialising $\vec{\theta}$ to some random values, then incrementally updating them by gradient descent

$$\delta\theta^i = -\alpha \frac{\partial \mathcal{L}}{\partial \theta^i}, \tag{1.70}$$

where the variable $\alpha < 1$ is called the learning rate. Typically we use batched data to ensure a measure of stochasticity, and use more complicated updating algorithms such as the ADAM optimiser [126].

As an example, suppose we are given a data set of in-data x_n and out-data y_n, with $n = 1, ..., N$, and we want to find a function that satisfies $f(x_n) \approx y_n \forall n$,

Figure 1.5: An example NN from $\mathbb{R}^n$ to $\mathbb{R}^m$ with two hidden layers and Tanh activation functions. Each box acts on the incoming vector and either applies an affine transformation, where the parameters are trainable, or applies an element-wise Tanh to each entry of the vector.

and also interpolates between the given data-points. We could consider the following loss function:

$$\mathcal{L}[f_{\vec{\theta}}] = \sum_{n=1}^{N} \left| f_{\vec{\theta}}(x_n) - y_n \right| . \tag{1.71}$$

Machine learning with this kind of problem is called supervised learning. In practical applications, some proportion of the data is not used for training and is instead reserved for validation and testing.

Alternatively, one could consider solving a differential equation $D_x \cdot f_{\vec{\theta}}(x) = 0$, with boundary conditions $BC_n[f_{\vec{\theta}}(x_n)] = 0$, for $n = 1, ..., N$. We could then consider the following loss function

$$\mathcal{L}[\pi_{\vec{\theta}}] = \alpha_1 \left(\int dx \left| D_x \cdot f_{\vec{\theta}}(x) \right| \right) + \alpha_2 \sum_{n=1}^{N} \left| BC_i[f_{\vec{\theta}}(x_n)] \right| , \tag{1.72}$$

where $\alpha_1, \alpha_2 > 0$ are some constants to be tuned. Such loss functions have been used to solve differential equations with machine learning [77–85, 127].

Usually, one takes $f_{\vec{\theta}}$ to be a feed-forward NN. These are a particularly convenient class of functions, given by the composition of affine transformations and element-wise non-linear activation functions in an alternating fashion. There are many choices of activation functions, but they are typically chosen to be functions with simple derivatives. An example NN is given in Fig. 1.5.

The trainable parameters of these neural networks are the matrices and vectors forming the affine transformation, and these are called the weights and biases in the machine learning literature.

One can prove that in the infinite width limit, even a single layer neural network can approximate any bounded continuous function, from a compact subspace of $\mathbb{R}^n$ to a compact subspace of $\mathbb{R}^m$.

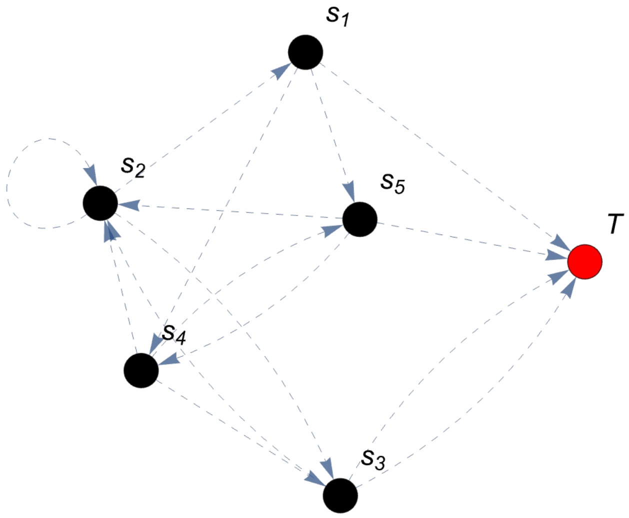


Figure 1.6: An example of an MDP with six states, one of which is terminal (indicated in red). There are three possible actions from each state, taking the agent to another state.

Reinforcement Learning

The main components of a Reinforcement Learning system are the *environment*, the *agents* and the neural network(s) [123, 128]. The latter are set up to learn certain properties of the environment, based on data delivered as the agent explores the environment. The mathematical underpinning of RL is provided by a *Markov decision process* (MDP), defined as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R})$. An example MDP can be seen in Fig. 1.6. Here $\mathcal{S}$ is a set which contains the environment's states, $\mathcal{A}$ is a set of maps $\alpha : \mathcal{S} \to \mathcal{S}$ which represent the actions, $\mathcal{P}$ provides a probability $\mathbb{P}(S = s'|S = s, A = \alpha)$ for a transition from state s to state s' via the action α, $\gamma \in [0,1]$ is called the *discount factor* and $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the *reward function*. Among the states in $\mathcal{S}$ a subset of so-called *terminal states* is singled out, which may, for example, consist of states with certain desirable properties. Within this set-up we can consider a sequence

$$s_0 \xrightarrow{\alpha_0, r_0} s_1 \xrightarrow{\alpha_1, r_1} s_2 \xrightarrow{\alpha_2, r_2} s_3 \cdots$$

of states s_t and actions α_t, producing rewards r_t, where $t = 0, 1, 2, \cdots$, which is referred to as an *episode*. In principle, an episode can have infinite length, although in practice a finite maximal episode length N_{ep} is imposed. If an episode arrives at a terminal state before it reaches its maximal number of steps it is stopped. The return G_t of a state s_t in such an episode is defined as

$$G_t = \sum_{k\geq 0} \gamma^k r_{t+k} \; . \tag{1.73}$$

The discount factor γ can be dialled to small values in order to favour short-term rewards dominating the return, or to values close to one so that longer-term rewards affect the return as well. The choice of action in a MDP is guided by a *policy* π, which provides probabilities $\pi(\alpha|s) = \mathbb{P}(A_t = \alpha|S_t = s)$ for applying a certain action α to a state s. Relative to such a policy, two important value functions, namely the *state value function* V_π and the *state-action value function* Q_π, can be defined as expectation values of the return.

$$V_\pi(s) = \mathbb{E}(G_t|S_t = s) \; , \qquad Q_\pi(s,\alpha) = \mathbb{E}(G_t|S_t = s, A_t = \alpha) \; . \tag{1.74}$$

The purpose of an RL system is to maximise a value function (state or state-action) over the set of possible policies. In practice, this can be realised in a number of ways which differ by which of the functions π, V_π and Q_π are represented by neural networks and how precisely these neural networks are trained via exploration of the environment. Common to all algorithms is an iterative approach, where a batch of data, in the form of triplets (s_t, a_t, G_t), is collected from episodes which are guided by the neural network(s) in their present state. This data is then used to update the neural network(s), followed by a further round of exploration and so on.

For our purposes, we will consider what is probably the simplest approach, a basic policy-based algorithm referred to as REINFORCE. This set-up contains a single neural network π_θ with weights θ which represents the policy π. Its inputs are states and the outputs are probabilities for actions. Exploration of the environment is guided by the policy, meaning the steps in an episode are selected based on π_θ, so

$$s_0 \xrightarrow{\pi_\theta} s_1 \xrightarrow{\pi_\theta} s_2 \xrightarrow{\pi_\theta} s_3 \cdots \; . \tag{1.75}$$

Data is collected by performing such episodes successively, so we can say that the system contains only one agent. According to the *policy-gradient theorem*, the neural network π_θ should be trained on the loss function

$$L(\theta) = Q_\pi(s,a)\ln(\pi_\theta(s,a)) \; , \tag{1.76}$$

where $Q_\pi(s,a)$ can, in practice, be replaced by the return G of the state s. Schematically, the algorithm then proceeds as follows.

(1) Initialise the policy network π_θ.

(2) Collect a batch of data triplets (s_t, a_t, G_t) from as many episodes (1.75) as required. New episodes start at random states s_0.

(3) Use this batch to update the weights θ of the policy network π_θ, based on the loss (1.76).

(4) Repeat from (2) until the loss is sufficiently small so that the policy has converged.

Example - Line Bundle Cohmomology

As an example, we apply RL to a simple environment which consists of the set of all line bundles (with entries in a certain range) on a given CY manifold. The goal is to identify line bundle with a given target index – this is equivalent to solving a cubic Diophantine equation. This environment is of course much simpler and smaller than other environments we tackle in the later chapters. However, it is qualitatively similar in that it is based on vector bundles whose topological properties we are attempting to engineer. It provides us with a first confirmation that RL is indeed capable of carrying out topological engineering. Moreover, the environment is small enough that it can easily be scanned systematically so that we have an independent check on the results obtained from RL. Given that we focus on a simple two-dimensional environment which facilitates graphical representation it is also a good opportunity to develop intuition and illustrate the workings of an RL system.

Our setting is a fixed CY three-fold X with Picard number $h = h^{1,1}(X)$ together with its line bundles $L = \mathcal{O}_X(\mathbf{k})$, labelled by h-dimensional integer vectors $\mathbf{k} = (k^1, \ldots, k^h)$. Let us be more precise about how this mathematical setting is mapped to the ingredients of a MDP. The environment is given by h-dimensional integer vectors

$$\mathcal{S} = \{\mathbf{k} \in \mathbb{Z}^h \, : \, |k_i| \leq k_{\rm max}\} \, , \tag{1.77}$$

which represent the line bundles $\mathcal{O}_X(\mathbf{k})$. To obtain a finite environment, we have imposed an upper bound $k_{\rm max}$ on the absolute values of the entries. The allowed actions involve increasing or decreasing one entry in $\mathbf{k}$ by one, so

$$\mathcal{A} = \{\mathbf{k} \mapsto \mathbf{k} \pm \mathbf{e}_i\} \, , \tag{1.78}$$

where $\mathbf{e}_i$ is the $i^{\rm th}$ standard unit vector in h dimensions. The definition of the reward is based on an *intrinsic state value*

$$v(\mathbf{k}) = -\frac{10\,|{\rm ind}(\mathcal{O}_X(\mathbf{k})) - \tau|}{h k_{\rm max}^3} \tag{1.79}$$

which (modulo normalisation factors) measures the negative difference of the index from the target index τ. The formula for the index of a line bundle was given in Sec. 1.4.2. Based on this intrinsic value, the reward is defined as

$$r_{s \mapsto s'} = \left\{ \begin{array}{ccc} (v(s') - v(s))^p & \text{if} & v(s') - v(s) > 0 \\ r_{\rm offset} & \text{if} & v(s') - v(s) \leq 0 \end{array} \right\} + r_{\rm step} + r_{\rm boundary} + r_{\rm terminal} \tag{1.80}$$

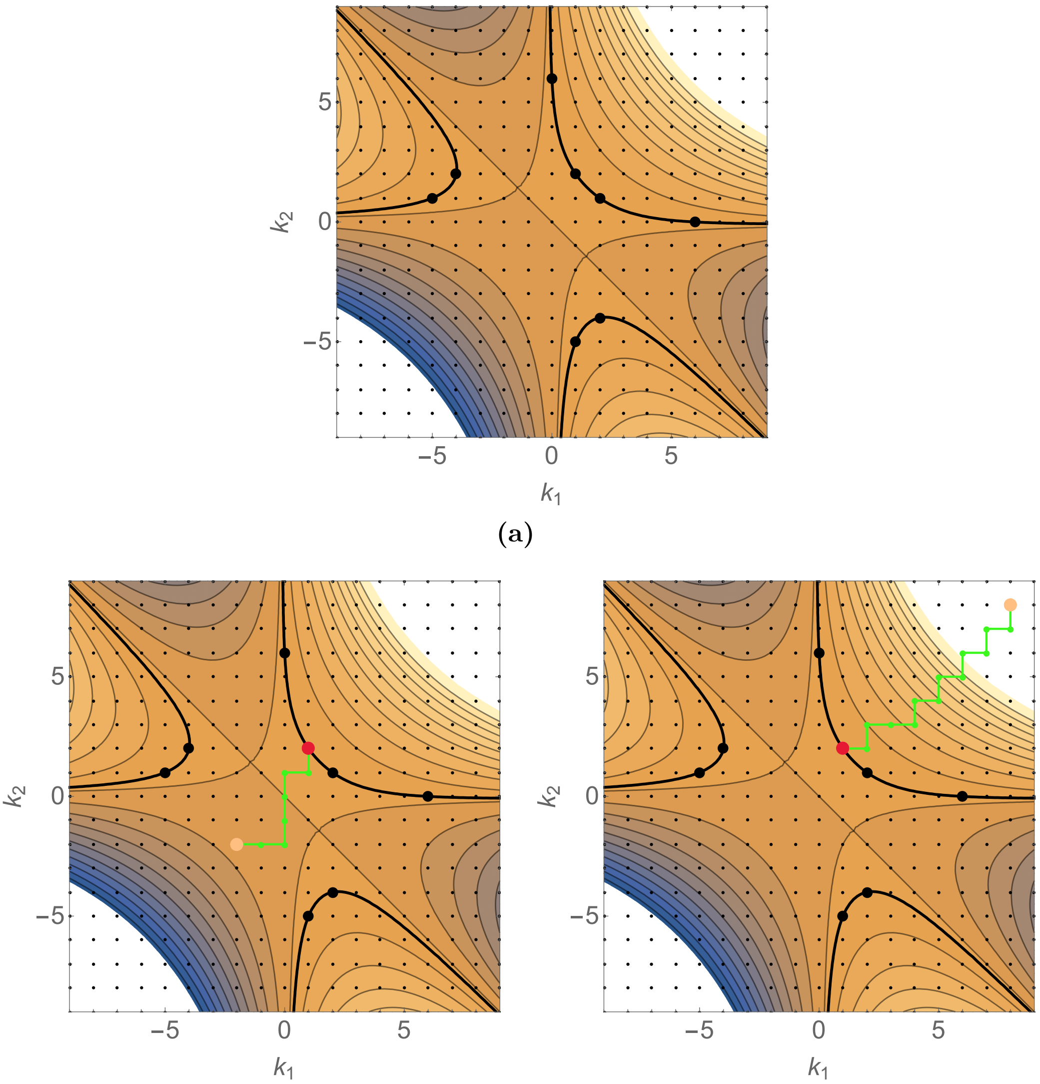
where $p \in \mathbb{R}$ is a suitably chosen power and $r_{\rm offset} < 0$ is a penalty for decreasing the intrinsic state value. To favour finding terminal states by short episodes a penalty $r_{\rm step} < 0$ is added for each step and actions which lead beyond the boundary set by $k_{\rm max}$ attract a penalty $r_{\rm boundary} < 0$. Finally, actions which lead to a terminal state are rewarded with a bonus $r_{\rm terminal} > 0$.

The above set-up can be applied to line bundles on any CY threefold but, for simplicity, we focus on a generic bicubic CY whose configuration matrix is given by

$$X = \begin{array}{c} \mathbb{P}^2 \\ \mathbb{P}^2 \end{array} \left[\begin{array}{c} 3 \\ 3 \end{array} \right] . \tag{1.81}$$

Since the Picard number equals $h = 2$ for this manifold, line bundles $L = \mathcal{O}_X(\mathbf{k})$ are specified by two dimensional integer vectors $\mathbf{k} = (k_1, k_2)$ and the explicit formula for the index is

$$\text{ind}(X, \mathcal{O}_X(k_1, k_2)) = \frac{3}{2}(k_1 + k_2)(2 + k_1 k_2) \, . \tag{1.82}$$



(a)

(b)

(c)

Figure 1.7: (a) Result of a systematic scan for line bundles on the bicubic three-fold, with the large back points representing line bundles with index equal to $\tau = 18$ (b) (c) Two sample episodes (green lines) obtained by following an RL-trained policy, starting at the yellow points and ending up in terminal states (red points).

We restrict the range of line bundle integers by setting $k_{\text{max}} = 9$ and we are looking for line bundles with target index $\tau = 18$. The parameters which enter the definition of the reward function (1.80) are chosen as

$$p = 1\,, \quad r_{\text{offset}} = -1\,, \quad r_{\text{step}} = 0\,, \quad r_{\text{boundary}} = -1\,, \quad r_{\text{terminal}} = 2\,.$$

The environment contains $19^2 = 361$ states so it is small enough to be scanned systematically. The result of such a scan is shown in Figure 1.7(a), where the eight black dots indicate the terminal states, that is, the line bundles with the required target index $\tau = 18$.

To realise the REINFORCE algorithm we need to supply a neural network π_θ which represents the policy. Its input are the two-dimensional vectors $\mathbf{k} = (k_1, k_2)$ and its output is a four-dimensional vector which provides the probabilities for the four possible actions. We use a standard feed-forward neural network with the eight-layer architecture shown in Fig 1.8, with input and output dimensions $d_0 = 2$ and $d_1 = 4$ and width $d = 16$.

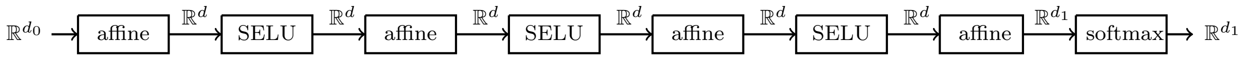


Figure 1.8: Structure of neural network used as a policy and value network.

Here “affine" refers to an affine layer performing the transformation $\mathbf{x} \to W\mathbf{x} + \mathbf{b}$ with weight W and bias $\mathbf{b}$, The layers “SELU" refer to the standard scaled exponential linear unit activation function

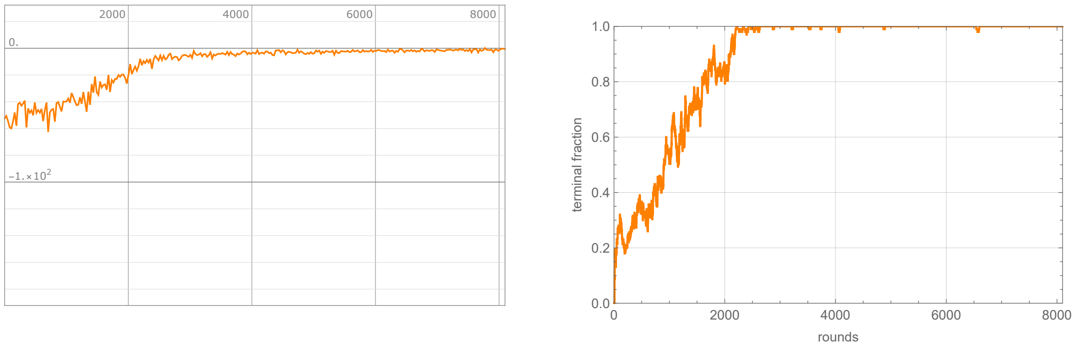
$$\rho(x) = \begin{cases} 1.0507x \ , & x \geq 0 \\ 1.7581(e^x - 1) \ , & x < 0 \end{cases} ,$$

while “softmax" is a softmax (normalised exponential) layer defined by

$$\sigma : \mathbb{R}^n \to [0,1]^n \ , \qquad \sigma(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}} \ .$$

It ensures that the output components are positive and sum up to one so they can be interpreted as probabilities.

To train this neural network, the agent is coupled to the line bundle environment which is explored in episodes of maximal length $t_{\max} = 16$ and with a discount factor $\gamma = 0.98$. Training data is supplied in batches of size 32 (two full episodes) and for stochastic gradient descent we use the ADAM optimiser with learning rate 1/5000. Training is accomplished on a single CPU and only takes a few minutes. The training measurements are shown in Fig. 1.9. Since the environment is quite small every state is sampled multiple times during training (this will be different for the much larger environments discussed in the later sections), so it is not surprising that all



(a) Loss vs batch number.

(b) Fraction of terminal episodes vs episode number.

Figure 1.9: Training metrics for line bundle environment on bicubic with a target index of 18.

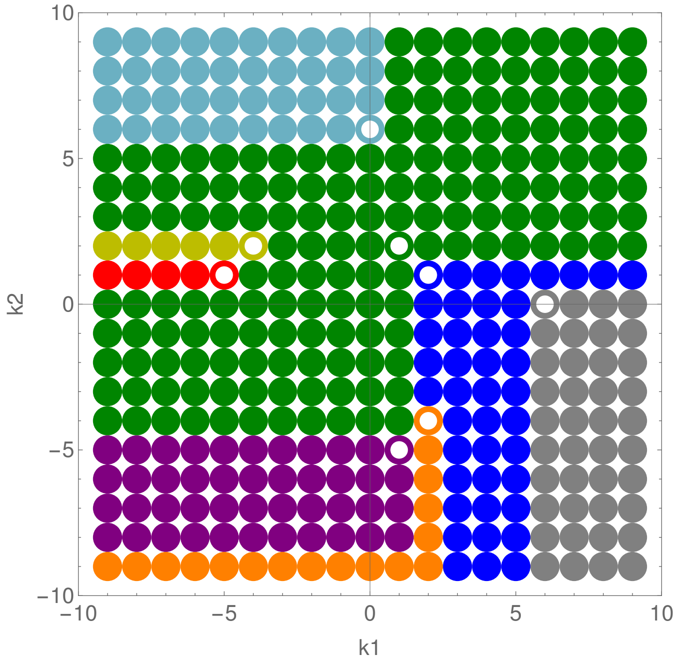


Figure 1.10: The basins of attraction for the terminal states of the line bundle environment, as encoded in the trained policy network. Each colour represents a basin of attraction with the corresponding terminal state indicated by the white circle in one corner of the region.

eight terminal states are found. After about 200 training rounds, when the loss goes to zero, the fraction of terminal episodes approaches 1 as is evident from Fig. 1.9 (b). At the same time, the average episode length decreases to about 6 steps. This means, the trained network guides episodes from any starting point to a terminal state on an efficient path with average length 6. Two examples for such paths are shown in Fig. 1.7 (b), (c). Similar results are obtained using the actor critic method.

Given the small environment size we can approach this more systematically and carry out an episode guided by the training network starting from every state. In this way, we find the basin of attraction for each terminal state and the result is shown in Fig. 1.10.

It shows that the basins of attraction can be of different sizes and can have complicated shapes. It is also interesting to note that the network does not always guide to the nearest terminal state. We only need to carry out one episode guided by the trained network starting in each domain to find all the terminal states. This compares favourably with a systematic scan of the entire environment.

In summary, training for the line bundle environment is quite successful and leads to a policy network which efficiently guides to the states with the desired target index for all starting points. This provides a first indication that RL is a suitable method for engineering topological quantities.

2

Automated Model Building

Contents

Before discussing applications of these search algorithms to string theory, we offer a toy example of "automated model building" from flavour physics [20, 129]. Namely we are going to explore the possibility of generating quark masses by introducing a number of scalar fields charged under some auxiliary $U(1)$ symmetries. The standard model fields are also charged under these symmetries, such that Yukawa couplings are generated when the new scalars acquire a vev. We will only use RL in this section, but we note that GA has also been applied in other physics contexts outside of string theory. Namely, for exploring models of cosmic inflation [118].

This is the only section in the reminder of this thesis where we will be looking at the standard model, as introduced in Sec. 1.2, rather than its supersymmetric extensions.

We begin by introducing the Froggatt-Nielsen models and discussing how to encode the relevant problem as an MDP in Sec. 2.1 and Sec. 2.2 respectively. We then present the results and conclude this chapter.

2.1 Froggatt-Nielsen Models

Before we discuss Froggatt-Nielsen models, we quickly review fermion masses in the standard model of particle physics, in order to set up notation and present the experimental data.

The standard model contains Yukawa interactions, which are responsible for generating the masses and mixing of quarks and leptons. In this chapter, we focus on the quark sector for simplicity, although we expect that our work can be generalised to include the lepton sector. The quark Yukawa couplings in the standard model take the form

$$\mathcal{L}_{\rm Yuk} = Y^u_{ij}\overline{Q}^i H^c u^j + Y^d_{ij}\overline{Q}^i H d^j + \text{h.c.}\ , \tag{2.1}$$

where Q_i are the left-handed quarks, u_i, d_i are the right-handed up and down type quarks and H is the Higgs doublet. We use indices $i, j, ... = 1, 2, 3$ to label the three families. Within the standard model, the Yukawa matrices Y^u and Y^d are not subject to any theoretical constraints - their (generally complex) values are inserted "by hand" in order to fit the experimental results for masses and mixing.

Once the charge-neutral component H^0 in the Higgs doublet develops a vev, $v = \langle H^0 \rangle$, the above Yukawa terms lead to Dirac mass terms with associated mass matrices

$$M_u = v\, Y^u\ , \qquad M_d = v\, Y^d\ . \tag{2.2}$$

These matrices need to be diagonalised,

$$M_u = U_u \hat{M}_u V_u^\dagger\ , \quad M_d = U_d \hat{M}_d V_d^\dagger\ , \tag{2.3}$$

where

$$\hat{M}_u = \text{diag}(m_u, m_c, m_t)\ , \quad \hat{M}_d = \text{diag}(m_d, m_s, m_b)\ , \tag{2.4}$$

by unitary matrices U_u, V_u, U_d, V_d in order to obtain the masses (m_u, m_c, m_t) for the up-type quarks and the masses (m_d, m_s, m_b) for the down-type quarks. The other observable quantity obtained from the quark Yukawa couplings is the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM}, defined by

$$V_{CKM} = U_u^\dagger U_d = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}. \tag{2.5}$$

The CKM matrix is unitary and can, hence, be written in terms of three angles $\theta_{12}, \theta_{13}, \theta_{23}$ and a phase δ, as in the above equation, where the abbreviations $s_{ij} = \sin(\theta_{ij})$ and $c_{ij} = \cos(\theta_{ij})$ have been used. The experimentally measured values for these quantities are given in Table 2.1. Inserting the mixing angles and the phase from Table 2.1 into the parametrisation (2.5) gives the numerical CKM matrix

m_u	m_d	m_c	m_s
$0.00216^{+0.00049}_{-0.00026}$	$0.00467^{+0.00048}_{-0.00017}$	1.27 ± 0.02	$0.093^{+0.011}_{-0.005}$
m_t	m_b	v	s_{12}
172.4 ± 0.07	$4.18^{+0.03}_{-0.02}$	~ 174	0.22650 ± 0.00048
s_{13}	s_{23}	δ	
$0.00361^{+0.00011}_{-0.00009}$	$0.04053^{+0.00083}_{-0.00061}$	$1.196^{+0.045}_{-0.043}$	

Table 2.1: Experimentally measured masses in GeV and mixing angles of quarks from [26].

$$|V_{CKM}| \approx \begin{pmatrix} 0.9740 & 0.2265 & 0.0036 \\ 0.2264 & 0.9732 & 0.0405 \\ 0.0085 & 0.0398 & 0.9992 \end{pmatrix}. \tag{2.6}$$

In the context of the standard model, the Yukawa matrices Y^u and Y^d in Eq. (2.1) have to be chosen to fit these experimental values for masses and mixing but this still leaves considerable freedom. Only 10 real constrains are imposed on the 36 real parameters which determine Y^u and Y^d.

Froggatt-Nielsen (FN) models provide a framework for adding more structure to the Yukawa sector of the standard model, in an attempt to remove some of this ambiguity and provide a theoretical explanation for the observed masses and mixing. Two main ingredients are added to the picture: a number of global $U(1)$ symmetries $U_a(1)$, where $a = 1, \ldots, r$, and a number of complex scalar fields ϕ_α, where $\alpha = 1, \ldots, \nu$, which are singlets under the standard model gauge group. The standard model fields as well as the scalar singlets are assigned $U_a(1)$ charges which we denote by $q_a(Q^i)$, $q_a(u^i)$, $q_a(d^i)$, $q_a(H)$ and $q_a(\phi_\alpha)$. In fact, to simplify matters, we assume that we have the same number of $U(1)$ symmetries and singlet fields, $\nu = r$, and that the a^{th} singlet ϕ^a is only charged under $U_a(1)$. The resulting singlet charges are then denoted by $q_a(\phi)$.

Given this set-up, the standard model Yukawa couplings (2.1) are no longer in general consistent with the $U_a(1)$ symmetries and should be replaced by

$$\mathcal{L}_{\text{Yuk}} = \sum_{i,j} \left(a_{ij} \phi_1^{n_{1,ij}} \cdots \phi_r^{n_{r,ij}} \overline{Q}^i H^c u^j \; + \; b_{ij} \phi_1^{m_{1,ij}} \cdots \phi_r^{m_{r,ij}} \overline{Q}^i H d^j \right) + \text{h.c.} \tag{2.7}$$

where $n_{a,ij}$ and $m_{a,ij}$ are non-negative integers. For a term (ij) in the up-quark sector to be invariant under $U_a(1)$ we require the conditions

$$n_{a,ij} = -\frac{q_a(\overline{Q}^i H^c u^j)}{q_a(\phi)} \; . \tag{2.8}$$

Hence, the term (ij) in the u-quark sector is allowed if the $n_{a,ij}$ given by Eq. (2.8) are non-negative integers for all $a = 1, \ldots, r$. In this case, the coefficient a_{ij} is of order one, otherwise it is set to zero. An analogous rule applies to the terms for the down-type quarks. Once the scalars ϕ_a develop vevs, $v_a = \langle \phi_a \rangle$, Yukawa couplings

$$Y^u_{ij} = a_{ij} v_1^{n_{1,ij}} \cdots v_r^{n_{r,ij}} \; , \qquad Y^d_{ij} = b_{ij} v_1^{m_{1,ij}} \cdots v_r^{m_{r,ij}} \; . \tag{2.9}$$

are generated [1]. The main model building idea in this setting is that moderately small singlet vevs v_a can generate the required large hierarchies in masses, in a way that is controlled by the integers $n_{a,ij}$ and $m_{a,ij}$ and, hence, ultimately, by the choices of $U_a(1)$ charges.

At this stage the environment of FN models consists of $U_a(1)$ charges for all fields, the singlet vevs v_a and the coefficients a_{ij}, b_{ij}. In principle, the singlet

---

[1] If these Yukawa couplings are generated at a high energy scale they have to be renormalised down to the electro-weak scale, in order to facilitate comparison with the experimental values. Since this typically leads to order one coefficients which have already been included via a_{ij}, b_{ij} we will not consider this explicitly.

vevs are meant to be fixed by a scalar potential but implementing this in detail adds another layer of model building. Instead, for a given choice of charges and coefficients a_{ij}, b_{ij}, we will fix the vevs v_a such that the model provides an optimal fit to the experimental masses and mixing. The non-zero coefficients a_{ij}, b_{ij} might be considered as part of the environment definition but, to keep things simple, we will fix those to specific numerical values of order one. While, in general, a_{ij} and b_{ij} can be complex, we simplify this scenario by only allowing them to take real values. Consequently, we will not attempt to fit the CP violating phase δ in the CKM matrix. As a further simplification, we require that the top Yukawa term $\bar{Q}^3 H^c u^3$ is present without any singlet insertions, a condition which seems reasonable given the size of the top Yukawa coupling. This requirement can be used to fixed the $U_a(1)$ charges of the Higgs multiplet as

$$q_a(H) = q_a(u^3) - q_a(Q^3) \ . \tag{2.10}$$

Altogether, this means a FN model within our set-up is specified by the charges choices

$$\left(\mathcal{Q}_a{}^I\right) = \left(q_a(Q^i), q_a(u^i), q_a(d^i), q_a(\phi)\right) \ , \tag{2.11}$$

which we have assembled into the $r \times 10$ integer charge matrix $\mathcal{Q}$. In practice, the charges in $\mathcal{Q}$ will be restricted to a certain range

$$q_{\rm min} \leq \mathcal{Q}_a{}^i \leq q_{\rm max} \ , \tag{2.12}$$

with $q_{\rm min}$ and $q_{\rm max}$ to be specified later. While this leads to a finite space of charge matrices and associated FN models, numbers can be considerable. For example, for $-q_{\rm min} = q_{\rm max} = 9$ we have $\sim 10^{13}$ models in the case of a single $U(1)$ symmetry and $\sim 10^{26}$ models for the case of two $U(1)$ symmetries.

The environment (2.11) of FN models has a number of permutation degeneracies, since the assignment of charges to families and the order of $U_a(1)$ symmetries does not carry physical meaning, although part of this symmetry is broken by designating Y^u_{33} the top Yukawa coupling. This means there is a permutation degeneracy isomorphic to

$$S_2 \times S_2 \times S_3 \times S_r \tag{2.13}$$

in the environment (2.11). For the purpose of RL we will not attempt to remove this redundancy, as this would complicate the constraints on the charges in $\mathcal{Q}$.

From the viewpoint of particle physics model building the task is now to investigate the model landscape defined by Eq. (2.11) and extract the phenomenologically promising cases. Considerable effort has been invested into this, since the original proposal of Froggatt and Nielsen [129–131]. It is precisely this task we wish to carry out using reinforcement learning.

2.2 Model building with reinforcement learning

We now explain how we propose to map the problem of FN model building onto the structure of reinforcement learning. We begin by describing the set-up of the RL environment.

2.2.1 The environment

Firstly, we must identify how the various ingredients of a MDP are realised in our context. We take the set $\mathcal{S}$ of states to consists of all FN models for a fixed number, r, of $U(1)$ symmetries and the same number of singlet fields. These models are represented by the $r \times 10$ integer charge matrices $\mathcal{Q}$ in Eq. (2.11), with entries restricted as in Eq. (2.12). The set $\mathcal{A}$ of actions α consists of the basic operations

$$\mathcal{Q}_a{}^I \xrightarrow{\alpha} \mathcal{Q}_a{}^I \pm 1\ , \tag{2.14}$$

that is, increasing or decreasing a single charge $\mathcal{Q}_a{}^I$ by one while keeping all other charges unchanged. These are deterministic actions so we do not need to introduce transition probabilities $\mathcal{P}$. The number of different actions is $2 \times r \times 10 = 20r$. For the discount factor γ we choose the value $\gamma = 0.98$.

Defining the reward function $\mathcal{R}$ requires a bit more effort. We start by defining the intrinsic value for a state $\mathcal{Q}$ as

$$\mathcal{V}(\mathcal{Q}) = -\min_{|v_a|\in I} \sum_{\mu} \left| \log_{10}\left(\frac{|\mu_{\mathcal{Q},v_a}|}{|\mu_{\text{exp}}|} \right) \right| . \tag{2.15}$$

Here, μ runs over the six quark masses as well as the entries of the CKM matrix, $\mu_{\mathcal{Q},v_a}$ is the value for one of these quantities predicted by the model with charge matrix $\mathcal{Q}$ and scalar fields vevs v_a, computed from Eqs. (2.9), (2.2), (2.3), (2.5) (using fixed random values of the order-one coefficients a_{ij}, b_{ij}), and μ_{exp} is its experimental value as given in Table (2.1) and Eq. (2.6). The minimisation is carried out over the scalar field vevs v_a, in a certain range $I = [v_{\text{min}}, v_{\text{max}}]$, with typical values $v_{\text{min}} = 0.01$ and $v_{\text{max}} = 0.3$. From this definition, the intrinsic value of a state $\mathcal{Q}$ is simply the (negative) total order of magnitude by which predicted masses and mixings deviate from the experimental ones, for optimal choices of the scalar field vevs.

A terminal state $\mathcal{Q}$ in our environment is one which is phenomenologically promising, that is, a state which gives rise to (roughly) the correct masses and mixings. More specifically, we call a state terminal if its intrinsic value $\mathcal{V}(\mathcal{Q})$ is larger than a certain threshold value $\mathcal{V}_0$ and if each individual deviation $-|\log_{10}(|\mu_{\mathcal{Q}}|/|\mu_{\text{exp}}|)|$ (computed for the scalar field vevs which minimise Eq. (2.15)) is larger than a threshold value $\mathcal{V}_1$. Since we have fixed our order-one parameters a_{ij}, b_{ij} these threshold values are chosen relatively generously, so as to not miss any promising models. For our computations, we have used $\mathcal{V}_0 = -10$ and $\mathcal{V}_1 = -1$.

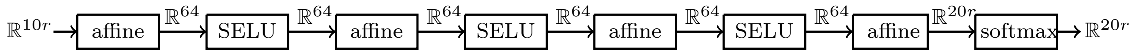
Based on this intrinsic value, the reward $\mathcal{R}(\mathcal{Q}, \alpha)$ for an action $\mathcal{Q} \overset{\alpha}{\to} \mathcal{Q}'$ of the form (2.14), connecting two states $\mathcal{Q}$ and $\mathcal{Q}'$, is defined by

$$\mathcal{R}(\mathcal{Q}, \alpha) = \begin{cases} \mathcal{V}(\mathcal{Q}') - \mathcal{V}(\mathcal{Q}) & \text{if} \quad \mathcal{V}(\mathcal{Q}') - \mathcal{V}(\mathcal{Q}) > 0 \\ \mathcal{R}_{\text{offset}} & \text{if} \quad \mathcal{V}(\mathcal{Q}') - \mathcal{V}(\mathcal{Q}) \leq 0 \end{cases} . \tag{2.16}$$

Here, $\mathcal{R}_{\text{offset}}$ is a fixed (negative) value which penalises a decrease of the intrinsic value, typically chosen as $\mathcal{R}_{\text{offset}} = -10$. In addition, if the new state $\mathcal{Q}'$ is terminal a terminal bonus $\mathcal{R}_{\text{term}}$, typically chosen as $\mathcal{R}_{\text{term}} = 100$, is added to the reward (2.16).

2.2.2 Neural network

To represent the policy π, we use a fully connected network f_θ with the following structure.



Here, “affine" refers to an affine layer performing the transformation $\mathbf{x} \to W\mathbf{x} + \mathbf{b}$ with weight W and bias $\mathbf{b}$, “SELU" is the standard scaled exponential linear unit activation function and “softmax" is a softmax layer which ensures that the output can be interpreted as a vector of probabilities which sum to one. The input of this network is the charge matrix $\mathcal{Q}$, in line with the input dimension of $10r$ while the output is a probability vector whose dimension, $20r$, equals the number of different actions (2.14).

Training data is provided in batches which consist of triplets $(\mathcal{Q}_t, \alpha_t, G_t)$, where the actions α_t are represented by a standard unit vector in $\mathbb{R}^{20r}$. The probability of an action can then be written as $\pi_\theta(\mathcal{Q}_t, \alpha_t) = \alpha_t \cdot f_\theta(\mathcal{Q}_t)$ and the loss (1.76) takes the form

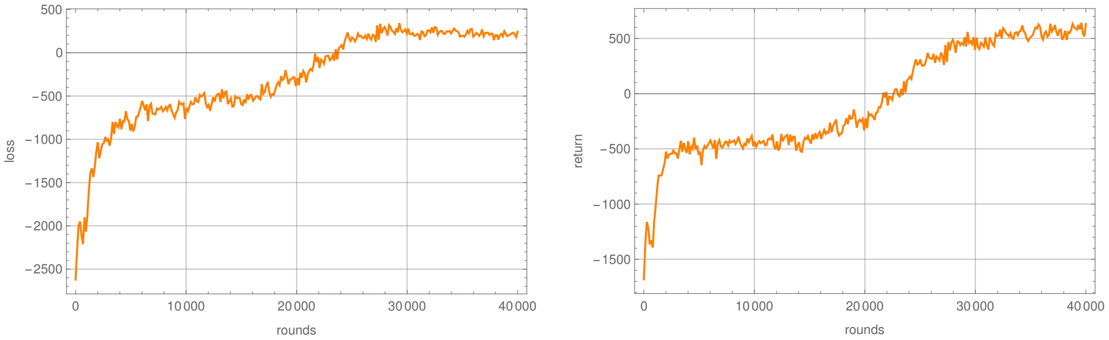
$$L(\theta) = G_t \ln(\alpha_t \cdot f_\theta(\mathcal{Q}_t) \ . \tag{2.17}$$

Based on this loss function, the above network is trained with the ADAM optimiser [126], using batch sizes of 32 and a typical learning rate of $\lambda = 1/4000$.

2.2.3 Agent

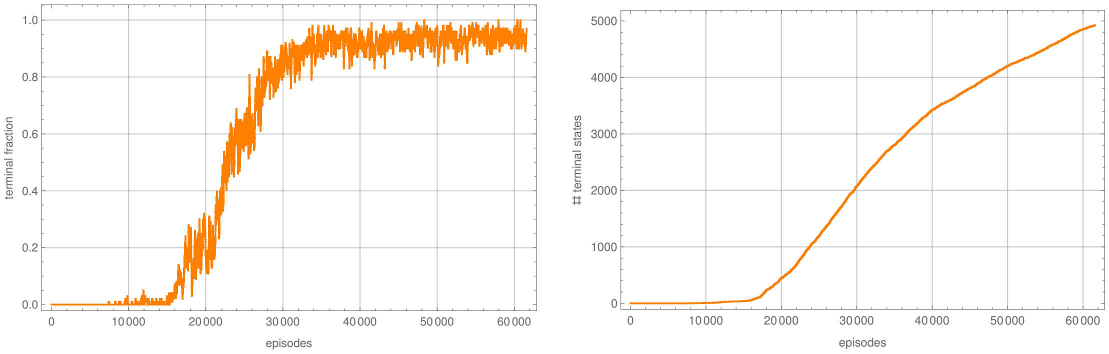
The FN environment will be explored by a single agent, following episodes (1.75) of maximal length $N_{\text{ep}} = 32$, and guided by the policy network π_θ. Each new episode is started from a random state, to improve exploration of the environment. Terminal states which are encountered during training are stored for later analysis.

The FN environment and the REINFORCE algorithm are realised as `Mathematica` packages, the latter based on the `Mathematica` suite of machine learning modules. For terminal states found during training or by applying the trained network we



(a) Loss vs batch number.

(b) Return vs batch number.



(c) Fraction of terminal episodes vs episode number.

(d) Number of terminal states vs episode number.

Figure 2.1: Training measurements for the case of a single $U(1)$ and $-q_{\rm min} = q_{\rm max} = 9$.

perform a further Monte Carlo analysis in the space of order one coefficients a_{ij}, b_{ij} (which were held fixed during training) in order to optimise their intrinsic value $\mathcal{V}(\mathcal{Q})$.

2.3 Results

In this section, we present the results we have obtained by applying the REINFORCE algorithm to the FN environment, as described in the previous section. We focus on the two cases of one $U(1)$ symmetry with one singlet scalars and two $U(1)$ symmetries with two singlet scalars, starting with the former.

2.3.1 One $U(1)$ symmetry

The entries of the 1×10 charge matrix $\mathcal{Q}$ are restricted as in Eq. (2.12), with $-q_{\rm min} = q_{\rm max} = 9$, so the environment contains $19^{10} \sim 10^{13}$ states. Training of the network in Section 2.2.2 takes about an hour on a single CPU and the measurements made during training are shown in Fig. 2.1. After an initial phase of exploration, lasting for about 15000 rounds, the network learns rapidly and the fraction of episodes which end in terminal states (plot (c) in the Fig. 2.1) rises to $> 90\%$ within 10000 rounds or so. This pattern is quite characteristic and persists under variation of the various pieces of meta data, including the depth and width of the network, the constants which enter the definition (2.16) of the reward and the definition of a terminal state. The result is also stable under modest variations of the learning rate $\lambda = 1/4000$, although too large learning rates ($\lambda > 1/1000$) suppress exploration and lead to convergence to the "wrong" policy. That is, a policy that does not find terminal states efficiently, if at all. The residual positive loss in Fig. 2.1 (a) can be attributed to the fact that frequently more than one efficient path to a terminal state exists. In other words, there are several very similar optimal policies.

During training, 4924 terminal states are found, which are reduced to 4630 after the redundancies due to the permutations (2.13) are removed. Episodes guided by the trained network, starting at a random state and with maximal length 32, lead to terminal states in 93% of cases, and the average episode length is 16.4.

The intrinsic values of these 4630 models found during training are optimised by performing a Monte-Carlo search over the order one coefficients a_{ij} and b_{ij}. In this way, we find 89 models $\mathcal{Q}$ with an intrinsic value $\mathcal{V}(\mathcal{Q}) > -1$. From these, the model with the highest intrinsic value is given by [2]

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 6 & 4 & 3 & -2 & 2 & 4 & -3 & -1 & -1 & 1 & 1 \end{array}\right) . \tag{2.18}$$

For a scalar vev $v_1 \simeq 0.224$ and the order one coefficients

$$\begin{gathered} (a_{ij}) \simeq \begin{pmatrix} -1.975 & 1.284 & -1.219 \\ 1.875 & -1.802 & -0.639 \\ 0.592 & 1.772 & 0.982 \end{pmatrix} , \\ (b_{ij}) \simeq \begin{pmatrix} -1.349 & 1.042 & 1.200 \\ 1.632 & 0.830 & -1.758 \\ -1.259 & -1.085 & 1.949 \end{pmatrix} , \end{gathered} \tag{2.19}$$

---

[2] Even though the Higgs charge is not part of the environment we include it here for convenience.

this model has an intrinsic value $\mathcal{V}(\mathcal{Q}) \simeq -0.598$ and leads to the mass matrices

$$\begin{aligned} M_u &\simeq \begin{pmatrix} 0.000 & 0.126 & -2.380 \\ 0.009 & -3.517 & -24.904 \\ 0.013 & 15.456 & 170.815 \end{pmatrix}, \\ M_d &\simeq \begin{pmatrix} -0.001 & 0.023 & 0.026 \\ 0.036 & 0.363 & -0.768 \\ -0.123 & -2.119 & 3.806 \end{pmatrix}. \end{aligned} \tag{2.20}$$

After diagonalisation, the resulting quark masses and mixings are

$$\begin{aligned} (m_u, m_c, m_t) &\simeq (0.003, 1.292, 173.358)\,\text{GeV} \\ (m_d, m_s, m_b) &\simeq (0.005, 0.066, 4.439)\,\text{GeV} \end{aligned} \tag{2.21}$$

$$V_{CKM} = \begin{pmatrix} 0.969 & 0.247 & 0.003 \\ -0.247 & 0.968 & 0.050 \\ 0.009 & -0.049 & 0.999 \end{pmatrix}, \tag{2.22}$$

in reasonable agreement with the values in Table 2.1 and Eq. (2.6). Further examples of models with a high intrinsic value found during training are listed in App. A.1.

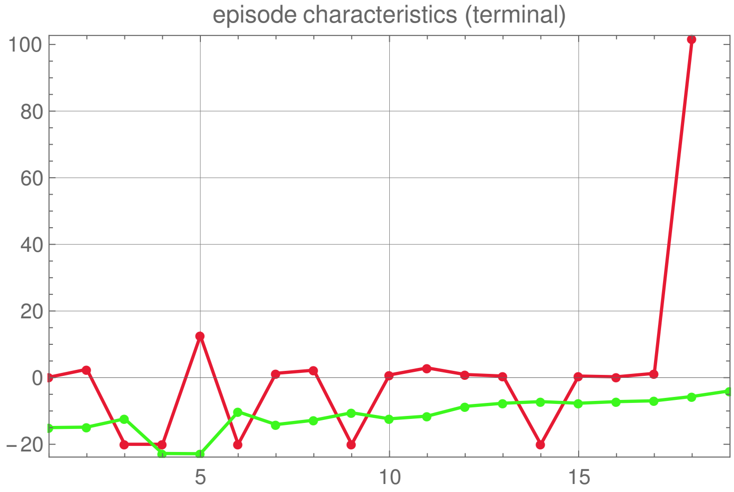
Of course, the trained network can be used to find new models. For example, consider starting with the initial state

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 0 & 2 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right). \tag{2.23}$$

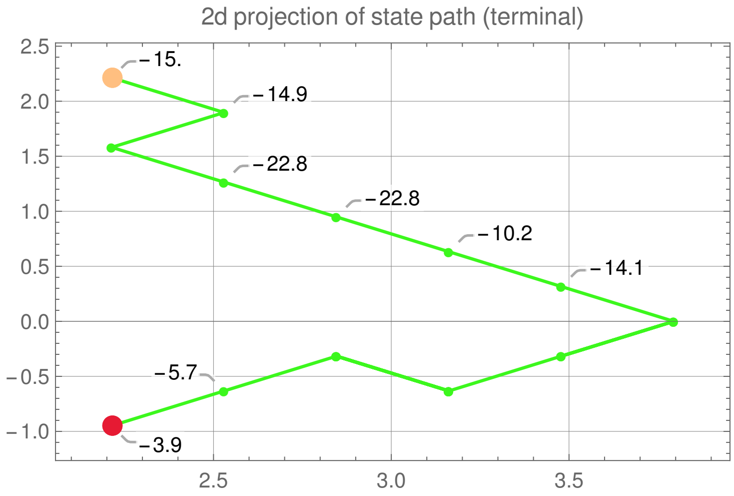
The optimal intrinsic value for this state, achieved for a singlet vev $v_1 \simeq 0.112$, is $\mathcal{V}(\mathcal{Q}) \simeq -15$, so this is definitely not a phenomenologically viable model. Using (2.23) as the initial state of an episode, guided by the trained network, it takes 18 steps to reach the terminal state

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 2 & 3 & 1 & 1 & 3 & 3 & -2 & -2 & -3 & 2 & 1 \end{array}\right), \tag{2.24}$$

with intrinsic value $\mathcal{V}(\mathcal{Q}) \simeq -3.94$ for a singlet vev $v_1 \simeq 0.056$. The intrinsic value and the reward along this episode, as well as a two-dimensional projection of the path mapped out during the episode is shown in Fig. 2.2.



(a) Intrinsic value (green) and reward (red) vs episode steps.



(b) Two-dimensional projection of the evolution of the charge matrices $\mathcal{Q}_t$ of the episode connecting the initial state (2.23) (yellow dot) with the final state (2.24) (red dot). This projection of the environment lattice is onto a random and misaligned plane. The axis labels are the distance from the origin of the lattice, along this plane in units of the lattice spacing. The labels indicate the intrinsic values of each state visited along the episode.

Figure 2.2: Characteristics of the episode connecting the states (2.23) and (2.24).

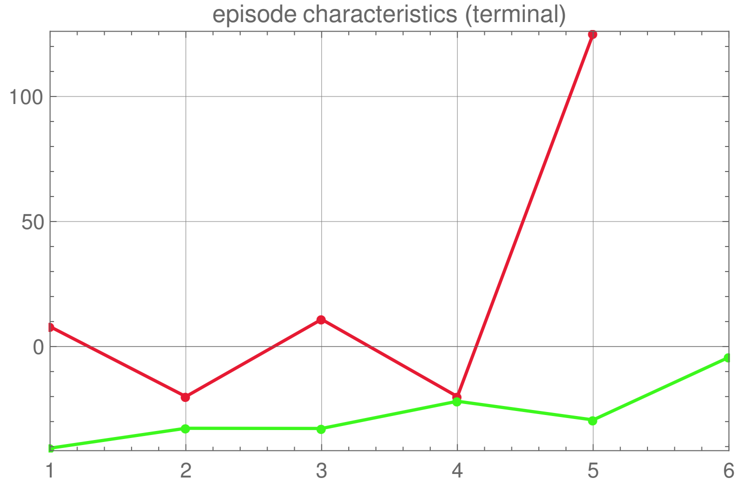
We can also test the trained network by checking whether it can guide us towards a model known in the literature, starting at a nearby state. For example, consider the model from Ref. [131], given by the charge matrix

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 3 & 2 & 0 & -3 & -1 & 0 & -3 & -2 & -2 & 0 & 1 \end{array}\right) . \tag{2.25}$$

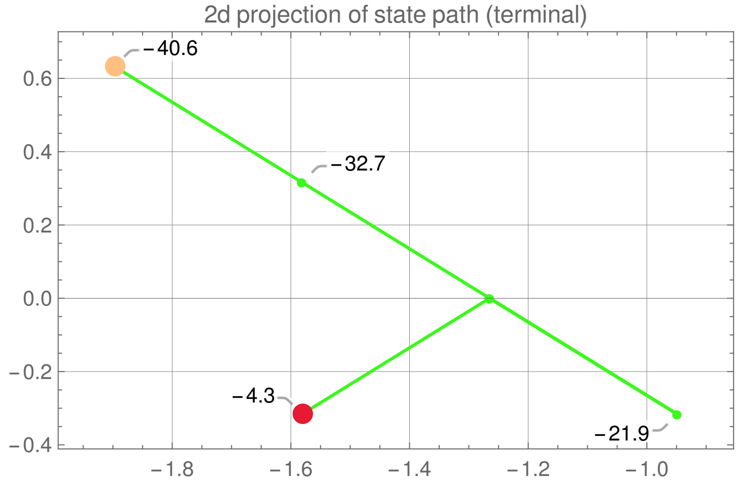
which has an intrinsic value of $\mathcal{V}(\mathcal{Q}) \simeq -4.3$ for a singlet vev $v_1 \simeq 0.159$. Suppose we use the initial state

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 4 & 2 & -3 & -3 & -1 & 0 & -3 & -2 & -2 & -3 & 2 \end{array}\right) . \tag{2.26}$$

which is a perturbation of the literature model (2.25) but, as is, does not amount to a potentially viable model. Generating an episode starting at the state (2.26) then leads to the literature model (2.25) in four steps, as indicated in Fig. 2.3.



(a) Intrinsic value (green) and reward (red) vs episode steps.



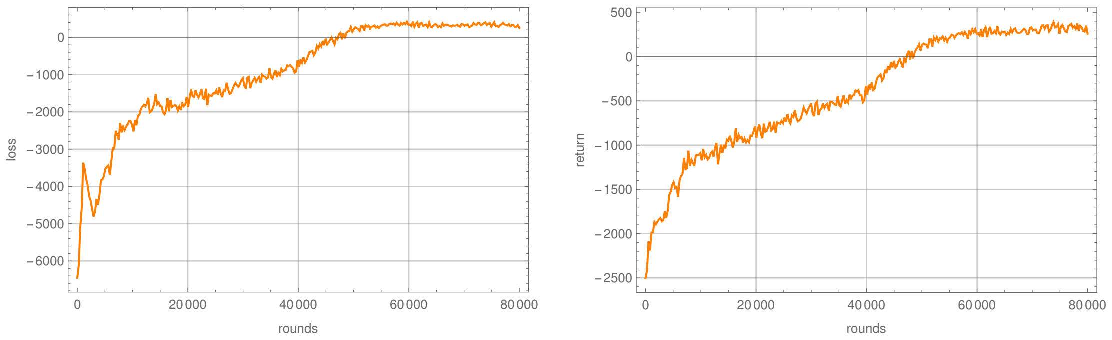
(b) Two-dimensional projection of the evolution of the charge matrices $\mathcal{Q}_t$ of the episode connecting the initial state (2.26) (yellow dot) with the final state (2.25) (red dot). This projection of the environment lattice is onto a random and misaligned plane. The axis labels are the distance from the origin of the lattice, along this plane in units of the lattice spacing. The labels indicate the intrinsic values of each state visited along the episode.

Figure 2.3: Characteristics of the episode connecting the states (2.26) and (2.25).

2.3.2 Two $U(1)$ symmetries

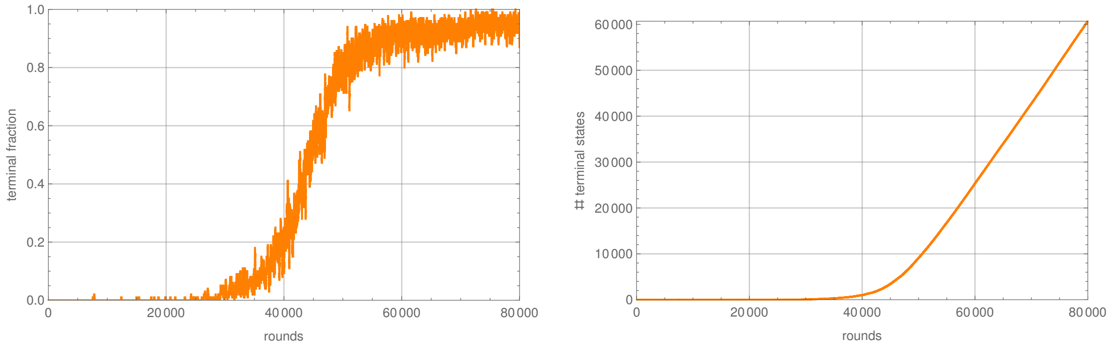
Next, we present results for an environment with two $U(1)$ symmetries and two singlet scalar fields. The entries of the 2×10 charge matrix $\mathcal{Q}$ are constrained as in Eq. (2.12) but we now consider a somewhat smaller range with $-q_{\rm min} = q_{\rm max} = 5$. This still leads to a considerably larger environment than previously, with a total of $11^{20} \sim 10^{21}$ states.

Training for this environment on a single CPU takes about 25 hours and leads to the measurements shown in Fig. 2.4. The networks finds 60686 terminal states which reduce to 57807 once the permutation redundancies (2.13) are removed. Episodes guided by the trained network and with maximal length 32 lead to terminal states in 95% of cases and the average episode length is 19.9 steps. After



(a) Loss vs batch number.

(b) Return vs batch number.



(c) Fractional of terminal episodes vs episode number.

(d) Number of terminal states vs episode number.

Figure 2.4: Training measurements for the case of a two $U(1)$ symmetries and $-q_{\rm min} = q_{\rm max} = 5$.

a Monte-Carlo optimisation of the order one coefficients a_{ij}, b_{ij} we find 2019 from the 57807 models found during training have an intrinsic value $\mathcal{V}(\mathcal{Q}) > -1$. The best of these has charge allocation

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 2 & 2 & 1 & -2 & 0 & 1 & -1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -2 & 0 & 0 & 1 \end{array}\right), \tag{2.27}$$

and an intrinsic value $\mathcal{V}(\mathcal{Q}) \simeq -0.390$, provided we choose singlet vevs $v_1 \simeq 0.079$, $v_2 \simeq 0.112$ and order one coefficients

$$\begin{gathered}(a_{ij}) \simeq \begin{pmatrix} -1.898 & 0.834 & -0.587 \\ -0.575 & -0.592 & 1.324 \\ -1.123 & -1.265 & 0.982 \end{pmatrix}, \\ (b_{ij}) \simeq \begin{pmatrix} -1.759 & 1.358 & 1.013 \\ -1.267 & 1.897 & -1.196 \\ 1.771 & 1.386 & -1.785 \end{pmatrix}. \end{gathered} \tag{2.28}$$

This results in the mass matrices

$$\begin{gathered} M_u \simeq \begin{pmatrix} -0.001 & 0.103 & -0.910 \\ -0.004 & -0.650 & 18.297 \\ -0.098 & -17.489 & 170.815 \end{pmatrix}, \\ M_d \simeq \begin{pmatrix} -0.002 & 0.019 & 0.020 \\ -0.012 & 0.234 & -0.208 \\ 0.218 & 2.149 & -3.910 \end{pmatrix}, \end{gathered} \tag{2.29}$$

and the masses and mixing

$$\begin{gathered} \begin{array}{rcl} (m_u, m_c, m_t) & \simeq & (0.002, 1.210, 172.679)\,\text{GeV} \\ (m_d, m_s, m_b) & \simeq & (0.005, 0.111, 4.476)\,\text{GeV} \end{array}, \\ V_{\text{CKM}} = \begin{pmatrix} 0.975 & -0.223 & 0.004 \\ 0.223 & 0.974 & 0.040 \\ -0.013 & -0.038 & 0.999 \end{pmatrix}, \end{gathered} \tag{2.30}$$

in rough agreement with the values in Table 2.6. More examples of promising models found by the network are listed in App. A.2.

We can also demonstrate that the trained network is capable of finding models which have been constructed in the literature. Consider the model from Ref. [131] which is described by the charge matrix

$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 3 & 0 & 0 & 1 & -1 & 0 & 1 & -4 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 & 0 & -2 & 1 & -1 & 0 & 0 & 1 \end{array}\right) . \qquad (2.31)$$

For singlet vevs $v_1 \simeq 0.158$ and $v_2 \simeq 0.028$ it is a terminal state with intrinsic value $\mathcal{V}(\mathcal{Q}) \simeq -4.1$ which, however, has not been found during training. To see that this model can be obtained we start an episode at a nearby state with charge matrix

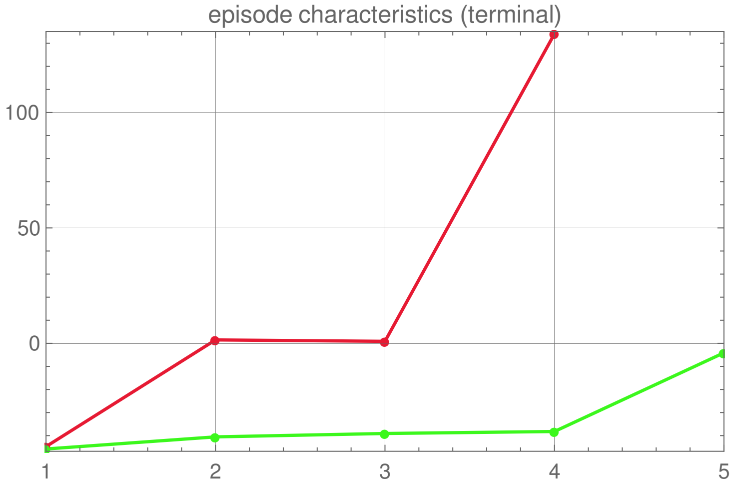
$$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 5 & 0 & 0 & 1 & -1 & 0 & 1 & -4 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 & 0 & -2 & 1 & -1 & 0 & 0 & 3 \end{array}\right) . \qquad (2.32)$$

The trained network then takes us from this state to the literature model (2.31) in three steps, as can be seen in Fig. 2.5.

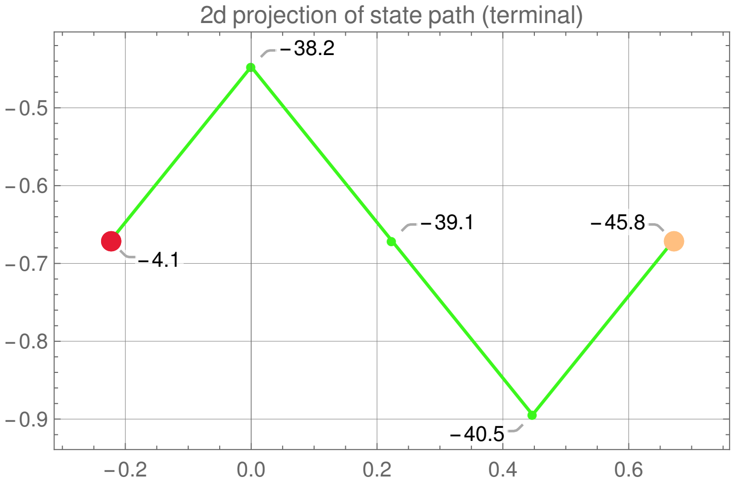
2.4 Summary

In this chapter, we have studied particle physics model building with reinforcement learning (RL). We have focused on a simple model building framework – Froggatt-Nielsen (FN) models for quark masses and mixing – and the simplest policy-based RL algorithm. This should be seen as a simple example for the basic method we consider the remaining chapters. The results show that successful model-building is indeed possible in this way. For both cases we consider, that is, for FN models with one $U(1)$ symmetries and two $U(1)$ symmetries, the network can be trained to settle on a highly efficient policy which leads to terminal states in $> 90\%$ of all cases and in an average number of < 20 steps. The trained networks can be used to find promising models from random initial states and it is capable of finding literature models, having started at a near-by state.

There are numerous extensions of this work. At the most basic level, the order one coefficients which enter FN model building might be incorporated into the RL environment, rather than being fixed to random values as we have



(a) Intrinsic value (green) and reward (red) vs episode steps.



(b) Two-dimensional projection of the evolution of the charge matrices $\mathcal{Q}_t$ of the episode connecting the initial state (2.32) (yellow dot) with the final state (2.31) (red dot). This projection of the environment lattice is onto a random and misaligned plane. The axis labels are the distance from the origin of the lattice, along this plane in units of the lattice spacing. The labels indicate the intrinsic values of each state visited along the episode.

Figure 2.5: Characteristics of the episode connecting the states (2.32) and (2.31).

done here for simplicity. It would also be desirable to study the performance of other RL algorithms, such as actor-critic set-ups and Q-learning, on the FN environment. Extending the environment to include lepton masses is another interesting direction. More generally, we can ask if other areas of particle physics model building can be approached in this way. For example, can RL be used for dark matter model building?

At its most ambitious, this line of thought suggests an RL environment which consists of large classes of quantum field theories, for example a large class of extensions of the standard model of particle physics, prescribed in some appropriate manner. The intrinsic value of such models might be determined by comparing their predictions with a wide range of experimental data. Realising such an environment would require significantly more theoretical preparation than was necessary for the FN environment as well as more computing power. However, the benefits of such a system might be considerable – it would allow exploring large classes of standard model extensions and their consistency with experimental data and may help to find the correct path for physics beyond the standard model.

3
Sums of Line Bundles

Contents

In this chapter, we consider the application of GA to sums of line bundles. These are some of the simplest geometric constructions for quasi-realistic models of particle physics from the heterotic string [62, 63, 132, 133], and should be considered as very special loci in the moduli space of non-Abelian bundles. Large scans have already been attempted in this setting, and so give us a benchmark for how effective GA can be as a tool for discrete optimisation. We also extend these results to other manifolds that would have been impossible to search exhaustively.

This chapter begins with an introduction to line bundle models in Sec. 3.1, before moving on to encoding these models for GA and the final result in Sec. 3.2 and Sec. 3.3 respectively.

3.1 Heterotic Line Bundle Models

As introduced in Chapter 1, the earliest and arguably one of the most promising proposal for string phenomenology is the $E_8 \times E_8$ heterotic string compactified on smooth Calabi-Yau threefolds with holomorphic vector bundles. In this context, the geometrical data describing the additional dimensions consists of a Calabi-Yau threefold X which reduces the number of large space-time dimension from 10 to 4, and a holomorphic bundle V on X, needed to break the $E_8 \times E_8$ gauge symmetry to the Standard Model gauge group or to one of its grand unification embeddings.

In our earlier discussions we introduced CICYs, as one particular class of Calabi-Yau manifolds, and the simplest vector bundles over them – line bundles. Direct sums of these line bundles offer the simplest option in constructing the vector bundle V. In this case, two of the major technical difficulties, checking slope-stability of the bundle and checking the low-energy spectrum, become manageable. Stability checks are relatively straightforward due to the split nature of the bundle, while computations of the spectrum are made virtually instantaneous by the discovery of line bundle cohomology formulae, as described in Sec. 1.4.2. As a result, deciding the physical viability of a heterotic line bundle sum model at the level of the particle spectrum (three families of quarks and leptons, the presence of a Higgs field and the absence of any exotic matter charged under the Standard Model gauge group) can be accomplished within a fraction of a second, something that has never been possible before. By comparison, traditional constructions in the literature have taken several years of laborious work to achieve a comparable level of analysis.

Another salient feature of line bundle models is the presence of additional $U(1)$-symmetries, which can restrict the allowed operators in the Lagrangian in a way that is robust against deformations away from line bundle sums [64]. In this way, the $U(1)$ symmetries give rise to Froggatt-Nielsen models of fermion masses and mixings within string theory. This is the same mechanism described in Chapter 2, but in this case with a very specific choice of charges which depend on the line bundles the multiplets descend from.

More concretely, throughout this chapter, V will be a rank-5 line bundle sum $V = \oplus_{a=1}^{5} L_a$ over a Calabi-Yau threefold X, so that the resulting model has $SU(5) \times S(U(1)^5)$ symmetry. The notation $L_a = \mathcal{O}_X(k_a)$ indicates a line bundle with first Chern class $c_1(L_a) = k_a^i J_i$, where k_a^i are the components of the integer vectors $k_a \in \mathbb{Z}^h$ and $(J_1, \ldots, J_h)$ is a suitably chosen basis of $H^2(X, \mathbb{Z})$, with dimension $h = h^{1,1}(X)$. The five integer vectors $(k_1, \ldots, k_5)$ uniquely specify the line bundle sum V. The manifold X will be assumed to admit a free action of a non-trivial discrete group Γ, such that the quotient manifold X/Γ has a non-trivial fundamental group (in fact, isomorphic to Γ). Given such a group action, there are, in general, several ways to break $SU(5)$ to the SM group using an appropriate discrete Wilson line on X/Γ. Fixing X, the aim will be to identify the line bundle sums V that satisfy the following constraints:

(C1) E_8 embedding $c_1(V) = \sum_{a=1}^{5} k_a \overset{!}{=} 0$.

In order to guarantee that the structure group of V is $S(U(1))^5$ and not smaller, no proper subsets of line bundles in V are allowed to have a vanishing first Chern class.

(C2) Anomaly cancellation

$$c_{2,i}(V) = -\frac{1}{2} d_{ijk} \sum_{a=1}^{5} k_a^j k_a^k \overset{!}{\leq} c_{2,i}(TX) \ ,$$

$\forall i = 1, \ldots, h$, where d_{ijk} denote the triple intersection numbers and $c_2(TX)$ the second Chern class of the tangent bundle of X, relative to the basis $(J_1, \ldots, J_h)$.

(C3) Supersymmetry/poly-stability There exists a non-trivial common solution t^i to the vanishing slopes

$$\mu(L_a) = d_{ijk} k_a^i t^j t^k \overset{!}{=} 0 \text{ for } a = 1, \ldots, 5$$

such that $J = t^i J_i$ is in the interior of the Kähler cone, which in our examples corresponds to $t^i > 0$. Solving the slope-zero equations is computationally expensive and this check is replaced by the weaker condition that each of the five matrices $M_a = (d_{ijk}k^i_a)$ has at least one positive and one negative entry. Moreover, the same should hold for every linear combination $v^a M_a$. In practice, considering all the vectors v^a with integer entries between -2 and 2 provides a strong enough check.

(C4) Spectrum: cohomology dimensions must satisfy

10-multiplets: $h^1(X,V) = 3|\Gamma|$
no $\overline{\mathbf{10}}$-multiplets: $h^2(X,V) = 0$
$\overline{\mathbf{5}}$-multiplets: $h^1(X,\wedge^2 V) = 3|\Gamma| + n_h$, $n_h > 0$
Higgs: $h^2(X,\wedge^2 V) = n_h$

Here $|\Gamma|$ is the order of the discrete group Γ and n_h represents the number of Higgs doublet pairs. In the absence of a cohomology formula, (C4) can be replaced by the weaker constraint (C4').

(C4') Chiral spectrum $\chi(X,V) = \chi(X,\wedge^2 V) = 3|\Gamma|$

(C5) Equivariance Require that V descends to a bundle on X/Γ. For symmetries acting trivially on the basis $(J_1,\dots,J_h)$ we require that the Euler characteristic of every (maximal) partial sum $\oplus_{a_i} L_{a_i}$ in V consisting of line bundles with identical first Chern classes, is divisible by $|\Gamma|$. For symmetries with a non-trivial action on the basis $(J_1,\dots,J_h)$, V must admit a partition into partial sums that are invariant under the induced action of Γ on $(J_1,\dots,J_h)$ and, moreover, the Euler characteristic of each partial sum must be divisible by $|\Gamma|$.

The GA scans discussed below have been carried out on four different Calabi-Yau threefolds realised as complete intersections in products of projective spaces. Using the standard notation for configuration matrices, with superscript indices on X indicating the Hodge numbers $(h^{1,1}(X), h^{1,2}(X))$ and a subscript index indicating the position in the CICY list [98], these four manifolds are described by the following defining matrices:

$$X^{(4,68)}_{7862} = \begin{array}{c}\mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\end{array}\left[\begin{array}{c}2\\2\\2\\2\end{array}\right] , \quad X^{(5,45)}_{7447} = \begin{array}{c}\mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\end{array}\left[\begin{array}{cc}1&1\\1&1\\1&1\\1&1\\1&1\end{array}\right]$$

$$X^{(6,30)}_{5302} = \begin{array}{c}\mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\end{array}\left[\begin{array}{ccc}0&1&1\\0&1&1\\1&1&0\\1&1&0\\1&0&1\\1&0&1\end{array}\right] , \quad X^{(7,27)}_{4071} = \begin{array}{c}\mathbb{P}^1\\ \mathbb{P}^2\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^1\\ \mathbb{P}^2\\ \mathbb{P}^3\end{array}\left[\begin{array}{cccccccc}1&1&0&0&0&0&0&0\\0&1&1&0&0&0&1&0\\0&0&1&0&0&1&0&0\\0&0&0&0&2&0&0&0\\0&0&0&1&1&0&0&0\\1&0&0&0&0&1&0&1\\0&0&0&1&1&0&1&1\end{array}\right] . \tag{3.1}$$

All four embeddings are favourable, in the sense that a basis $(J_1, \dots J_h)$ of $H^2(X, \mathbb{Z})$ can be obtained by pulling back to X the Kähler classes of the h projective factors. Line bundle cohomology formulae on the manifolds X_{7862} and X_{7447}, used to implement the constraints (C4) in the GA searches, are presented in Appendix B of [21]. For the manifolds X_{5302} and X_{4071} cohomology formulae are not yet available and we have used the weaker spectrum constraint (C4’). The first three manifolds admit symmetries of orders 2 and 4 which leave the basis $(J_1, \dots J_h)$ invariant, while X_{4071}, admits a free action by $\mathbb{Z}_2$ which maps $(J_1, J_2, J_3, J_4, J_5, J_6, J_7) \mapsto (J_1, J_6, J_3, J_4, J_5, J_2, J_7)$.

3.2 Encoding Sums of Line Bundles for GA

Fixing the manifold X, a sum of five line bundles V is specified by $4h$ integers $(k^i_a)^{i=1,\dots,h}_{a=1,\dots 4}$, where the condition (C1) is used to fix the fifth line bundle in terms of the first four. There are no a priori bounds on these $4h$ integers. However, previous experience from systematic scans [133, 134] indicates that only a relatively small range is relevant, as bundles involving larger integers either violate the anomaly cancellation condition or fail to match the required Euler characteristic. We choose this range as $k_a^i \in \{-2^n+1, \ldots, 2^n\}$, so that every integer can be encoded by $n+1$ bits without redundancy, and a complete model is described by a bit list of length $N_{\text{bits}} = 4h(n+1)$. In practice, we take $n = 3$ for the first three manifolds and $n = 2$ for the manifold X_{4071}.

The classic GA algorithm was introduced in Sec. 1.5.1. To decide how successful a particular individual is, GA requires we define a *fitness function* $f : \mathbb{F}_2^{N_{\text{bits}}} \to \mathbb{R}$ on this set of binary strings, which indicates how close the corresponding bundle comes to satisfying conditions (C1)–(C5). The detailed definition of f is presented in Sec 3.2.1. The population is then evolved via the three main evolutionary ingredients: selection, breeding and mutation. In this work a single point cross-over performs well enough, in which a cut is made at a single random point and the 'tails' swapped, and our implementation includes *elitism*, which means that the fittest individual in every generation is copied to the next generation without modification. We explored the possibility of copying multiple individuals from the previous generation, but found little change in the performance.

3.2.1 The Fitness Function

The fitness function $f(X, V)$ is the measure of how close a line bundle sum V over a smooth Calabi-Yau threefold X comes to satisfy the constraints (C1)–(C5) of Sec. 3.1. It receives several contributions,

$$f = f_{\text{anom}} + f_{\text{ind}} + f_{\text{slope}} + f_{\text{equiv}} + f_{\text{str. gp}} + f_{\text{spec}} \ , \tag{3.2}$$

which we now discuss in turn. The contribution associated with the cancellation of anomalies is

$$f_{\text{anom}} = 10 \sum_{i=1}^{h} \frac{\min(c_{2,i}(V) - c_{2,i}(TX), 0)}{h k_{\text{max}}^2 \text{rk}(V)}, \tag{3.3}$$

where $k_{\text{max}} = 2^n$ is the (absolute) maximal line bundle integer allowed in the search. The sum contains h terms that are quadratic in the line bundle integers, hence the pre-factor $(h k_{\text{max}}^2)^{-1}$. The factor of 10 and the numerical factors appearing below in the expressions for the other contributions to the fitness function are arranged such that all contributions are of roughly the same order of magnitude for a typical bundle V.

The contribution from the Euler characteristic of V is

$$f_{\text{ind}} = -100 \frac{|\text{ind}(V) + 3|\Gamma||}{h k_{\text{max}}^3 \text{rk}(V)} . \tag{3.4}$$

The necessary slope-0 checks discussed under (C3) involve a number of matrices which are required to have both positive and negative entries. If a number n_{pos} of these matrices are found to have non-negative entries only and a number n_{neg} are found to have non-positive entries only, there is a (negative) contribution to the fitness function equal to

$$f_{\text{slope}} = -\frac{n_{\text{pos}} + n_{\text{neg}}}{10} . \tag{3.5}$$

For equivariance, in the case of symmetries acting trivially on the second cohomology of X we have a contribution

$$f_{\text{equiv}} = -\sum_{\text{distinct } L \subset V} m(L) \chi(X, L) \text{mod} |\Gamma| , \tag{3.6}$$

where $m(L)$ is the number of times L appears in the line bundle sum V.

Symmetries with a non-trivial action Γ on the second cohomology permute non-trivially the projective space factors in the embedding of X, which amounts to a permutation of line bundle integers in V. The fitness contribution from equivariance then has to take into account two aspects. On the one hand, in the ideal case the permutation induced by Γ on V will produce a bundle that can be identified with V up to re-orderings of the line bundles. The failure to achieve this is measured by summing over the absolute values of the differences between the line bundle integers in V and the integers obtained after applying the Γ-permutation and a line bundle re-ordering, and minimising over all possible re-orderings. We call this minimal sum M, which in the ideal case vanishes. Furthermore, provided that $M = 0$, we compute an equivariance contribution analogous in spirit to (3.6) above, with the difference that we first partition the line bundles in V into parts that are formed from the cycles of the Γ-permutation, compute their Euler characteristics, mod out by $|\Gamma|$ and sum over all the parts. We denote this value as N. With these considerations, the total fitness contribution from equivariance is taken to be

$$f_{\text{equiv}} = \begin{cases} -|\Gamma| - M & M \neq 0 \\ -N & M = 0 \end{cases}, \tag{3.7}$$

where in the case $M \neq 0$ we have added a default penalty of $-|\Gamma|$, corresponding to the maximal penalty that can be accrued from N when $M = 0$. This default penalty provides an incentive to evolve towards achieving $M = 0$.

The contribution from the constraint on the structure group is given by

$$f_{\text{equiv}} = -\frac{|S|}{10} , \tag{3.8}$$

where S is the collection of subsets of line bundles in V whose sum of first Chern class vanishes.

The contribution corresponding to the spectrum, on manifolds where a cohomology formula is available, is computed as

$$\begin{aligned} f_{\text{spec}} =& -1000\frac{h^0(X,V)+h^3(X,V)+h^0(X,\wedge^2 V)+h^3(X,\wedge^2 V)}{hk_{max}^3 \text{rk}(V)^2} \\ & -100\frac{h^2(X,V)}{hk_{max}^3} + \frac{\theta(h^2(X,\wedge^2 V) - 1/2) - 1}{10} \\ & -5\frac{\max(h^2(X,\wedge^2 V)/|\Gamma| - 2, 0)}{hk_{max}^3 \text{rk}(V)} , \end{aligned} \tag{3.9}$$

where the first term corresponds to the requirement that the zeroth and the top cohomologies of V have to vanish in view of slope-stability, while the other terms correspond to the exact spectrum constraints in (C4).

Concretely, the second term corresponds to the absence of $\overline{\mathbf{10}}$-multiplets, the third term penalises the lack of Higgs multiplets, while the fourth term penalises the presence of more than two pairs of Higgs doublets. There is no further contribution from the $\mathbf{10}$-multiplets and the $\overline{\mathbf{5}}$-multiplets, given the f_{ind} contribution in (3.4) above. In the absence of a cohomology formula we set $f_{\text{spec}} = 0$, since the index constraint (C4') has already been taken care of in (3.4).

3.3 Results

We have implemented the classic genetic algorithm and the line bundle environment (performing the binary encoding and the computation of the fitness function) in C, and the code is available here [125, 135]. We performed 7 different searches, as summarised in Table 3.1. Each search was divided into a large number of genetic episodes, with every episode containing 300 generations of 300 individuals each. The mutation rate was set to 0.5%, and $\alpha = 3$.

Manifold	h	$\|\Gamma\|$	Range	GA	Scan	Found	Explored
7862	4	2	[-7,8]	5	5	100%	10^{-10}
7862	4	4	[-7,8]	30	31	97%	10^{-10}
7447	5	2	[-7,8]	38	38	100%	10^{-14}
7447	5	4	[-7,8]	139	154	90%	10^{-14}
5302	6	2	[-7,8]	403	442	93%	10^{-19}
5302	6	4	[-7,8]	722	897	80%	10^{-19}
4071	7	2	[-3,4]	11,937	N/A	N/A	10^{-14}

Table 3.1: Summary of results for the 7 GA searches. The table compares the number of models found here (GA) with numbers found in previous comprehensive searches (scan) for manifolds with $h < 7$, both as actual numbers and as percentages. For the first three manifolds these numbers refer to the models that pass a sufficient criterion for poly-stability, performed after the GA search. The last column indicates the fraction of the environment explored in the GA search.

3.3.1 The manifolds X_{7862}, X_{7447} and X_{5302}

Systematic and comprehensive scans on these manifolds have been previously carried out in Ref. [133]. On the manifold X_{5302} a search using reinforcement learning was carried out in Ref. [136]. Our purpose here is to gauge the GA performance as a heuristic method of searching. The results are surprising. For the manifold X_{7862} with $h^{1,1}(X) = 4$, the environment contains $\sim 10^{19}$ line bundle sums[1]. All

---

[1] The comprehensive scan of Ref. [133] on environments of this size was only possible due to the split nature of the bundle, which implied that vast regions of the solution space could be discarded by imposing constraints on individual line bundles, pairs of line bundles etc. The present GA search does not make use of such simplifications.

$\mathbb{Z}_2$-models and 97% of the $\mathbb{Z}_4$-models were found after visiting a fraction of 10^{-10} of this environment. For the manifold X_{7447} with $h^{1,1}(X) = 5$, the size of the environment is $\sim 10^{24}$. All $\mathbb{Z}_2$-models and 90% of the $\mathbb{Z}_4$-models were found after visiting an even smaller fraction of 10^{-14} of the environment. Most impressively, for the manifold X_{5302} with $h^{1,1}(X) = 6$ the environment contains $\sim 10^{29}$ bundles and after visiting only a tiny fraction of 10^{-19} of it, 93% of the $\mathbb{Z}_2$-models and 80% of the $\mathbb{Z}_4$-models were found.

In figure 3.1 we present the saturation curve for the number of inequivalent $\mathbb{Z}_4$-models found in the GA search on X_{7447} as a function of the number of states visited. All the other cases are qualitatively similar. An important common feature of these saturation curves, relevant for evaluating the performance of the GA, is that the initial rate of finding new viable models is of order 1 (inequivalent) models per 100 episodes. This implies that, although the size of the environment increases by several orders of magnitude with every additional Kähler parameter, while the number of viable models is expected to increase only by an order of magnitude, the initial rate at which GA identifies these is independent of the number of Kähler parameters.

The computational time required for a genetic episode is $O(10)$ seconds on a single CPU and displays a linear increment with the number of Kähler parameters ($\sim$23 seconds for X_{5302}, compared to $\sim$17.5 seconds for X_{7447} and $\sim$12 s for X_{7862}). This means that each of the searches mentioned above finished within a few hours on a cluster of 100 CPUs.

3.3.2 The manifold X_{4071}

The manifold X_{4071} pushes the search for realistic string models of particle physics into a new realm of larger Picard numbers. The favourable representation shown in Eq. (3.1) was taken from the maximally favourable CICY list of Ref. [137] and the $\mathbb{Z}_2$-symmetry from the recent classification of cyclic freely acting symmetries undertaken in Ref. [101]. The group $\mathbb{Z}_2$ acts non-trivially on the second cohomology of the manifold, leading to a more involved bundle equivariance check as discussed

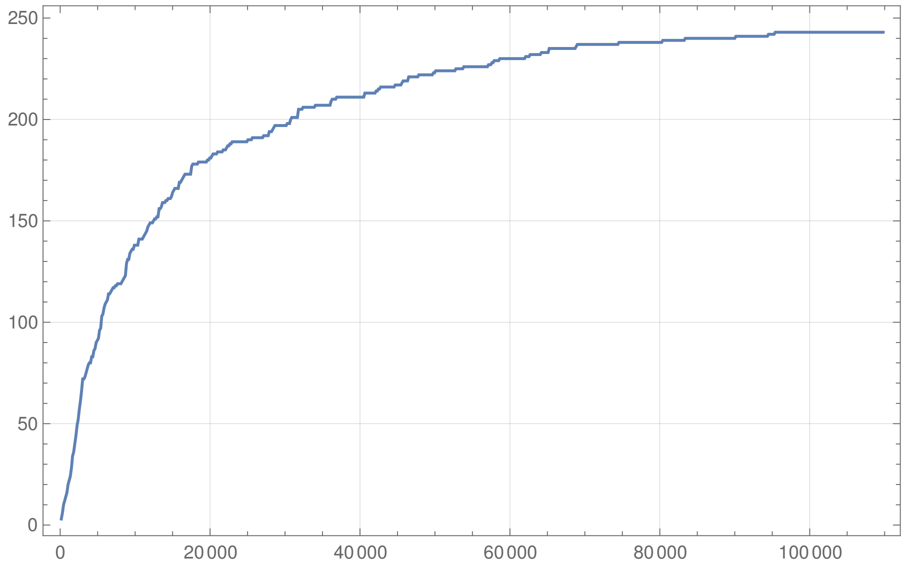


Figure 3.1: Saturation plot for the GA search on X_{7447} with $|\Gamma| = 4$ and $h^{1,1}(X_{7447}) = 5$. The horizontal axis represents the number of genetic episodes, in each episode a number of 90,000 states being visited. The vertical axis corresponds to the number of inequivalent models found in the search satisfying the necessary criterion (C3) for poly-stability. The computational time for a genetic episode is $O(10)$ seconds on a standard machine.

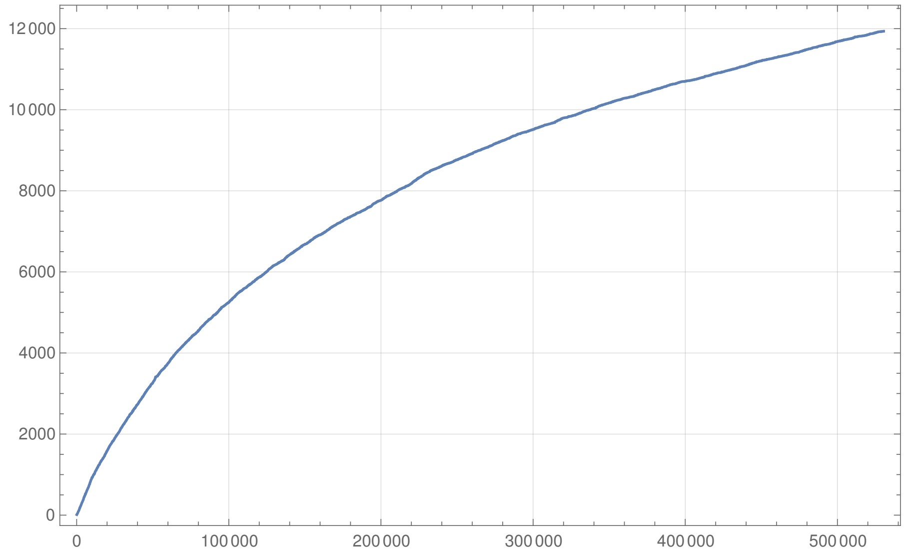


Figure 3.2: Saturation plot for the GA search on X_{4071} with $|\Gamma| = 2$ and $h^{1,1}(X_{4071}) = 7$. The horizontal axis represents the number of genetic episodes, each episode containing a number of 90,000 visited states. The vertical axis corresponds to the number of inequivalent models found in the search, satisfying the necessary criterion (C3) for poly-stability. The computational time for a genetic episode is $O(1)$ minute on a standard machine.

in Section 3.1. To our knowledge, this is the first instance when a symmetry of this type was considered for the purpose of a large-scale search.

The classification of Ref. [101] gives examples of favourable CICYs with numbers of Kähler parameters as large as 15 that admit freely acting cyclic symmetries. All of these examples involve non-trivial actions on the second cohomology and the methods discussed here for the manifold X_{4071} are directly applicable in these cases. Performing searches on such manifolds would very likely result in a plethora of viable string models.

The saturation curve for the number of inequivalent models found on X_{4071} as a function of the number of states visited is shown in figure 3.2. The plot indicates that after 500,000 genetic episodes saturation has not been reached. However, by doubling the computational time a good degree of saturation would likely be achieved.

Due to the more involved equivariance checks, the computational time required for a single genetic episode was slightly longer than for the previous manifolds and averaged at around 1 minute.

3.4 Summary

The main lesson emerging from the present chapter is that the size of the string landscape is no longer a major impediment in the way of constructing quasi-realistic string models of Particle Physics. Using genetic algorithms we were able to scrutinise spaces of string compactifications of sizes as large as $\sim 10^{30}$ and identify the vast majority of good solutions residing in these spaces after visiting only a tiny fraction of the total number of states. This has been carried out for heterotic line bundle models on Calabi-Yau threefolds with $4, 5, 6$ and 7 Kähler parameters but the basic methodology applies to other string constructions as well.

Extending the methods to manifolds with larger numbers of Kähler parameters (for example the favourable CICYs of Ref. [101] which include examples with up to 15 Kähler parameters) is perfectly achievable. While finding all the models that satisfy the conditions (C1)–(C5) of Section 3.1 may be impossible on such manifolds (the expected numbers of viable models being too large even to store), our results suggest that GAs provide the means to produce string models on demand. The method can also be extended to manifolds from the Kreuzer-Skarke list [18], provided that more examples with a non-trivial fundamental group are found [138]. In order to impose the full spectrum constraints (as we did here for the manifolds X_{7862} and X_{7447}), explicit line bundle cohomology formulae need to be obtained, using a combination of algebro-geometric methods [139, 140] and machine learning techniques [112, 141].

It is important to stress that the constraints (C1)–(C5) of Section 3.1 do not represent a complete list of requirements, but only lead to a broad brush version of the Standard Model. The formidable success of GAs and the immediate access to cohomology data provided by the line bundle cohomology formulae offer the possibility of substantially refining the requirements. One of the many ways in which this can be accomplished would proceed by first identifying a number of Froggatt-Nielsen models with four flavour $U(1)$ symmetries that can explain the observed hierarchies of fermion masses and mixings. The successful Froggatt-Nielsen models would then correspond to specific $U(1)$-charge assignments for the bundle moduli fields, measured by $H^2(X, V \otimes V^*)$, which would be searched for along with imposing the other constraints. Other requirements that can be included are the absence of fast proton decay operators, as well as various model-dependent constraints for neutrino physics. Such additional requirements will likely lead to a large reduction in the number of viable models, making complete (and more targeted) GA searches possible even for models with a larger number of Kähler parameters. From the results of this chapter, the expectation is that, despite GA being a heuristic method, nearly all the models returned from such a search will be almost complete. As a result, if such a search returns no models, it can be concluded that there is likely no realistic particle physics in the environment.

Finally, as mentioned at the start of this chapter, these models should be considered as very special loci in moduli space which simplify a number of the calculations. Although the chiral spectrum remains unchanged as we move away from the line bundle locus, it is important to find realisations of the full non-Abelian bundles as these will play a key role in finding the correct Yukawa couplings and stabilising bundle moduli [74–76, 85].

4

Monad Bundles

Contents

We now consider the more ambitious goal of constructing non-Abelian vector bundles over Calabi-Yau manifolds. As mentioned earlier, the line bundle sums constructed in Chapter 3 should be considered as special loci in the space of non-Abelian bundles [105, 142].

Monad bundles are a particularly convenient approach to constructing non-Abelian bundles, as they are also constructed from line bundles. We introduce monad bundles and their application to heterotic compactifications in Sec. 4.1, before searching the space of monad bundles with RL and GA in Sec. 4.2 and

Sec. 4.3 respectively. The results of the two methods are then compared in Sec. 4.4 before we summarise in Sec. 4.5.

4.1 Heterotic Models from Monad Bundles

Explicit bundle constructions frequently rely on holomorphic line bundles as basic building blocks. As described in Sec. 1.4.2, holomorphic line bundles are classified by their first Chern class. Relative to an integral basis $(J_1, \ldots, J_h)$ of $H^2(X)$, the first Chern class of a line bundle $L \to X$ can be represented by an integer vector $\mathbf{k} \in \mathbb{Z}^h$, such that $c_1(L) = k^i J_i$, hence the notation $L = \mathcal{O}_X(\mathbf{k})$.

Monad bundles are defined by a short exact sequence

$$0 \longrightarrow V \longrightarrow B \xrightarrow{f} C \longrightarrow 0 \;, \qquad B = \bigoplus_{i=1}^{r_B} \mathcal{O}_X(\mathbf{b}_i) \;, \quad C = \bigoplus_{a=1}^{r_C} \mathcal{O}_X(\mathbf{c}_a) \;, \tag{4.1}$$

where, for our purposes, B and C are each taken to be line bundle sums. Thus such monad bundles are specified by a pair (B, C) of line bundle sums which can be identified with an integer matrix $(\mathbf{b}_1, \ldots, \mathbf{b}_{r_B}, \mathbf{c}_1, \ldots, \mathbf{c}_{r_C})$ of size $h \times (r_B + r_C)$. The bundle V is isomorphic to the kernel, $\mathrm{Ker}(f)$ of the monad map f which can be thought of as a $r_C \times r_B$ matrix with polynomial entries $f_{ai} \in \Gamma(\mathcal{O}_X(\mathbf{c}_a - \mathbf{b}_i))$.

A given choice of (B, C) might lead to a monad sheaf, rather than a monad bundle. This happens when the degeneracy locus of f, that is, the locus on X where the rank of f is less than maximal, is non-trivial. Since we are considering smooth, geometrical models we should avoid sheafs, so we have to check "bundleness" of V by computing the dimension

$$d_{\mathrm{deg}} = \text{dimension of degeneracy locus of } f. \tag{4.2}$$

Typically, the matrix f exhibits a pattern of zero and non-zero entries, depending on whether $\Gamma(\mathcal{O}_X(\mathbf{c}_a - \mathbf{b}_i))$ is trivial or non-trivial, and for otherwise generic choices, the dimension d_{deg} can be determined from this pattern. The model is acceptable iff $d_{\mathrm{deg}} = -1$.

A monad map with too many zero entries might also lead to a split bundle, that is, to a bundle V whose structure group is a proper sub-group of $SU(4)$, for example

$S(U(2)\times U(2))$. These split bundles lead to a GUT group larger than $SO(10)$ and we do not consider this possibility here. To this end, we introduce the number

$$n_{\text{split}} = \text{number of splits of the bundle}\, V \tag{4.3}$$

which can be determined from the pattern of the monad map. It is also useful to introduce the number

$$n_{\text{trivial}} = \text{number of trivial line bundles } \mathcal{O}_X \text{ in } (B, C)\ . \tag{4.4}$$

As mentioned earlier, Γ-equivariance of V is a non-trivial constraint but a full check based on presently available methods is not compatible with an efficient computational realisation. Instead, we will perform a strong necessary check for B and C to possess a Γ-equivariant structure. This is done by checking that each unique line bundle L in B or C has an index divisible by $|\Gamma|$. If a line bundle L appears with multiplicity m in either B or C, we require that

$$\text{ind}(L^{\oplus m}) \quad \text{is divisible by} \quad |\Gamma|\ . \tag{4.5}$$

To check the other constrains we require the following formulae for the Chern classes of a monad bundle:

$$\begin{array}{rcl}
\text{rk}(V) & = & \text{rk}(B) - \text{rk}(C) \overset{!}{=} 4 \\
c_1^k(V) & = & c_1^k(B) - c_1^k(C) = \sum_{i=1}^{r_B} b_i^k - \sum_{a=1}^{r_C} c_a^k \overset{!}{=} 0 \\
c_{2k}(V) & = & \text{ch}_{2k}(C) - \text{ch}_{2k}(B) = \frac{1}{2} d_{klm} \left(\sum_{a=1}^{r_C} c_a^l c_a^m - \sum_{i=1}^{r_B} b_i^l b_i^m \right) \overset{!}{\leq} c_{2k}(TX) \\
\text{ind}(V) & = & \sum_{q=0}^{3} (-1)^q h^q(X, V) = \frac{1}{2} c_3(V) = \text{ch}_3(B) - \text{ch}_3(C) \\
 & = & \frac{1}{6} d_{klm} \left(\sum_{i=1}^{r_B} b_i^k b_i^l b_i^m - \sum_{a=1}^{r_C} c_a^k c_a^l c_a^m \right) \overset{!}{=} -3|\Gamma|\ .
\end{array} \tag{4.6}$$

The conditions appearing on the right-hand sides of these equations originate from the requirement that V has an $SU(4)$ structure group, from the anomaly cancellation condition (1.34) and from the constraint $\text{ind}(V) = n_{\mathbf{16}} - n_{\overline{\mathbf{16}}}$ on the chiral asymmetry.

Calculating cohomology for monad bundles requires the long exact sequence in cohomology associated to the monad sequence (4.1). Such computations can require ranks of maps and can be time-consuming. For this reason, checking Hoppe's criterion for poly-stability in full is not feasible while running a GA. Instead, we will check if Hoppe's criterion is violated and, hence, that V is non-supersymmetric if

$$h^0(B) - h^0(C) > 0 \quad \text{or} \quad h^0(B^*) - h^0(C^*) > 0 \; . \tag{4.7}$$

Fortunately, analytic formulae for zeroth line bundle cohomology exist for both manifolds under considerations [22], so the above expressions can be checked efficiently.

4.2 Learning Heterotic Monad Bundles

We now move on to our main interest, namely applying RL to an environment of monad bundles. One of the properties which needs to be incorporated in this context is the index of the bundle, and the experience with the previous line bundle environment suggests how to accomplish this.

We concentrate on $SU(4)$ monad bundles on two of the CICY three-folds, namely the bicubic CY, X_{7884}, and the triple trilinear CY, X_{7669}, which are given by the defining matricies

$$X_{7884} = \begin{array}{c} \mathbb{P}^2 \\ \mathbb{P}^2 \end{array} \left[\begin{array}{c} 3 \\ 3 \end{array} \right] \; , \; X_{7669} = \begin{array}{c} \mathbb{P}^2 \\ \mathbb{P}^2 \\ \mathbb{P}^2 \end{array} \left[\begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right] . \tag{4.8}$$

Both of these admit freely acting $\mathbb{Z}_3 \times \mathbb{Z}_3$ discrete symmetries at certain loci in complex structure moduli space and can therefore be used for $SO(10)$ model building. We expect our methods can be applied to other CICYs and quite likely other classes of manifolds and types of bundles as well, but our purpose here is not to be exhaustive. Rather, we would like to show that RL can successfully engineer string models with prescribed properties in a context where systematic scans would fail.

On a CY three-fold X with Picard number $h = h^{1,1}(X)$, our environment consists of monad bundles of the form (4.1) given by two line bundle sums (B, C), with fixed ranks $r_B = \mathrm{rk}(B)$ and $r_C = \mathrm{rk}(C)$ such that $r_B - r_C = 4$. We also build in the Chern class condition $c_1(V) = 0$, that is, we restrict the environment to pairs (B, C) with $c_1(B) = c_1(C)$. Concretely, we think of these states as $h \times (r_B + r_C)$

integer matrices $(B, C) = (\mathbf{b}_1, \ldots, \mathbf{b}_{r_B}, \mathbf{c}_1, \ldots, \mathbf{c}_{r_C}) = (b_i^k, c_a^k)$, where each column corresponds to a line bundle. In practice, the size of these integers has to be limited to a certain range, so the environment is defined as

$$\mathcal{S} = \left\{ (\mathbf{b}_1, \ldots, \mathbf{b}_{r_B}, \mathbf{c}_1, \ldots, \mathbf{c}_{r_C}) \,|\, b_{\text{min}} \leq b_i^k \leq b_{\text{max}},\ c_{\text{min}} \leq c_a^k \leq c_{\text{max}}, \sum_{i=1}^{r_B} \mathbf{b}_i = \sum_{a=1}^{r_C} \mathbf{c}_a \right\} \tag{4.9}$$

The actions should amount to a "minimal" modification of a state (B, C) but also have to leave the condition $c_1(B) = c_1(C)$ intact. This can be accomplished by adding ± 1 to one entry of a line bundle in B and perform the same action in the same row for a line bundle in C, while keeping all other line bundles unchanged. Hence, the action space is of the form

$$\mathcal{A} = \{ \mathbf{b}_i \mapsto \mathbf{b}_i \pm \mathbf{e}_k,\ \mathbf{c}_a \mapsto \mathbf{c}_a \pm \mathbf{e}_k \,|\, i = 1, \ldots, r_b,\ a = 1, \ldots, r_C,\ k = 1, \ldots, h \} \ , \tag{4.10}$$

where $\mathbf{e}_k$ is the k^{th} standard unit vector in h dimensions.

To define the reward we proceed as we did for the line bundle environment and first define an intrinsic state value $v(B, C)$. In addition to the deviation of the index from the target it has several other contributors whose purpose is to incentivise the desired properties of the monad bundle. They include penalties for violating the anomaly condition, non-bundleness, equivariance and stability. Trivial line bundles appearing in both B and C can be dropped so we penalise their appearance in order to avoid such trivial configurations. Finally, if the monad map is too simple, the bundle V might split, so the structure group is not $SU(4)$ but a non-trivial sub-group thereof. Since we would like to obtain genuine $SU(4)$ bundles such split monad bundles receive an additional penalty. The detailed form of these various contributions is given in Table 4.1. In terms of this intrinsic state value function, the reward is then defined exactly as for the line bundle environments, that is, by Eq. (1.80). The various parameters in this formula are chosen as

$$p = 1.2 \ , \quad r_{\text{offset}} = -2 \ , \quad r_{\text{step}} = -1 \ , \quad r_{\text{boundary}} = -2 \ , \quad r_{\text{terminal}} = 10 \ .$$

Ideally, the value function should include more and more sophisticated properties of V, such as, for example, the full particle spectrum (rather than just the chiral

property	term in $v(B,C)$	comment
index match	$-\frac{2\lvert\mathrm{ind}(V)-\tau\rvert}{hM^3}$	$\tau=-3\lvert\Gamma\rvert$ is the target index, $\mathrm{ind}(V)$ computed from Eq. (4.6)
anomaly	$\frac{1}{hM^2}\sum_{i=1}^{h}\min\left(c_{2i}(TX)-c_{2i}(V),0\right)$	no penalty if anomaly condition satisfied, $c_{2i}(V)$ computed from Eq. (4.6)
bundleness	$-(d_{\mathrm{deg}}+1)$	d_{deg} = dimension of degeneracy locus as discussed in Sec. 4.1; if the degeneracy locus is empty, d_{deg} is to be taken as -1
split bundle	$-n_{\mathrm{split}}$	n_{split} = number of splits in V
equivariance	$-\sum_{U\subset B,C}\mathrm{mod}(\mathrm{ind}(U),\lvert\Gamma\rvert)$	U runs over all line bundles in B,C or blocks of same line bundles, as discussed in Sec. 4.1
trivial bundle	$-n_{\mathrm{trivial}}$	n_{trivial} = number of trivial line bundles
stability V	$-\frac{\max(0,h^0(X,B)-h^0(X,C))}{hM^3}$	tests Hoppe's criterion for V, cohomologies from formulae in Sec. 1.4.2
stability V^*	$-\frac{\max(0,h^0(X,B^*)-h^0(X,C^*))}{hM^3}$	tests Hoppe's criterion for V^*, cohomologies from formulae in Sec. 1.4.2

Table 4.1: Contributions to the intrinsic value for the monad environment. The intrinsic value $v(B,C)$ is the sum of all eight terms and $M=\max(b_{\max},c_{\max})$.

asymmetry) and a detailed stability check. However, including these properties would require carrying out cohomology calculations during training. Given that these calculations are currently based on commutative algebra methods this is not feasible, as it would lead to an unpredictable slow-down of the training process. Progress in this direction can be made if analytical formulae for monad cohomology can be derived, in analogy with the formulae for line bundle cohomology discussed in Section 1.4.2. This may well be possible but, at present, such formulae are not known. Hence, for the time being, we limit ourselves to training on the properties listed in Table 4.1. More detailed calculations of the spectrum and stability checks will only be carried out after training and for the terminal states found by the RL system.

The monad environment has considerable degeneracy. Permutations of the line bundles in B and C of course do not change the monad bundle, so we have a permutation symmetry $S_{r_B}\times S_{r_C}$. Also, depending on the underlying manifold, there can be an additional discrete symmetry $H \subset S_h$ which permutes the rows of (B, C), so the total symmetry group is

$$H \times S_{r_B} \times S_{r_C} \tag{4.11}$$

In fact, for the bicubic CY, we have $H = S_2$ and for the triple trilinear CY, $H = S_3$.

Finally, we should add a comment on how we sample the initial states for episodes. Naively, one might choose a flat distribution on the environment to choose these states. However, model building experience shows that successful models tend to have entries which are relatively small (typically 0, ±1 or ±2). Since we are trying to match relatively small numbers, such as the number of families, this is perhaps not surprising. For this reason it is helpful to choose a distribution which favours small entries k in initial states (B, C) and we have opted for

$$P(k) \sim \frac{1}{1 + |k|^2} \; .$$

This environment is realised as a MATHEMATICA package which is coupled to either the REINFORCE or the actor-critic package. The policy network (as well as the value network in the actor-critic case) are fully connected neural networks of the type shown in Fig. 1.8. The input dimension is $d_0 = h(r_B + r_C - 1)$, the number of independent entries in a state (B, C) subject to the contraint $c_1(B) = c_1(C)$, and the output dimension is $d_1 = 2h(r_B + r_C)$, the size of the action space. For the network width we have chosen $d = 64$. We use the ADAM optimiser with a learning rate of 1/3500. The maximal episode length is $t_{\max} = 32$, batch sizes are 64 and the discount factor is set to $\gamma = 0.98$ for REINFORCE and to $\gamma = 0.6$ for actor-critic.

4.2.1 Results on the bicubic with $(r_B, r_C) = (6, 2)$

Our first example is for the bicubic CY with ranks $(r_B, r_C) = (6, 2)$ and entries in the range

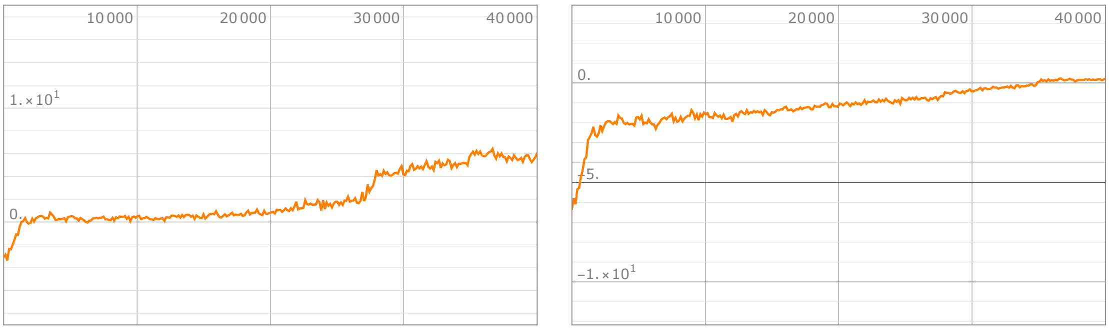
$$-3 = b_{\min} \leq b_i^k \leq b_{\max} = 5 \; , \qquad 0 = c_{\min} \leq c_a^k \leq c_{\max} = 5 \; , \tag{4.12}$$

which amounts to an environment with about 10^{14} states. This is already quite sizeable and, given that monad bundles do not allow for simplifications such as checks carried out for each line bundle, a systematic scan of this environment is not feasible. Moreover, terminal states are very rare; for instance, by randomly sampling 10^9 models, no terminal state are typically found.

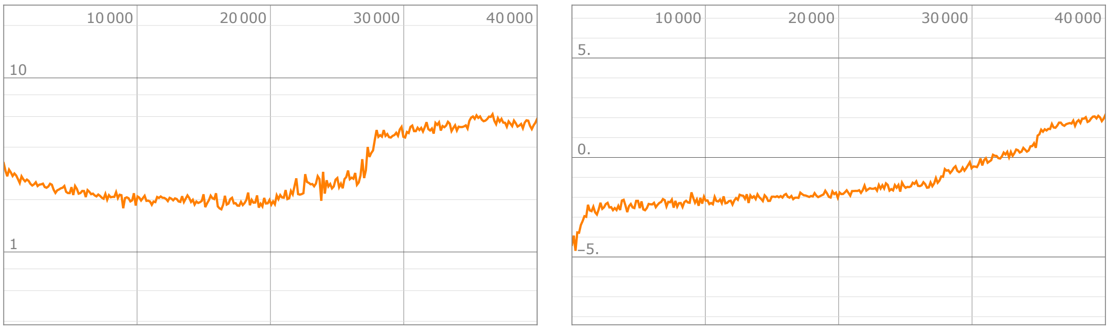
In the following we present the results we have obtained with the actor-critic algorithm but results for REINFORCE are, in fact, quite similar. The measurements taken during 40000 rounds of training (about an hour on a single CPU) are shown in Fig. 4.1. The most impressive indicator is Fig. 4.1(f) which shows the fraction of terminal episodes. At about 30000 rounds this fraction quickly rises to a value close to 1, showing that every starting state is guided to a terminal state and it turns out that this happens within 19 steps on average. We note that this is achieved by sampling only a tiny fraction of about $\sim 10^{-8}$ of the environment's states. During training a few hundred terminal states are found (see Fig. 4.1(e)). After removing redundancies due to the symmetry (4.11) this number reduces to 59 terminal states.

The trained network can be used to search for further terminal states. Running about 1000 episodes from random starting points and guided by the trained network two more terminal states are found (after removing redundancies) which brings the total to 61. The episodes generated by the trained network have, in general, a standard order in which the various requirements for a terminal state are satisfied. Specifically the network tries to satisfy the anomaly cancelation and the rudimentary stability conditions first. The network consistently adjust for the equivariance and the index conditions last (though sometimes joint last). This may be considered an indication to how a model builder may try to construct such a model with pen and paper. This is demonstrated in Fig. 4.2, by considering 1000 terminal states generated using the trained network.

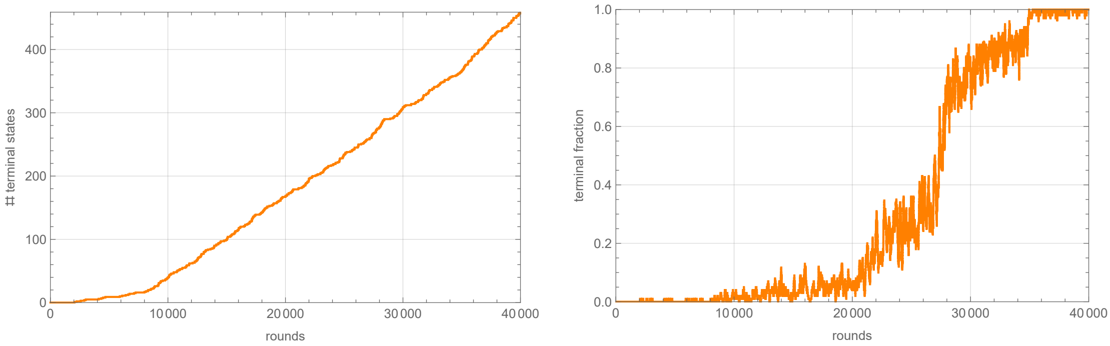
The 61 monad bundle models found in this RL scan have been further checked. We found that 47 of these have $h^1(X,V)=27$ families before taking the $\mathbb{Z}_3\times\mathbb{Z}_3$-quotient and no anti-families, that is, $h^2(X,V)=0$. Further stability checks, testing the injection of line bundles with entries in the range $-6,\dots,6$ into V and V^*, shows that many of these models are unstable. However, 18 models survive these fairly extensive checks. These models include the semi-positive monad with



(a) Loss vs batch number. **(b)** Policy loss vs batch number.



(c) Value loss vs batch number. **(d)** TD return vs batch number.



(e) Number of terminal states vs episode number. **(f)** Terminal fraction vs episode number.

Figure 4.1: Training metrics for the bicubic monad environment with $(r_B, r_C) = (6, 2)$.

$B = \mathcal{O}_X(0,1)^{\bigoplus 3} \oplus \mathcal{O}_X(1,0)^{\bigoplus 3}$ and $C = \mathcal{O}_X(1,1) \oplus \mathcal{O}_X(2,2)$ found in Ref. [143]. In fact, the latter model is the only semi-positive monad found by the network, so it is likely the only semi-positive standard model with $(r_B, r_C) = (6, 2)$ on the bicubic.

4.2.2 Results on the bicubic with $(r_B, r_C) = (7, 3)$

Raising our ambition moderately higher, we consider next a monad environment on the bicubic with larger ranks, $(r_B, r_C) = (7, 3)$, but with a range of integer entries as in Eq. (4.12), leading to about 10^{16} states. Both REINFORCE and actor-critic

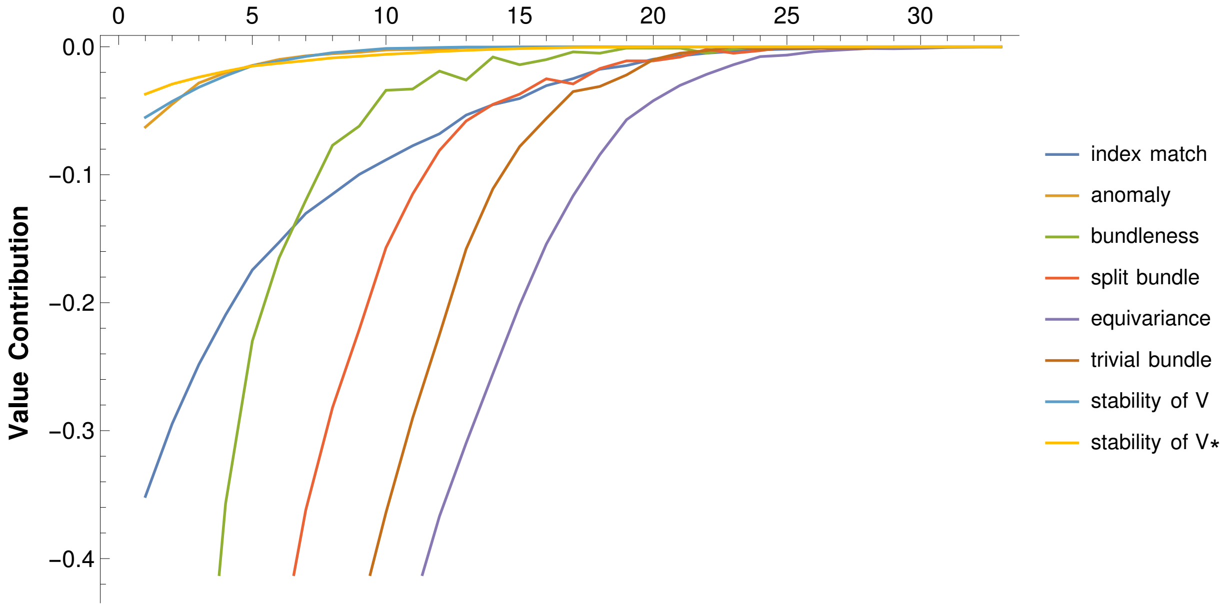
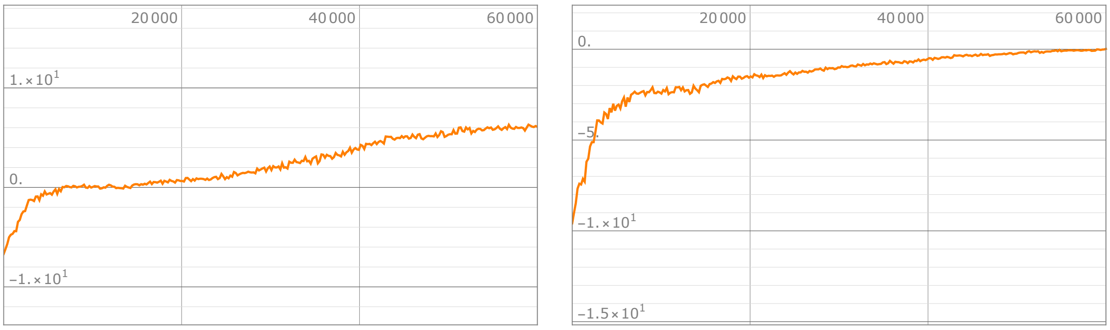


Figure 4.2: The different contribtutions to the intrinsic value for $(r_b, r_c) = (6, 2)$ bicubic models. This data is averaged over 1000 termianl states using the trained network.

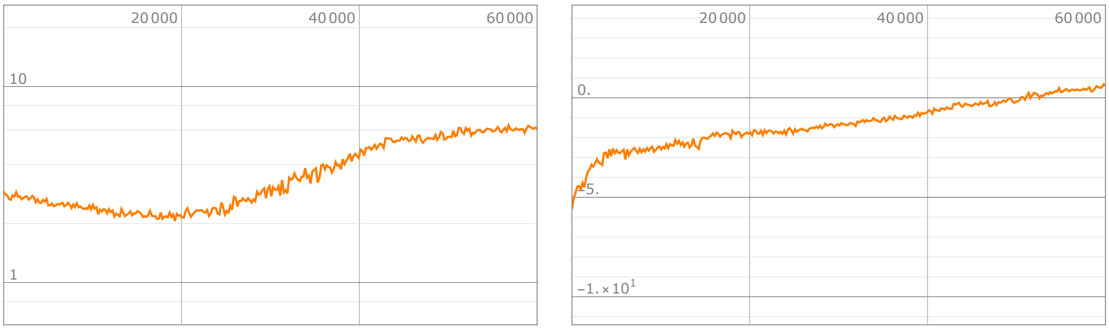
algorithms lead to similar results and, for definiteness, we present the latter.

The network has been trained for 60000 rounds, sampling only a tiny fraction of about 10^{-10} of the environment's states. Training takes about two hours on a single CPU and training measurements are shown in Fig. 4.3. As in the previous case, we observe a transition to a policy with a terminal fraction close to 1 during the training process (Fig. 4.3(f)), with the average episode length decreasing to about 20.5. Fig. 4.3 shows that about 3000 terminal states are found during training. After removing redundancies, this number reduces to 141, of which 129 have 27 families and no anti-families. It is worth mentioning that 101 of these models have a common line bundle in B and C and are, hence, equivalent to models with $(r_B, r_C) = (6, 2)$. A further 5 models have two common line bundles in B and C are are equivalent to models with $(r_B, r_C) = (5, 1)$.

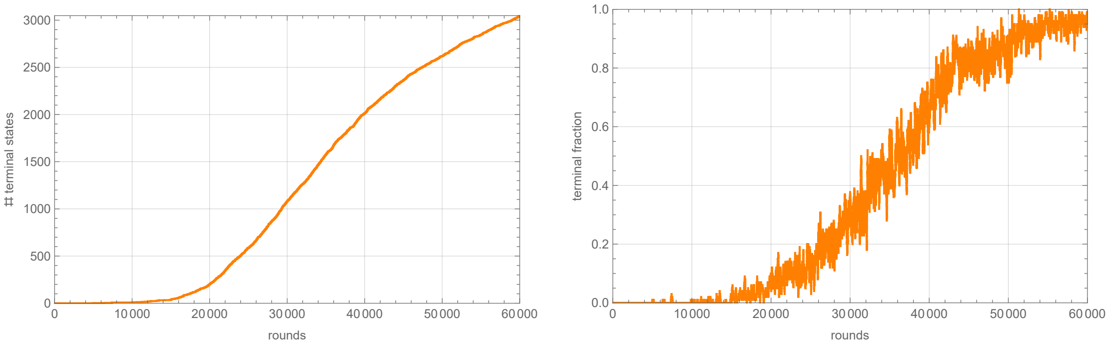
As before, the trained network can be used to find more terminal states. After running 1000 episodes guided by the trained network, removing redundancies and checking the cohomology for the terminal states, one more model is found, bringing the total number to 130. A stability check, looking at injection of line bundles



(a) Loss vs batch number. **(b)** Policy loss vs batch number.



(c) Value loss vs batch number. **(d)** TD return vs batch number.



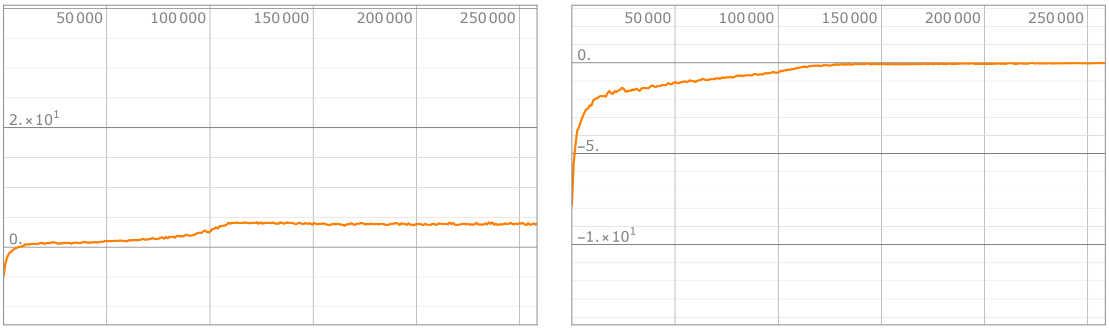
(e) Number of terminal states vs episode number. **(f)** Terminal fraction vs episode number.

Figure 4.3: Training metrics for the bicubic monad environment with $(r_B, r_C) = (7, 3)$.

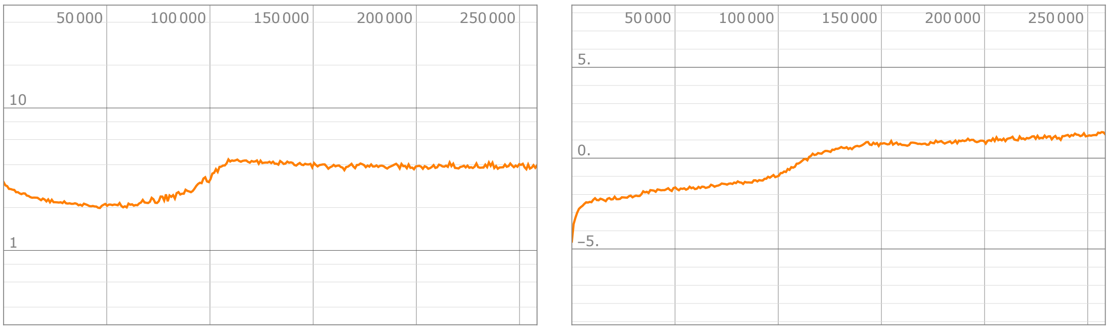
with entries between ±5 into V and $V^\star$, rejects 61 of these models as unstable, leaving 69 models which pass.

4.2.3 Results on the triple trilinear manifold with $(r_B, r_C) = (6, 2)$

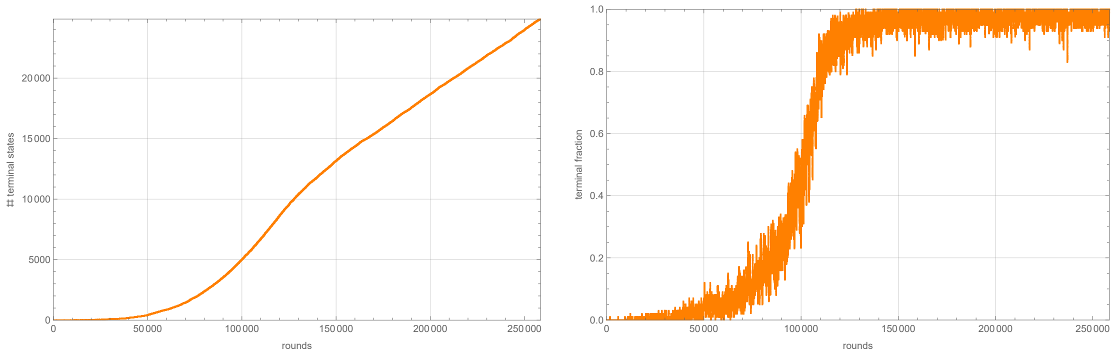
Finally, we consider monads on the triple tri-linear CY X_{7669} (see eqn (4.8)) with line bundle sum ranks $(r_B, r_C) = (6, 2)$ and entries bounded as before, see Eq. (4.12). Since the Picard number is $h = 3$, the environment is much larger than for



(a) Loss vs batch number. **(b)** Policy loss vs batch number.



(c) Value loss vs batch number. **(d)** TD return vs batch number.



(e) Number of terminal states vs episode number. **(f)** Terminal fraction vs episode number.

Figure 4.4: Training metrics for a monad environment on the triple tri-linear CY with $(r_B, r_C) = (6, 2)$.

the previous example and contains about 10^{20} states. As in the previous case, the manifold admits a freely acting $\mathbb{Z}_3 \times \mathbb{Z}_3$ symmetry, hence we are looking for models with 27 families.

For this manifold the actor-critic algorithm proves more efficient. $258,500$ rounds of training take about a day on a single CPU and training measurements are shown in Fig. 4.4. Only a tiny fraction of about 10^{-13} of the environment's states has been sampled during training.

Similar to the previous example, Fig. 4.4(f) shows a dramatic increase of the terminal fraction to a value close to 1. At the same time, the average episode length decreases to a value of about 22.5. After removing redundancies, 12819 terminal states are found during training. Out of these, 7638 models have 27 families, no anti-families and have vanishing first and third cohomologies as required by stability.

As before, the trained network can be used to find further terminal states. After running 1000 episodes with the trained network, removing redundancies and checking for anti-families, 174 models are found, taking the total to 7812. This data set has been included as ancillary material in the arXiv submission of [22].

Performing stability checks for these models is computationally intense, due to the higher Picard number. For this reason, we have refrained from performing stability checks for the entire list of 7812 models.

4.3 Breeding Heterotic Monad Bundls

We now turn to GA as an alternative search method. The translation of the above RL environemnt is very straight forward. To realise the GA we choose the range of entries such that $b_{\max} - b_{\min} = 2^{n_B} - 1$ and $c_{\max} - c_{\min} = 2^{n_C} - 1$, representing every integer in B and C in a binary encoding, with n_B and, respectively, n_C bits, resulting in a bit sequence of total length $\ell = h(r_B n_B + (r_C - 1)n_C)$. The typical size of a population consists of $N_{\text{pop}} \simeq 100 - 300$ states, each represented by such a binary encoding of the matrix (B, C). The population is then evolved by crossing and mutation, as discussed in Section 1.5.1.

With the fitness function specified by the intrinsic value $v(B, C)$, as given in Sec. 4.2, we found that rank-weighting selection, elitism and a simple one-point cross breeding work well.

The GA was implemented in two ways, as a MATHEMATICA and as a Python package, in both cases being coupled to the MATHEMATICA package realising the environment. The two realisations lead to similar results.

4.3.1 A typical GA run on the bicubic

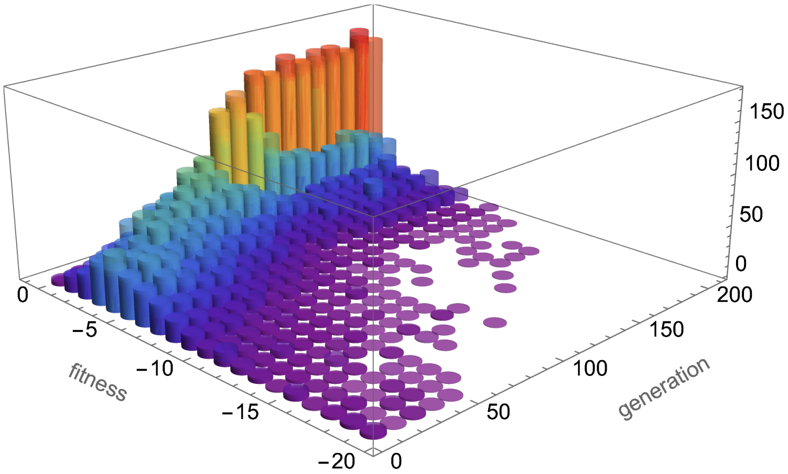
We first consider monad bundles on the bicubic manifold with $\mathrm{rk}(B) = 6$ and $\mathrm{rk}(C) = 2$. We choose $b_{\mathrm{min}} = -3$, $c_{\mathrm{min}} = 0$ and $b_{\mathrm{max}} = 4$, $c_{\mathrm{max}} = 7$, such that each integer entry in the matrix (B, C) can be represented by a 3-bit binary sequence counting up from the fiducial points, b_{min} and c_{min}. For instance, an entry of -3 in B will be represented by the sequence $(0, 0, 0)$, while an entry of 4 in B corresponds to $(1, 1, 1)$. Similarly, an entry of 0 in C is represented as $(0, 0, 0)$, while an entry of 7 in C corresponds to $(1, 1, 1)$.

Working with integers that can be encoded on 3 bits is convenient for the purpose of comparing the GA and RL techniques. However, it is possible and in fact easy to further increase the range of line bundle integers, as the corresponding augmentation in the computational time is not significant.

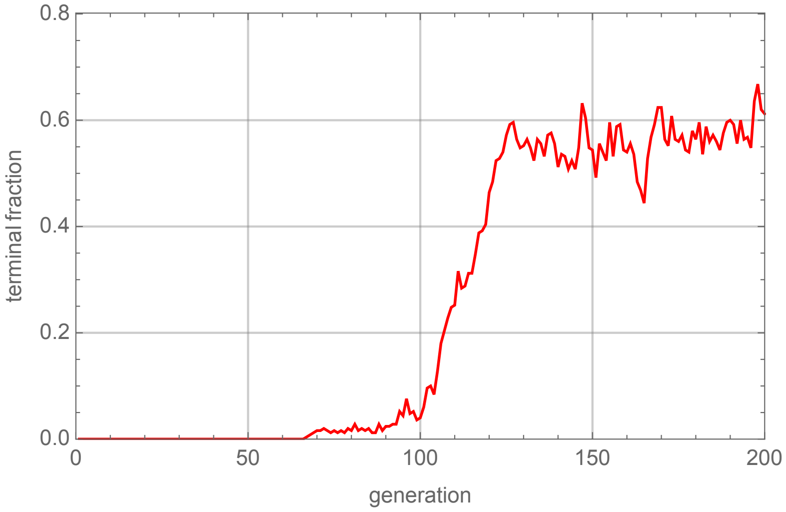
As mentioned above, the environment has a large degeneracy. Equivalent models arise from permuting the two $\mathbb{P}^2$-factors in the bicubic embedding, as well as from permuting the line bundles in B and C. This amounts to a group of order $2! \cdot 6! \cdot 2! = 2800$. Of course, this does not necessarily imply that every perfect model can be found in 2800 different sites, since some permutations may leave particular matrices (B, C) unchanged, but it nevertheless gives some idea of the degree of degeneracy present in the environment.

It is useful to illustrate the performance of the GA graphically. We set the size of the population to $N_{\mathrm{pop}} = 250$ randomly initialised states and let the GA run for 200 generations with a mutation rate of 0.004 and the parameter α from (1.66) is set to $\alpha = 3$. The time taken by such a run is relatively small (<100 seconds on a laptop). Figure 4.5 shows a typical evolution of the population. The histogram on the left shows how the population progresses towards greater fitness (intrinsic value). The plot on the right shows the fraction of perfect models in the population, that is, the fraction of models with a value $v(B, C) = 0$. By generation 150 more than half of the population corresponds to perfect models. For this run, a total of 200×250 states have been visited, many of them multiple times. Of these

12, 665 correspond to perfect models with 48 being distinct. After eliminating redundancies, 18 non-equivalent models remain.



(a) Fitness histogram: number of individuals as a function of generation and fitness.



(b) Fraction of perfect models vs generation.

Figure 4.5: Performance measures for a typical GA initialisation on the bicubic.

The fraction of perfect models can be easily increased by lowering the mutation rate. However, as one might anticipate, this also induces a degree of stagnation, with a tendency for the same states to emerge repeatedly.

A few remarks are in order. Firstly, to appreciate the performance of the algorithm, note that the size of the search space is

$$8^{14} \simeq 4.4 \times 10^{12} \, .$$

By comparison, the number of states visited in the above run, namely $50,000$, represents only a tiny fraction of the space. However, the GA was capable of finding 48 perfect states, while a random search over millions of states would typically lead to no perfect states at all.

Secondly, as already noted, the GA has a tendency of visiting the same states multiple times. It is interesting to plot the total number of perfect states found after n generations as a function of n. For our illustrative run such a plot is shown in Figure 4.6, which suggests that there is no additional benefit in letting the population evolve beyond a certain generation ($n \simeq 150$).

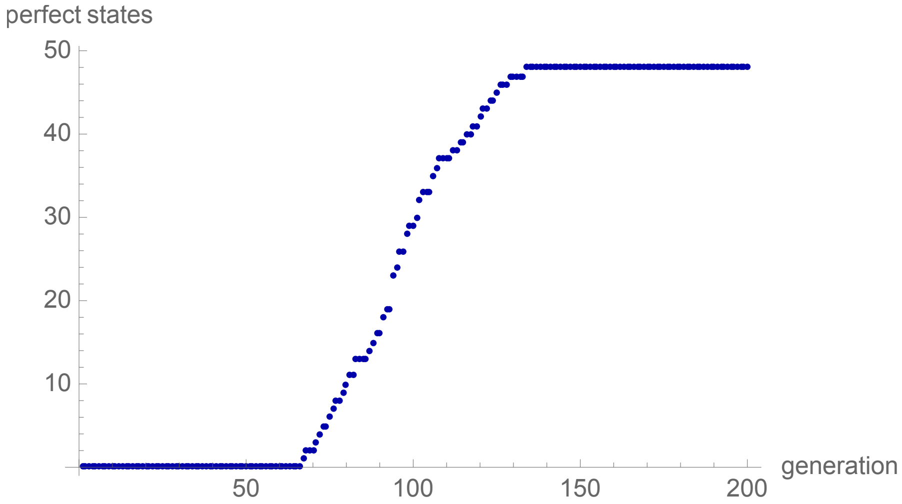


Figure 4.6: Saturation of the number of perfect states found in a typical GA run on the bicubic.

Finally, it is useful to compute the degeneracy of the 18 states that remain after removing redundancies. By performing all the allowed permutations a number of $19,080$ states are obtained. This should be compared with the product $18 \times 2800 = 50,400$ which turns out to be an overestimate by more than a factor of 2. Moreover, what this computation indicates is that a single run of the GA is not enough if the aim was to find all the perfect states available in the environment and that $1,000$ further runs, which would take about 1 day on a single machine, would be just about enough to find the other redundant representations of the 18 states found in the first run. Of course, many more new states, not related to the 18 by permutations, would be expected to arise in such a search, which implies that a comprehensive search would require several, possibly tens of thousands of GA runs. With $10,000$ runs this would amount to exploring $\sim 0.01\%$ of the environment.

4.3.2 A quasi-comprehensive GA search

The illustrative results presented above show that GAs can be efficiently used to identify monad bundle models on the bicubic manifold that satisfy the phenomenological requirements discussed in Sec. 1.3. A natural question to ask is whether the GA can find *all* (or almost all) perfect models available within the environment. Answering this question is difficult, mainly because the environment is too large to be subjected to a systematic scan which would ensure that no perfect state was missed. The computational time needed for a systematic scan within the environment bounds can be easily estimated to about 200 core years, which although attainable, is by no means something that one would easily invest in. A partial answer to the question can be achieved by making use of the redundancy present in the environment. The procedure is as follows. We perform 10,000 runs keeping the same settings as above, namely $N_{\text{pop}} = 250$, the length of a run equal to 200 generations, a mutation rate of 0.004 and $\alpha = 3$. The required computation time for this is about 10 core days[1]. We extract the number of perfect states found after n runs and then ask how many of these are left after removing redundancies. Now, if the GA is capable of finding almost all the states, then the expectation would be that the redundancy reduced number of perfect states after n runs would saturate as a function of the total number of perfect states found after n runs. In other words the GA is only finding versions of models that it has found before. The plot in Figure 4.7 confirms this expectation. The search lead to a number of $96,705$ perfect states, out of which 639 are inequivalent (not related by a permutation). Performing all possible permutations on the 639 states leads to $481,680$ perfect states that must be present in the environment. If allowed more time, the GA would of course be expected to find all of these. However, the saturation plot in Figure 4.7 as well as the plots in Figure 4.8 suggest that there are not many more inequivalent states to be found.

---

[1] The scan was carried out on the Hydra Computer Cluster in the Dept. of Physics at the University of Oxford.

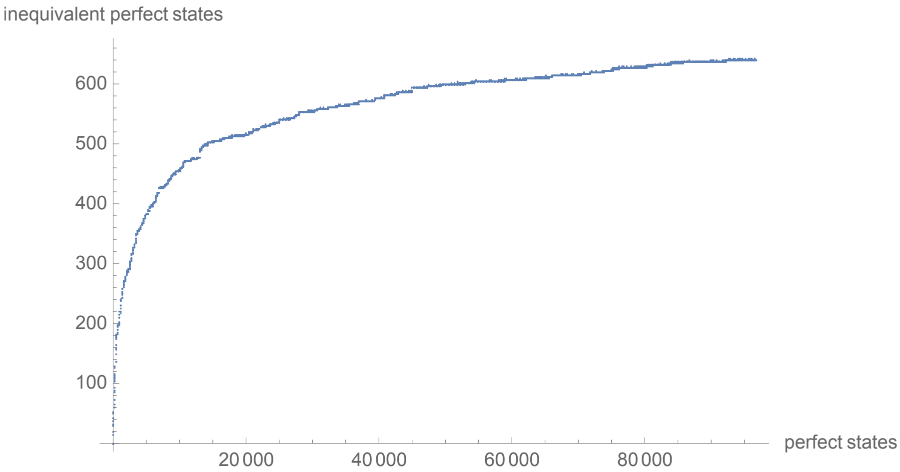


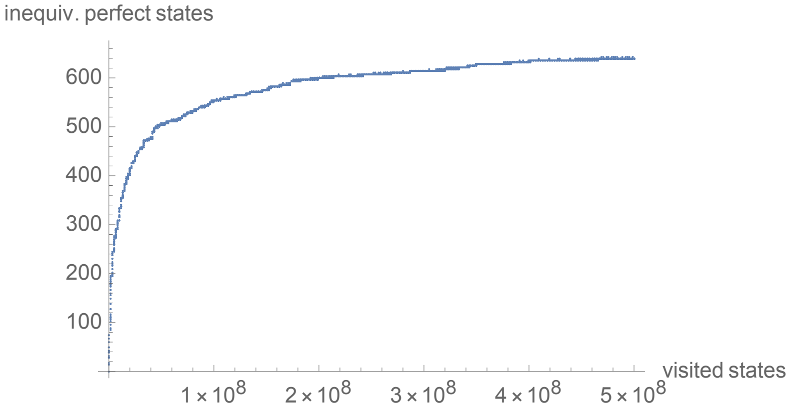
Figure 4.7: GA scan results. Saturation of the redundancy reduced number of perfect states.

We take this as an indication that the search has already successfully found most of the inequivalent perfect states.

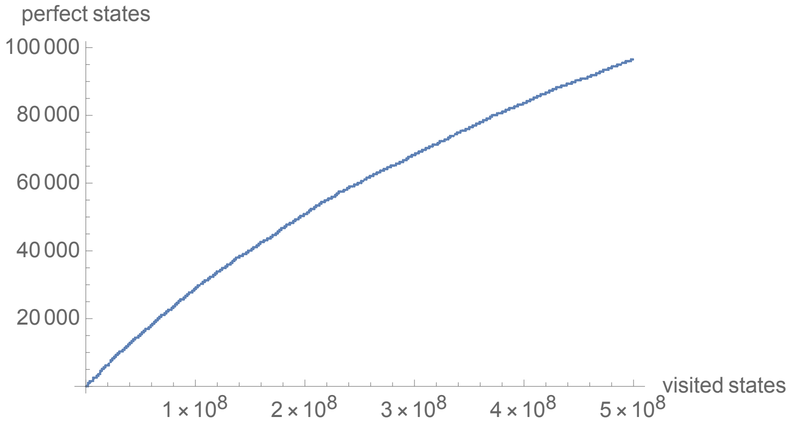
4.3.3 Enlarging the search space for the bicubic

So far the minimal and the maximal values of the entries in B and C were fixed such that all integers can be encoded on 3 bits. Upgrading to integers that can be encoded on 4 bits leaves the computational time required for a GA run virtually unchanged. Thus enlarging the range of line bundle integers in B and C would not pose a problem for the performance of the GA implementation used here. But is there a good reason for enlarging the environment?

To answer this question we need to see if the perfect models found are well within the boundary of the environment or if a significant fraction of them lies close to the boundary. The bar charts in Figure 4.9 show that in fact most of the models lie on the boundary (for instance, 451 out of the 639 inequivalent models include the integer 4 in B). This suggests the need for increasing further the entries in B and C. However, before embarking on such an enterprise, it is worth looking deeper into the quality of the monad bundles found. The requirement of stability imposes that $h^0(V) = h^3(V) = 0$. Moreover, in order to avoid the presence of $\overline{\mathbf{16}}$-multiplets, the bundle has to satisfy $h^2(V) = 0$. These cohomological constraints are too



(a) Number of inequivalent perfect states found as a function of the number of states visited.



(b) Number of perfect states found as a function of the number of states visited.

Figure 4.8: GA search results on the bicubic. The number of inequivalent perfect states saturates, while the total number of perfect states found is still far from saturation after 10,000 GA runs.

computationally intensive to have included them in the search, however they can be carried out on the set of 639 inequivalent models found on the bicubic manifold.

Moreover, many of the models found contain a repeated line bundle in B and C. While this is not a problem in itself, such models are equivalent to monad bundles with $\text{rk}(B) = 5$ and $\text{rk}(C) = 1$. We eliminate these from the discussion, which leaves 214 models of genuine $\text{rk}(B) = 6$ and $\text{rk}(C) = 2$ type. Out of these, 67 have the right cohomologial properties mentioned above. Further stability checks reduce this number to 29 models. The distribution of entries in B for the remaining 29 models is shown in blue in Figure 4.10 and indicates that most of the models have all the line bundle entries in the range $\{-2, -1, 0, 1, 2\}$. In view of this, the

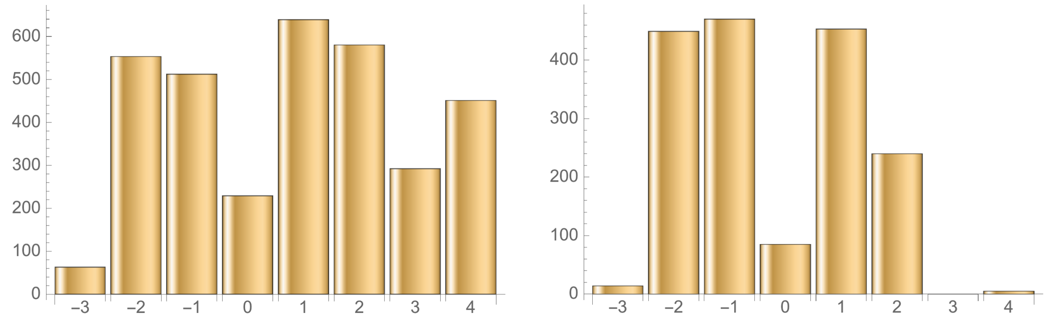


Figure 4.9: Bar charts showing the number of inequivalent perfect states (B, C) that have a certain integer present in B (left chart) and in C (right chart).

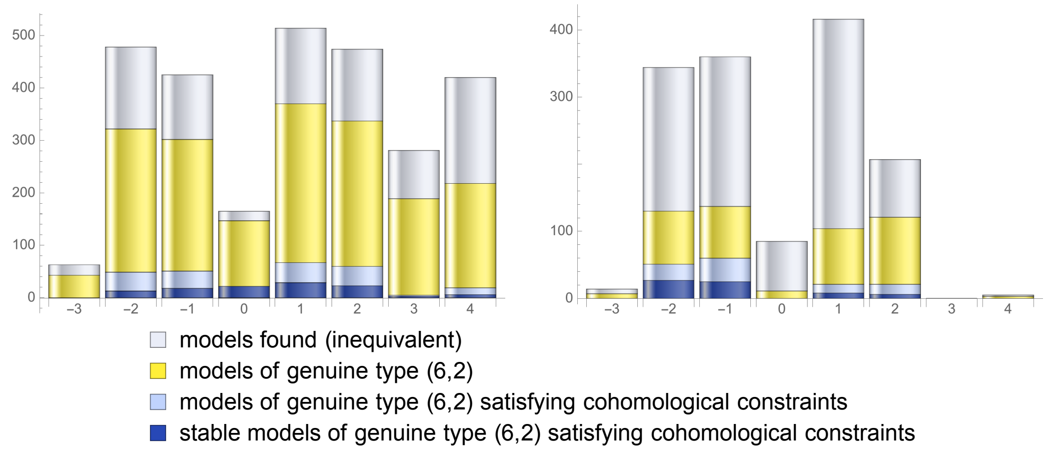


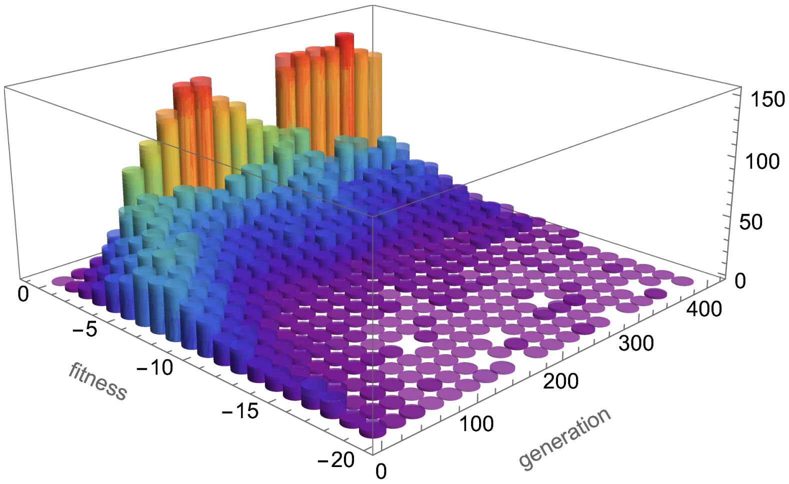
Figure 4.10: Distribution of models across the range of integers in B (top chart) and C (bottom chart). The number of models is significantly reduced when additional requirements are imposed. Very few models are left on the boundary of the environment.

expectation is that very few, if any, additional models satisfying the additional constraints would be found by further enlarging the size of the environment.

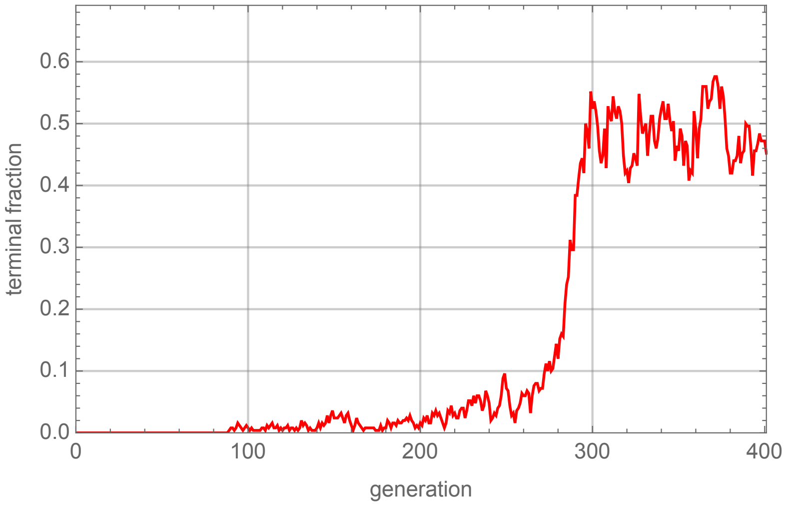
4.3.4 A typical GA run on the triple trilinear

As in the previous section we consider monad bundles with $\mathrm{rk}(B) = 6$ and $\mathrm{rk}(C) = 2$ and set $b_{\min} = -3$, $c_{\min} = 0$ and $b_{\max} = 4$, $c_{\max} = 7$, such that each integer entry in the matrix (B, C) can be represented by a 3-bit binary sequence. For the triple trilinear manifold the degeneracy of the environment is even larger than in the case of the bicubic, due to the six possible permutations of the three $\mathbb{P}^2$-factors in the ambient space. The corresponding group has order 8640.

As in the previous case, we illustrate the performance of the GA graphically. We keep the size of the population to $N_{\text{pop}} = 250$ randomly initialised states. The GA needs to run a little longer until a significant fraction of the population corresponds to perfect states. Evolutionary episodes of 400 generations with a mutation rate of 0.004 and $\alpha = 4$ seem to work well. The time taken by such a run is relatively small ($\sim$450 seconds on a single machine).



(a) Fitness histogram. Height: number of individuals. Front axis: fitness (intrinsic value). Depth: generation number.



(b) Fraction of models that achieved ‘perfection’ vs generation.

Figure 4.11: Performance measures for a typical GA initialisation on the triple trilinear.

Figure 4.11 shows a typical evolution of a population of models on the triple trilinear manifold. By generation 300 about half of the population corresponds to perfect models. For this run, a total of 400×250 states have been visited, many of them multiple times. Of these $15,377$ correspond to perfect models, with 43 being distinct.

After eliminating redundancies, 41 non-equivalent models remain.

The size of the search space is

$$8^{21} \simeq 10^{19} \, .$$

which is orders of magnitude larger than the bi-cubic space. The number of states visited during the above run, namely $100,000$ states (not necessarily distinct), represents a minuscule fraction of the space. Nevertheless, it is remarkable that 43 perfect states were found, and these perfect states lead to 41 inequivalent models. The total number of perfect states found after n generations as a function of n is plotted in Figure 4.12, which suggests that the number of generations, 400, is suitably chosen given the other hyper-parameters. This continued efficacy despite the much larger search space suggests that the difficulty of finding a solution is increasing only polynomially for the GA as one might have hoped.

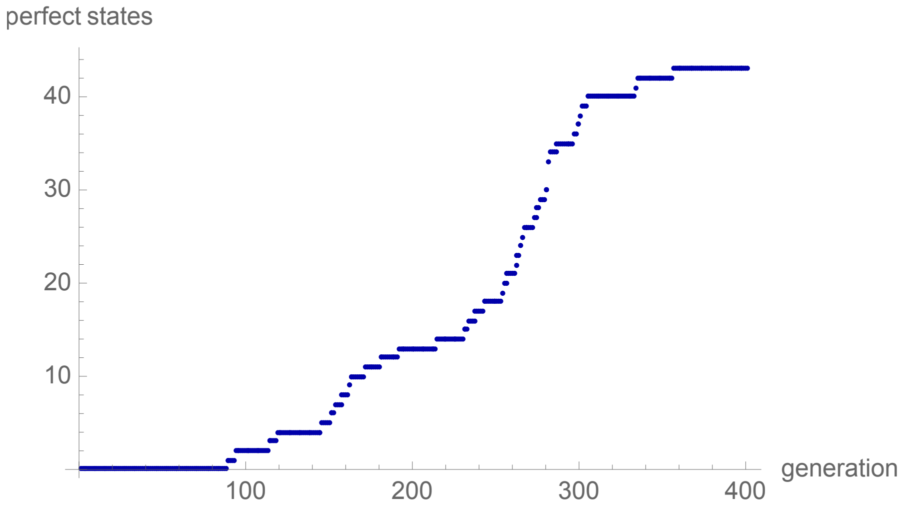


Figure 4.12: Saturation of the number of perfect states found in a typical GA run on the triple trilinear CY manifold.

4.3.5 GA search on the triple trilinear

The environment being much larger than in the case of the bicubic manifold, we have not attempted to reach the same degree of comprehensiveness as before.

After running for 15 core days, the GA found $\sim 100,000$ perfect states, three quarters of which are inequivalent. Figure 4.13 shows that the number of inequivalent

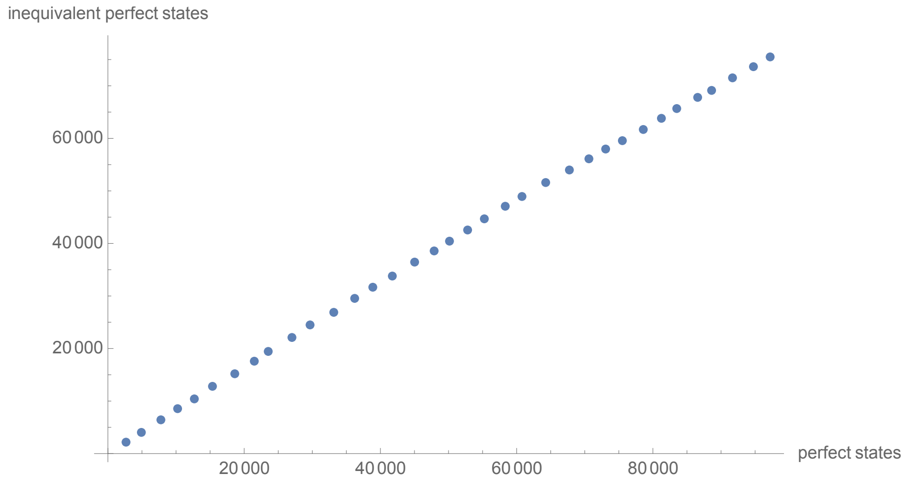


Figure 4.13: GA search results on the triple trilinear manifold. The creation of each point has taken 10 core hours and corresponds to 10^8 visited states. The number of inequivalent perfect states has not saturated after $35 \cdot 10^8$ visited states.

states is far from saturation at the end of the run, suggesting that many more other inequivalent perfect states are still to be found within the environment.

4.4 GA scan versus RL scan

Although the RL and GA methods rely on the same intrinsic value function, their underlying philosophies are very different, which makes for an interesting comparison. In particular we focus on the (6,2) monads on the bicubic, and take the previously trained network. The RL agent is capable of finding terminal states from virtually any random starting point within episodes of average length ∼30.

In order to get a preliminary assessment of the performance of the trained network we let it run for $10,000$ episodes which takes about 16 minutes on a single machine. The run produced 2460 terminal states, out of which only 72 are distinct. The repetitions are not evenly distributed over the 72 distinct states. In fact, one of these states accounts for almost half of the terminal states found. Incidentally, this particular state corresponds to a monad bundle that can be reduced to a $\text{rk}(B) = 5$ and $\text{rk}(C) = 1$ case. The genuine $\text{rk}(B) = 6$ and $\text{rk}(C) = 2$ are found with much smaller frequencies, most of them appearing only once in the dataset. Thus, although the network is capable of finding terminal states, a small number of these states have very large basins of attraction and are found much more often than the others. This is a source of inefficiency for the RL implementation.

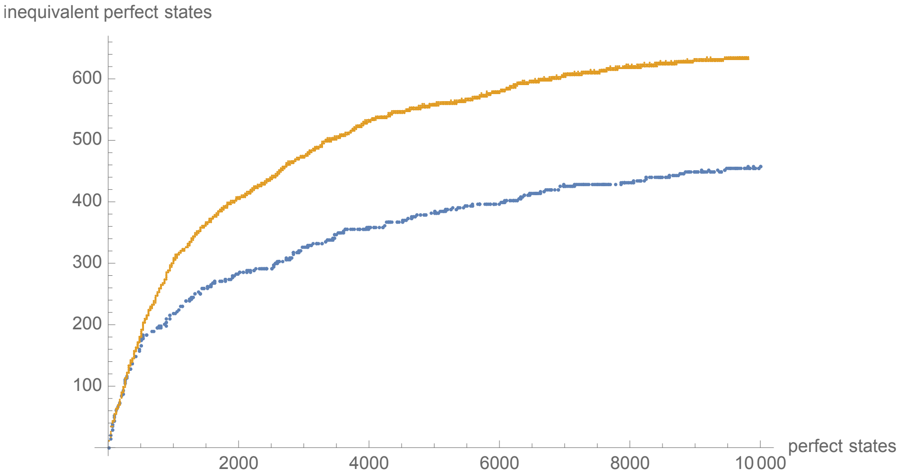
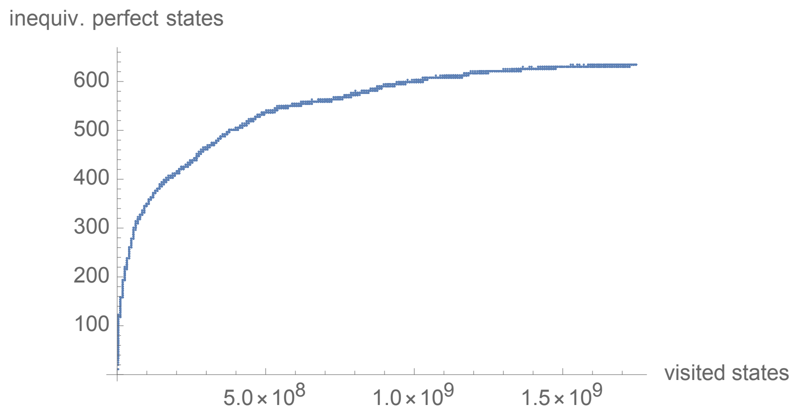


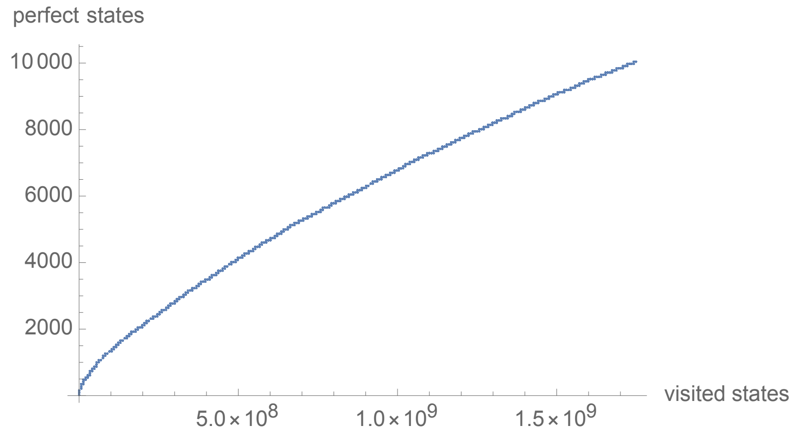
Figure 4.14: Saturation of the redundancy reduced number of perfect states. Orange curve: RL search. Blue curve: GA search. The RL search took 35 core days, while the part of the GA search shown here took only 1 core day.

We let the trained network search for terminal states from $1.7 \cdot 10^9$ random initial states. This search, which took 35 core days produced ∼10, 000 perfect states, out of which 643 were inequivalent. By comparison, the computational time for the GA scan was 10 core days and lead to ∼100, 000 perfect states, out of which 639 were inequivalent. The saturation curve of the number of inequivalent models as a function of the total number of models found at various stages of the RL search is shown in Figure 4.14 together with the corresponding curve for the first 10, 000 perfect states found during the GA search. Figure 4.15 shows the progression of the number of (inequivalent) models found with the number of states visited.

A number of remarks can now be made. Firstly, the GA search is more efficient at finding perfect states than the RL search. Roughly the same number of inequivalent perfect states have been found in the two searches, though the RL search took 3.5 times longer than the GA. Secondly, the GA is more prolific, by more than an order of magnitude compared to the RL search, at finding redundant perfect states. This is very likely a by-product of crossover, which is prone to producing redundant states by permuting various parts of the genome. Note that various techniques such as crowding penalties might be considered to remove this inefficiency.



(a) Number of inequivalent perfect states found as a function of the number of states visited.



(b) Number of perfect states found as a function of the number of states visited.

Figure 4.15: RL search results on the bicubic. The number of inequivalent perfect states saturates, while the total number of perfect states found is still far from saturation after $1.7 \cdot 10^9$ RL episodes.

Finally, there is a substantial overlap between the inequivalent models found in the two searches, with only ∼50 models in each complement. This brings the total number of inequivalent models to 689. The fact that two such different methods lead to virtually the same dataset of models is another even stronger indication that it has been possible to achieve a high degree of comprehensiveness by scanning only a tiny fraction of the environment.

While both methods produce nearly identical data sets of perfect models after sufficient running time their trajectories during evolution/training appear to be quite different. Using nonlinear visualisation techniques such as Sammon maps we find that the set of perfect models found at an early stage is quite different for the two methods. In particular, as illustrated by Fig. 4.14, GAs tend to produce a significantly larger degree of redundancy than RL.

4.5 Summary

In this chapter we have explored the efficacy of genetic algorithms (GAs) and reinforcement learning (RL) as search methods for probing the landscape of four-dimensional $\mathcal{N} = 1$ heterotic compactifications. Both methods turn out to be very successful for engineering the compactification data and for finding perfect models, that is, models which resemble the Standard Model. The spaces of models explored here were extremely large ($\sim 10^{12}$ in the case of the bicubic manifold and $\sim 10^{19}$ for the triple trilinear manifold). In the case of the bicubic manifold a systematic scan, although feasible, would be extremely time consuming, while for the triple trilinear manifold, a systematic scan is practically unfeasible.

GA was able to find tens of models within the first few minutes of running on a single machine. The method is not only efficient, but also highly parallelisable. After 10 days of running, the GA accomplished a high degree of comprehensiveness in finding all the perfect models available on the bicubic manifold. Indirect evidence in support of this claim was obtained by exploiting the degeneracy of the environment, as well as by comparison of the results of the GA search with those obtained through reinforcement learning (RL). Although the two methods are very different, they lead to datasets of inequivalent models that have an overlap of over 90%. This suggests that the details of the optimisation process are not really essential once the processes begin to saturate, provided that they share the same incentives.

Our comparative analysis of the two search methods revealed that GA is, overall, more efficient by about an order of magnitude in finding perfect states than RL. However, GA is also more prone to finding models that are related by redundant permutations, a feature which is likely related to the crossover mechanism. In principle techniques such as crowding penalties (with the redundancy incorporated) might improve this aspect. An important factor contributing to the lower output of the RL search was the fact that certain perfect states had very large basins of attraction and consequently were found many times, while other perfect states had small basins of attraction and needed many trial episodes to be found.

The combined RL and GA searches on the bicubic manifold led to some 700 inequivalent perfect models. Only a tenth of these correspond to monad bundles of genuine $(\mathrm{rk}(B), \mathrm{rk}(C)) = (6, 2)$ type with the correct cohomology dimensions. Further stability checks reduce the number of viable models by another factor of 2 – the corresponding monad bundles have been tabulated in Appendix B. Unfortunately these checks had to be performed at the end of the search due to the limitations imposed by cohomology computations. However, it may be the case that a quick method for computing monad bundle cohomology can be developed, for instance in the form of an analytic formula similar to those found for line bundle sums. If possible, this would facilitate the implementation of further checks at the search stage, which would significantly improve the performance of both GAs and RL.

Improving the performance of the search methods is of particular importance for large environments, such as that corresponding to the triple trilinear manifold. In the absence of more stringent notion of perfect model, achieving any degree of comprehensiveness is difficult in such large environments; consequently, partial searches such as the one presented in Section 4.3.5 may miss many relevant models.

Looking more broadly, one is irresistibly drawn to the question of whether the optimisation processes that we have been exploring here might have a bearing on physics in the early Universe. It is know that topological and small-instanton transitions between different vacua exist in string theory, and one may wonder if the dynamics of moving between different vacua may be modelled by RL or similar. In this context, each state would be a distinct vacua of string theory, and the actions would correspond to transitions between them. Of course it is not known what (if any) incentives, leading to an appropriate reward function, might favour the physics that we observe in our world, although there have over the years been varyingly speculative suggestions (and we do not need to add to the list). But if this is the case and if the idea of a multiverse is correct, then the results that we observe with

GA and RL searches do at least support the idea that selection mechanisms can successfully operate within the vast search space, and indeed that they do so in a quite convergent way. This suggests the possibility of alternative pictures, quite distinct from that of a randomly populated multiverse, in which the properties of this world would be more common than one would naively expect.

I'm an astronomer, not a doctor! I mean, I am a doctor, but I'm not that kind of doctor. I have a doctorate, it's not the same thing. You can't help people with a doctorate. You just sit there and you're useless!

— Doctor Doppler *Treasure Planet (2002)*

5

Conclusion

The purpose of this thesis was to explore the use of contemporary computational methods to navigate the string landscape. More specifically, reinforcement learning (RL) and genetic algorithms (GA) were used to navigate different compactifications of the $E_8 \times E_8$ heterotic string, with the aim of finding realisations of the standard model – at least, at the level of the massless particle spectrum.

We were successful in this aim, and applied these methods to monad bundles and sums of line bundles. In both cases many new models were discovered, alongside several that were already present in the literature. It is important to stress that these methods can find applicability outside of string theory, and such an environment was presented in Chapter 2 as a warm-up for the more technical string theory examples. This toy example explored the space of Froggatt-Nielsen models of flavour physics using RL. Significant work has been conducted for these kind of models since the original proposal of Froggatt and Nielsen [129–131], and the RL system was successfully able to reproduce and extend some of these results within a few hours.

We applied GA to sums of line bundles in Chapter 3. These are, in some sense, the simplest non-standard embedding models one can construct from the heterotic string. As a result, systematic scans have been carried out over this environment in the past, leaving us with a useful benchmark with which to compare. The results are summarised in Tab. 3.1, where we see that GA was able to find almost all the models on each considered manifold while only exploring a tiny fraction of the total space – at its most extreme, for one particular case GA was able to find 93% of models while only exploring 10^{-19} of the total environment. We were also able to explore and find models on manifolds with $h_{11} > 6$, which involve sufficiently large environments to defy systematic scans. The main takeaway from this chapter was that the size of the string landscape should no longer be considered a major hurdle in constructing quasi-realistic models of particle physics. Targeted search methods for discrete optimisation, such as the ones presented here, can be performed instead of exhaustive scans.

In Chapter 4, we then applied both GA and RL to monad bundles, which are a subset of non-Abelian bundles. Monad bundles are particularly convenient for our purposes, as they are constructed from line bundles. Although these models have been considered in the literature, only one model was known prior to this work. On the bicubic manifold, GA was able to find new models within a number of minutes, and both RL and GA reach saturation within a few days. The models found by both methods were largely the same, suggesting that almost all models were found. The 28 models found on the bicubic are presented in Appendix B, which includes the one previously known model. The triple tri-linear manifold was also considered, and $\mathcal{O}(200)$ models were found, but for this case neither method was able to reach saturation in a realistic time scale.

Looking forward, it would be interesting to apply these algorithms to other problems in physics – from string theory or otherwise. For example, could these methods be applied to F-theory models of particle physics, for stabilising moduli, or even models of dark matter candidates? At its most ambitious, one can imagine having environments consisting of large families of quantum field theories for physics beyond the standard model, where a diverse set of experimental data is used to define an appropriate value to each model. Of course, such an environment would be an immense undertaking, requiring input from many different fields of science.

Turning to the string models found in this thesis, there is a large amount of further analysis that can be carried out. These models have only been checked at the level of the massless spectrum but, as mentioned in Sec. 1.3.3, one would also like to be able to calculate the Yukawa couplings and stabilise moduli with these models. Both of these tasks will require significant technical progress before they can be fully addressed, but some advances have been made along these lines using other modern computational methods [85, 144]. However, the key take away from this thesis should be that finding the particle spectrum of the standard model (at least in the context of supersymmetric compactifications of the heterotic string) is essentially a solved problem, in part due to the computational methods presented here.

A typical Wednesday morning [145].

Appendices

A
Some Froggatt-Nielsen Models

Contents

A.1 Example models for one $U(1)$ symmetry

In this appendix we list some models with a single $U(1)$ symmetry with a high intrinsic value $\mathcal{V}(\mathcal{Q})$, found during chapter 2.

<table>
<tr><td>charges</td><td>$\mathcal{Q}=\left(\begin{array}{c|ccc|ccc|ccc|c|c} & Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline q & 6 & 4 & 3 & -2 & 2 & 4 & -3 & -1 & -1 & 1 & 1 \end{array} ss \right)$</td></tr>
<tr><td>$\mathcal{O}(1)$</td><td>$(a_{ij})\simeq\begin{pmatrix} -1.975 & 1.284 & -1.219 \\ 1.875 & -1.802 & -0.639 \\ 0.592 & 1.772 & 0.982 \end{pmatrix}, (b_{ij})\simeq\begin{pmatrix} -1.349 & 1.042 & 1.200 \\ 1.632 & 0.830 & -1.758 \\ -1.259 & -1.085 & 1.949 \end{pmatrix}$</td></tr>
<tr><td>VEV, Value</td><td>$v_1\simeq 0.224\,,\qquad \mathcal{V}(\mathcal{Q})\simeq -0.598$</td></tr>
</table>

<table>
<tr><td>charges</td><td>$\mathcal{Q}=\left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 1 & 2 & 0 & -1 & -3 & 1 & -3 & -5 & -4 & 1 & 1 \end{array}\right)$</td></tr>
<tr><td>$\mathcal{O}(1)$</td><td>$(a_{ij})\simeq\begin{pmatrix} -0.601 & 1.996 & 0.537 \\ -0.976 & -1.498 & -1.156 \\ 1.513 & 1.565 & 0.982 \end{pmatrix}, (b_{ij})\simeq\begin{pmatrix} 0.740 & -1.581 & -1.664 \\ -1.199 & -1.383 & 0.542 \\ 0.968 & 0.679 & -1.153 \end{pmatrix}$</td></tr>
</table>

<table>
<tr><td>VEV, value</td><td>$v_1 \simeq 0.158$, $\mathcal{V}(\mathcal{Q}) \simeq -0.621$</td></tr>
</table>

<table>
<tr><td>charges</td><td>$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 7 & 8 & 4 & 2 & -1 & 4 & 0 & -1 & 0 & 0 & 1 \end{array}\right)$</td></tr>
<tr><td>$\mathcal{O}(1)$</td><td>$(a_{ij}) \simeq \begin{pmatrix} -1.027 & -1.234 & 1.914 \\ 1.027 & -0.525 & -0.921 \\ -1.995 & 1.317 & 0.982 \end{pmatrix}, (b_{ij}) \simeq \begin{pmatrix} -1.223 & -1.322 & 1.121 \\ 1.004 & 0.945 & 1.851 \\ -1.930 & -1.071 & -1.720 \end{pmatrix}$</td></tr>
<tr><td>VEV, value</td><td>$v_1 \simeq 0.316$, $\mathcal{V}(\mathcal{Q}) \simeq -0.642$</td></tr>
</table>

<table>
<tr><td>charges</td><td>$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|c} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi \\ \hline 3 & 4 & 1 & -3 & -6 & -1 & 1 & -1 & -1 & -2 & 1 \end{array}\right)$</td></tr>
<tr><td>$\mathcal{O}(1)$</td><td>$(a_{ij}) \simeq \begin{pmatrix} 1.226 & -0.747 & 1.017 \\ 1.473 & 1.350 & 1.776 \\ -1.575 & 0.988 & 0.982 \end{pmatrix}, (b_{ij}) \simeq \begin{pmatrix} -1.012 & 1.947 & -1.941 \\ -1.834 & -1.410 & 1.170 \\ -0.614 & -0.872 & 1.405 \end{pmatrix}$</td></tr>
<tr><td>VEV, value</td><td>$v_1 \simeq 0.224$, $\mathcal{V}(\mathcal{Q}) \simeq -0.721$</td></tr>
</table>

A.2 Example models for two $U(1)$ symmetries

In this appendix we list some models with a single $U(1)$ symmetry with a high intrinsic value $\mathcal{V}(\mathcal{Q})$, found during chapter 2.

<table>
<tr><td>charges</td><td>$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 2 & 2 & 1 & -2 & 0 & 1 & -1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -2 & 0 & 0 & 1 \end{array}\right)$</td></tr>
<tr><td>$\mathcal{O}(1)$</td><td>$(a_{ij}) \simeq \begin{pmatrix} -1.898 & 0.834 & -0.587 \\ -0.575 & -0.592 & 1.324 \\ -1.123 & -1.265 & 0.982 \end{pmatrix}, (b_{ij}) \simeq \begin{pmatrix} -1.759 & 1.358 & 1.013 \\ -1.267 & 1.897 & -1.196 \\ 1.771 & 1.386 & -1.785 \end{pmatrix}$</td></tr>
<tr><td>VEVs, value</td><td>$(v_1, v_2) \simeq (0.079, 0.112)$, $\mathcal{V}(\mathcal{Q}) \simeq -0.390$</td></tr>
</table>

<table>
<tr><td>charges</td><td>$\mathcal{Q} = \left(\begin{array}{ccc|ccc|ccc|c|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 3 & 3 & 0 & 1 & 0 & 1 & -2 & -1 & -1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & -2 & 2 & 0 & -2 & -1 & 0 & 0 & 1 \end{array}\right)$</td></tr>
<tr><td>$\mathcal{O}(1)$</td><td>$(a_{ij}) \simeq \begin{pmatrix} 0.715 & -1.366 & -1.988 \\ 1.005 & 1.497 & 0.576 \\ 1.767 & -1.194 & 0.982 \end{pmatrix}, (b_{ij}) \simeq \begin{pmatrix} 1.195 & -1.352 & -1.410 \\ -0.979 & -1.860 & -0.521 \\ -1.932 & -0.946 & 0.912 \end{pmatrix}$</td></tr>
<tr><td>VEVs, value</td><td>$(v_1, v_2) \simeq (0.224, 0.224)$, $\mathcal{V}(\mathcal{Q}) \simeq -0.425$</td></tr>
</table>

<table>
<tr><td>charges</td><td>$\mathbb{Q} = \left(\begin{array}{ccc|ccc|ccc|c|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 2 & 3 & 1 & 1 & -1 & 1 & -1 & -1 & -2 & 0 & 1 & 0 \\ 2 & 2 & 1 & -1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \end{array}\right)$</td></tr>
</table>

$\mathcal{O}(1)$	$(a)_{ij}) \simeq \begin{pmatrix} 1.603 & 1.793 & 1.589 \\ 0.653 & -1.887 & -0.869 \\ -1.609 & -0.679 & 0.982 \end{pmatrix}, (b_{ij}) \simeq \begin{pmatrix} 1.369 & -1.107 & 1.151 \\ -1.658 & -1.831 & 1.868 \\ 1.819 & 1.395 & -0.813 \end{pmatrix}$
VEVs, Value	$(v_1, v_2) \simeq (0.316, 0.158)$, $\mathcal{V}(\mathcal{Q}) \simeq -0.480$

charges	$\mathcal{Q} = \left(\begin{array}{ccc\|ccc\|ccc\|c\|cc} Q_1 & Q_2 & Q_3 & u_1 & u_2 & u_3 & d_1 & d_2 & d_3 & H & \phi_1 & \phi_2 \\ \hline 1 & 3 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & -1 & 1 & 0 \\ 3 & 1 & 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right)$
$\mathcal{O}(1)$	$(a_{ij}) \simeq \begin{pmatrix} -1.231 & 1.029 & 0.753 \\ 0.955 & -1.997 & 1.507 \\ -1.265 & 1.447 & 0.982 \end{pmatrix}, (b_{ij}) \simeq \begin{pmatrix} 1.685 & 1.397 & -0.842 \\ 0.877 & 0.777 & -1.451 \\ -1.544 & 1.425 & 1.359 \end{pmatrix}$
VEVs, Value	$(v_1, v_2) \simeq (0.158, 0.079)$, $\mathcal{V}(\mathcal{Q}) \simeq -0.488$

B

Some Monad bundles on the bicubic

The table below contains the 28 monad bundles of genuine type $(r_B, r_C) = (6, 2)$ obtained on the bicubic manifold X, leading to models with 27 families and no anti-families. For these bundles equivariance with respect to the $\mathbb{Z}_3 \times \mathbb{Z}_3$ action on the manifold has been checked at the level of index divisibility for the line bundles in B and C. A number of necessary stability checks have been carried out. More precisely, we have found all the line bundles with entries in the range $[-3, 3]$ that inject into V and have checked that there exists a non-empty locus in the Kähler cone where all these line bundles have a negative slope, thus leaving V slope poly-stable.

In order to save space we use the notation $\mathcal{O}(k_1, k_2)$ instead of $\mathcal{O}_X(k_1, k_2)$. All the line bundles are to be understood on the bicubic threefold.

B	C
$\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(2,-1)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,2)$	$\mathcal{O}(1,4)\oplus\mathcal{O}(2,1)$
$\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)$	$\mathcal{O}(1,1)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(2,-1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(1,-2)\oplus\mathcal{O}(1,2)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(2,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-2,1)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,2)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(2,2)\oplus\mathcal{O}(2,2)$

$\mathcal{O}(-2,2)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,2)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,2)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(2,-2)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(0,2)\oplus\mathcal{O}(0,2)\oplus\mathcal{O}(0,2)\oplus\mathcal{O}(1,-2)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,2)\oplus\mathcal{O}(-1,2)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(2,-1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(2,-1)\oplus\mathcal{O}(2,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(2,0)\oplus\mathcal{O}(2,0)\oplus\mathcal{O}(2,0)$	$\mathcal{O}(1,1)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,2)\oplus\mathcal{O}(2,-1)$	$\mathcal{O}(1,4)\oplus\mathcal{O}(2,1)$
$\mathcal{O}(-2,2)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,4)\oplus\mathcal{O}(2,-1)\oplus\mathcal{O}(2,-1)$	$\mathcal{O}(1,5)\oplus\mathcal{O}(2,1)$
$\mathcal{O}(-2,1)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,2)\oplus\mathcal{O}(1,4)$	$\mathcal{O}(1,5)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,2)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,4)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(1,5)$
$\mathcal{O}(-1,2)\oplus\mathcal{O}(-1,2)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,-1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,4)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(1,5)$
$\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(1,-2)\oplus\mathcal{O}(1,2)\oplus\mathcal{O}(1,4)$	$\mathcal{O}(1,5)\oplus\mathcal{O}(2,2)$
$\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(1,-2)\oplus\mathcal{O}(1,4)\oplus\mathcal{O}(4,1)$	$\mathcal{O}(1,5)\oplus\mathcal{O}(5,1)$
$\mathcal{O}(-1,2)\oplus\mathcal{O}(-1,2)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(2,-1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,1)$
$\mathcal{O}(-2,2)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,2)$	$\mathcal{O}(1,4)\oplus\mathcal{O}(2,1)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)\oplus\mathcal{O}(1,1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(1,4)$
$\mathcal{O}(-1,1)\oplus\mathcal{O}(-1,2)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,2)$	$\mathcal{O}(1,1)\oplus\mathcal{O}(1,4)$
$\mathcal{O}(-2,1)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,0)\oplus\mathcal{O}(1,4)\oplus\mathcal{O}(4,1)$	$\mathcal{O}(1,5)\oplus\mathcal{O}(5,1)$
$\mathcal{O}(-1,2)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(0,1)\oplus\mathcal{O}(2,-1)\oplus\mathcal{O}(2,-1)$	$\mathcal{O}(1,2)\oplus\mathcal{O}(2,1)$

References

[1] Edward Witten. “Introducing M-Theory at Stony Brook”. In: (1995). https://www.math.stonybrook.edu/Videos/IMS/video.php?f=35-Witten.

[2] Juan Martin Maldacena. “The Large N limit of superconformal field theories and supergravity”. In: *Adv. Theor. Math. Phys.* 2 (1998), pp. 231–252. arXiv: `hep-th/9711200`.

[3] Edward Witten. “Anti-de Sitter space and holography”. In: *Adv. Theor. Math. Phys.* 2 (1998), pp. 253–291. arXiv: `hep-th/9802150`.

[4] Andrew Strominger and Cumrun Vafa. “Microscopic origin of the Bekenstein-Hawking entropy”. In: *Phys. Lett. B* 379 (1996), pp. 99–104. arXiv: `hep-th/9601029`.

[5] Wolfgang Lerche, Cumrun Vafa, and Nicholas P. Warner. “Chiral Rings in N=2 Superconformal Theories”. In: *Nucl. Phys. B* 324 (1989), pp. 427–474.

[6] P. Candelas et al. “Vacuum configurations for superstrings”. In: *Nucl. Phys. B* 258 (1985), pp. 46–74.

[7] Petr Horava and Edward Witten. “Heterotic and type I string dynamics from eleven-dimensions”. In: *Nucl. Phys. B* 460 (1996), pp. 506–524. arXiv: `hep-th/9510209`.

[8] Petr Horava and Edward Witten. “Eleven-dimensional supergravity on a manifold with boundary”. In: *Nucl. Phys. B* 475 (1996), pp. 94–114. arXiv: `hep-th/9603142`.

[9] Edward Witten. “Strong coupling expansion of Calabi-Yau compactification”. In: *Nucl. Phys. B* 471 (1996), pp. 135–158. arXiv: `hep-th/9602070`.

[10] Joseph Polchinski. “Dirichlet Branes and Ramond-Ramond charges”. In: *Phys. Rev. Lett.* 75 (1995), pp. 4724–4727. arXiv: `hep-th/9510017`.

[11] G. Aldazabal et al. “D-branes at singularities: A Bottom up approach to the string embedding of the standard model”. In: *JHEP* 08 (2000), p. 002. arXiv: `hep-th/0005067`.

[12] Cumrun Vafa. “Evidence for F theory”. In: *Nucl. Phys. B* 469 (1996), pp. 403–418. arXiv: `hep-th/9602022`.

[13] Bobby Samir Acharya. “M theory, Joyce orbifolds and superYang-Mills”. In: *Adv. Theor. Math. Phys.* 3 (1999), pp. 227–248. arXiv: `hep-th/9812205`.

[14] Sujay Ashok and Michael R. Douglas. “Counting flux vacua”. In: *JHEP* 01 (2004), p. 060. arXiv: `hep-th/0307049`.

[15] Washington Taylor and Yi-Nan Wang. “The F-theory geometry with most flux vacua”. In: *JHEP* 12 (2015), p. 164. arXiv: `1511.03209 [hep-th]`.

[16] Aditi Chandra et al. “Enumerating Calabi-Yau Manifolds: Placing bounds on the number of diffeomorphism classes in the Kreuzer-Skarke list”. In: (Oct. 2023). arXiv: `2310.05909 [hep-th]`.

[17] Naomi Gendler et al. “Counting Calabi-Yau Threefolds”. In: (Oct. 2023). arXiv: `2310.06820 [hep-th]`.

[18] Maximilian Kreuzer and Harald Skarke. “Complete classification of reflexive polyhedra in four-dimensions”. In: *Adv. Theor. Math. Phys.* 4 (2000), pp. 1209–1230. arXiv: `hep-th/0002240`.

[19] David Silver et al. “Mastering the game of Go with deep neural networks and tree search”. In: *nature* 529.7587 (2016), pp. 484–489.

[20] T. R. Harvey and A. Lukas. “Quark Mass Models and Reinforcement Learning”. In: *JHEP* 08 (2021), p. 161. arXiv: `2103.04759 [hep-th]`.

[21] Steve A. Abel et al. “Decoding Nature with Nature’s Tools: Heterotic Line Bundle Models of Particle Physics with Genetic Algorithms and Quantum Annealing”. In: *Fortsch. Phys.* 72.2 (2024), p. 2300260. arXiv: `2306.03147 [hep-th]`.

[22] Andrei Constantin, Thomas R. Harvey, and Andre Lukas. “Heterotic String Model Building with Monad Bundles and Reinforcement Learning”. In: *Fortsch. Phys.* 70.2-3 (2022), p. 2100186. arXiv: `2108.07316 [hep-th]`.

[23] Steven Abel et al. “Evolving Heterotic Gauge Backgrounds: Genetic Algorithms versus Reinforcement Learning”. In: *Fortsch. Phys.* 70.5 (2022), p. 2200034. arXiv: `2110.14029 [hep-th]`.

[24] Steven Abel et al. “String Model Building, Reinforcement Learning and Genetic Algorithms”. In: *Nankai Symposium on Mathematical Dialogues: In celebration of S.S.Chern’s 110th anniversary.* Nov. 2021. arXiv: `2111.07333 [hep-th]`.

[25] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266.

[26] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01.

[27] Stuart Raby. “SUSY GUT Model Building”. In: *Eur. Phys. J. C* 59 (2009), pp. 223–247. arXiv: `0807.4921 [hep-ph]`.

[28] Paul Langacker. *The standard model and beyond.* 2010.

[29] Luis E. Ibanez and Angel M. Uranga. *String theory and particle physics: An introduction to string phenomenology.* Cambridge University Press, Feb. 2012.

[30] Sidney R. Coleman and J. Mandula. “All Possible Symmetries of the S Matrix”. In: *Phys. Rev.* 159 (1967). Ed. by A. Zichichi, pp. 1251–1256.

[31] J. Polchinski. *String theory. Vol. 2: Superstring theory and beyond.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, Dec. 2007.

[32] Jean-Loup Gervais and B. Sakita. “Field Theory Interpretation of Supergauges in Dual Models”. In: *Nucl. Phys. B* 34 (1971). Ed. by K. Kikkawa, M. Virasoro, and S. R. Wadia, pp. 632–639.

[33] D. V. Volkov and V. P. Akulov. “Is the Neutrino a Goldstone Particle?” In: *Phys. Lett. B* 46 (1973), pp. 109–110.

[34] V. P. Akulov and D. V. Volkov. “Goldstone fields with spin 1/2”. In: *Theor. Math. Phys.* 18 (1974), p. 28.

[35] J. Wess and J. Bagger. *Supersymmetry and supergravity.* Princeton, NJ, USA: Princeton University Press, 1992.

[36] S. Ferrara, J. Wess, and B. Zumino. “Supergauge Multiplets and Superfields”. In: *Phys. Lett. B* 51 (1974), p. 239.

[37] V. Bargmann and Eugene P. Wigner. “Group Theoretical Discussion of Relativistic Wave Equations”. In: *Proc. Nat. Acad. Sci.* 34 (1948), p. 211.

[38] Steven Weinberg. *The Quantum theory of fields. Vol. 1: Foundations.* Cambridge University Press, June 2005.

[39] Steven Weinberg. *The quantum theory of fields. Vol. 3: Supersymmetry.* Cambridge University Press, June 2013.

[40] Tomas Ortin. *Gravity and Strings.* 2nd ed. Cambridge Monographs on Mathematical Physics. Cambridge University Press, July 2015.

[41] Stephen P. Martin. “A Supersymmetry primer”. In: *Adv. Ser. Direct. High Energy Phys.* 18 (1998). Ed. by Gordon L. Kane, pp. 1–98. arXiv: `hep-ph/9709356`.

[42] Polytope24. “A schematic illustration of the relationship between M-theory, the five superstring theories, and eleven-dimensional supergravity.” In: (2014). https://commons.wikimedia.org/wiki/File:Limits_of_M-theory.png.

[43] Paul A. M. Dirac. “An Extensible model of the electron”. In: *Proc. Roy. Soc. Lond. A* 268 (1962), pp. 57–67.

[44] G. Veneziano. “Construction of a crossing - symmetric, Regge behaved amplitude for linearly rising trajectories”. In: *Nuovo Cim. A* 57 (1968), pp. 190–197.

[45] Michael B. Green, John H. Schwarz, and Edward Witten. *Superstring Theory Vol. 1: 25th Anniversary Edition.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, Nov. 2012.

[46] Edward Witten. “String theory dynamics in various dimensions”. In: *Nucl. Phys. B* 443 (1995), pp. 85–126. arXiv: `hep-th/9503124`.

[47] Tom Banks et al. “M theory as a matrix model: A Conjecture”. In: *Phys. Rev. D* 55 (1997), pp. 5112–5128. arXiv: `hep-th/9610043`.

[48] J. Polchinski. *String theory. Vol. 1: An introduction to the bosonic string.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, Dec. 2007.

[49] Michael B. Green, John H. Schwarz, and Edward Witten. *Superstring Theory Vol. 2: 25th Anniversary Edition.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, Nov. 2012.

[50] David J. Gross et al. “The Heterotic String”. In: *Phys. Rev. Lett.* 54 (1985), pp. 502–505.

[51] Andre Lukas, Burt A. Ovrut, and Daniel Waldram. “On the four-dimensional effective action of strongly coupled heterotic string theory”. In: *Nucl. Phys. B* 532 (1998), pp. 43–82. arXiv: `hep-th/9710208`.

[52] Andre Lukas et al. “Heterotic M theory in five-dimensions”. In: *Nucl. Phys. B* 552 (1999), pp. 246–290. arXiv: `hep-th/9806051`.

[53] Lara B. Anderson and Mohsen Karkheiran. “TASI Lectures on Geometric Tools for String Compactifications”. In: *PoS* TASI2017 (2018), p. 013. arXiv: `1804.08792 [hep-th]`.

[54] Lara Briana Anderson. “Heterotic and M-theory Compactifications for String Phenomenology”. In: (Aug. 2008). arXiv: `0808.3621 [hep-th]`.

[55] D.D. Joyce. *Compact Manifolds with Special Holonomy.* Oxford mathematical monographs. Oxford University Press, 2000. URL: `https://books.google.co.uk/books?id=c3P-YUD8GZQC`.

[56] Shing-Tung Yau. “On the ricci curvature of a compact kähler manifold and the complex monge-ampére equation, I”. In: *Communications on Pure and Applied Mathematics* 31.3 (1978), pp. 339–411.

[57] S. K. Donaldson. “Anti Self-dual Yang-Mills Connections Over Complex Algebraic Surfaces and Stable Vector Bundles”. In: *Proc. Lond. Math. Soc.* 50 (1985), pp. 1–26.

[58] K. Uhlenbeck and S. T. Yau. “On the existence of hermitian-yang-mills connections in stable vector bundles”. In: *Commun. Pure Appl. Math.* 39.S1 (1986), S257–S293.

[59] Lara B. Anderson et al. “Hypercharge Flux in Heterotic Compactifications”. In: *Phys. Rev. D* 91.4 (2015), p. 046008. arXiv: `1411.0034 [hep-th]`.

[60] Eric R. Sharpe. “Discrete torsion and gerbes. 1.” In: (Sept. 1999). arXiv: `hep-th/9909108`.

[61] Eric R. Sharpe. “Discrete torsion and gerbes (2)”. In: (Sept. 1999). arXiv: `hep-th/9909120`.

[62] Lara B. Anderson et al. “Two Hundred Heterotic Standard Models on Smooth Calabi-Yau Threefolds”. In: *Phys. Rev. D* 84 (2011), p. 106005. arXiv: `1106.4804 [hep-th]`.

[63] Lara B. Anderson et al. “Heterotic Line Bundle Standard Models”. In: *JHEP* 06 (2012), p. 113. arXiv: `1202.1757 [hep-th]`.

[64] Evgeny I. Buchbinder, Andrei Constantin, and Andre Lukas. “The Moduli Space of Heterotic Line Bundle Models: a Case Study for the Tetra-Quadric”. In: *JHEP* 03 (2014), p. 025. arXiv: `1311.1941 [hep-th]`.

[65] Edward Witten. “Dimensional Reduction of Superstring Models”. In: *Phys. Lett. B* 155 (1985), p. 151.

[66] Philip Candelas and Xenia de la Ossa. “Moduli Space of Calabi-Yau Manifolds”. In: *Nucl. Phys. B* 355 (1991), pp. 455–481.

[67] Edward Witten. “Nonperturbative superpotentials in string theory”. In: *Nucl. Phys. B* 474 (1996), pp. 343–360. arXiv: `hep-th/9604030`.

[68] Evgeny I. Buchbinder, Ron Donagi, and Burt A. Ovrut. “Vector bundle moduli superpotentials in heterotic superstrings and M theory”. In: *JHEP* 07 (2002), p. 066. arXiv: `hep-th/0206203`.

[69] Evgeny Buchbinder et al. "Heterotic Instanton Superpotentials from Complete Intersection Calabi-Yau Manifolds". In: *JHEP* 10 (2017), p. 032. arXiv: `1707.07214 [hep-th]`.

[70] Evgeny I. Buchbinder et al. "Heterotic Instantons for Monad and Extension Bundles". In: *JHEP* 02 (2020), p. 081. arXiv: `1912.07222 [hep-th]`.

[71] Hans Peter Nilles. "Gaugino condensation and SUSY breakdown". In: *Cargese School of Particle Physics and Cosmology: the Interface.* Feb. 2004, pp. 397–433. arXiv: `hep-th/0402022`.

[72] Ignatios Antoniadis and M. Quiros. "On the M theory description of gaugino condensation". In: *Phys. Lett. B* 416 (1998), pp. 327–333. arXiv: `hep-th/9707208`.

[73] Andre Lukas, Burt A. Ovrut, and Daniel Waldram. "Gaugino condensation in M theory on s**1 / Z(2)". In: *Phys. Rev. D* 57 (1998), pp. 7529–7538. arXiv: `hep-th/9711197`.

[74] Stefan Blesneag, Evgeny I. Buchbinder, and Andre Lukas. "Holomorphic Yukawa Couplings for Complete Intersection Calabi-Yau Manifolds". In: *JHEP* 01 (2017), p. 119. arXiv: `1607.03461 [hep-th]`.

[75] Stefan Blesneag et al. "Holomorphic Yukawa Couplings in Heterotic String Theory". In: *JHEP* 01 (2016), p. 152. arXiv: `1512.05322 [hep-th]`.

[76] S. Blesneag. "Holomorphic Yukawa Couplings in Heterotic String Theory". PhD thesis. Oxford U., 2021. arXiv: `2204.01165 [hep-th]`.

[77] Magdalena Larfors et al. "Learning Size and Shape of Calabi-Yau Spaces". In: (Nov. 2021). arXiv: `2111.01436 [hep-th]`.

[78] Magdalena Larfors et al. "Numerical metrics for complete intersection and Kreuzer–Skarke Calabi–Yau manifolds". In: *Mach. Learn. Sci. Tech.* 3.3 (2022), p. 035014. arXiv: `2205.13408 [hep-th]`.

[79] Anthony Ashmore, Yang-Hui He, and Burt A. Ovrut. "Machine Learning Calabi–Yau Metrics". In: *Fortsch. Phys.* 68.9 (2020), p. 2000068. arXiv: `1910.08605 [hep-th]`.

[80] Lara B. Anderson et al. "Moduli-dependent Calabi-Yau and SU(3)-structure metrics from Machine Learning". In: *JHEP* 05 (2021), p. 013. arXiv: `2012.04656 [hep-th]`.

[81] Vishnu Jejjala, Damian Kaloni Mayorga Pena, and Challenger Mishra. "Neural network approximations for Calabi-Yau metrics". In: *JHEP* 08 (2022), p. 105. arXiv: `2012.15821 [hep-th]`.

[82] Michael R. Douglas, Subramanian Lakshminarasimhan, and Yidi Qi. "Numerical Calabi-Yau metrics from holomorphic networks". In: (Dec. 2020). arXiv: `2012.04797 [hep-th]`.

[83] Anthony Ashmore et al. "Calabi-Yau Metrics, Energy Functionals and Machine-Learning". In: *International Journal of Data Science in the Mathematical Sciences* 1.1 (2023), pp. 49–61. arXiv: `2112.10872 [hep-th]`.

[84] Mathis Gerdes and Sven Krippendorf. "CYJAX: A package for Calabi-Yau metrics with JAX". In: *Mach. Learn. Sci. Tech.* 4.2 (2023), p. 025031. arXiv: `2211.12520 [hep-th]`.

[85] Andrei Constantin et al. “Computation of Quark Masses from String Theory”. In: (Feb. 2024). arXiv: `2402.01615 [hep-th]`.

[86] Liam McAllister and Fernando Quevedo. “Moduli Stabilization in String Theory”. In: (Oct. 2023). arXiv: `2310.20559 [hep-th]`.

[87] Oliver DeWolfe et al. “Type IIA moduli stabilization”. In: *JHEP* 07 (2005), p. 066. arXiv: `hep-th/0505160`.

[88] Shamit Kachru et al. “De Sitter vacua in string theory”. In: *Phys. Rev. D* 68 (2003), p. 046005. arXiv: `hep-th/0301240`.

[89] Vijay Balasubramanian et al. “Systematics of moduli stabilisation in Calabi-Yau flux compactifications”. In: *JHEP* 03 (2005), p. 007. arXiv: `hep-th/0502058`.

[90] Andre Lukas and Cyril Matti. “G-structures and Domain Walls in Heterotic Theories”. In: *JHEP* 01 (2011), p. 151. arXiv: `1005.5302 [hep-th]`.

[91] Lara B. Anderson et al. “Stabilizing All Geometric Moduli in Heterotic Calabi-Yau Vacua”. In: *Phys. Rev. D* 83 (2011), p. 106011. arXiv: `1102.0011 [hep-th]`.

[92] Andre Lukas, Zygmunt Lalak, and Eirik Eik Svanes. “Heterotic Moduli Stabilisation and Non-Supersymmetric Vacua”. In: *JHEP* 08 (2015), p. 020. arXiv: `1504.06978 [hep-th]`.

[93] Tristan Hubsch. *Calabi-Yau manifolds: A Bestiary for physicists.* Singapore: World Scientific, 1994.

[94] K. Hori et al. *Mirror symmetry.* Vol. 1. Clay mathematics monographs. Providence, USA: AMS, 2003.

[95] Brian R. Greene. “String theory on Calabi-Yau manifolds”. In: *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 96): Fields, Strings, and Duality.* June 1996, pp. 543–726. arXiv: `hep-th/9702155`.

[96] Victor V. Batyrev. “Dual polyhedra and mirror symmetry for Calabi-Yau hypersurfaces in toric varieties”. In: *J. Alg. Geom.* 3 (1994), pp. 493–545. arXiv: `alg-geom/9310003`.

[97] Yang-Hui He et al. “Heterotic Model Building: 16 Special Manifolds”. In: *JHEP* 06 (2014), p. 077. arXiv: `1309.0223 [hep-th]`.

[98] P. Candelas et al. “Complete Intersection Calabi-Yau Manifolds”. In: *Nucl. Phys. B* 298 (1988), p. 493.

[99] Paul S. Green, Tristan Hubsch, and Carsten A. Lutken. “All Hodge Numbers of All Complete Intersection Calabi-Yau Manifolds”. In: *Class. Quant. Grav.* 6 (1989), pp. 105–124.

[100] Volker Braun. “On Free Quotients of Complete Intersection Calabi-Yau Manifolds”. In: *JHEP* 04 (2011), p. 005. arXiv: `1003.3235 [hep-th]`.

[101] James Gray and Juntao Wang. “Free quotients of favorable Calabi-Yau manifolds”. In: *JHEP* 07 (2022), p. 116. arXiv: `2112.12683 [hep-th]`.

[102] R. Hartshorne. *Algebraic Geometry.* Graduate Texts in Mathematics. Springer, 1977. URL: `https://books.google.co.uk/books?id=3rtX9t-nnvwC`.

[103] P. Griffiths and J. Harris. *Principles of Algebraic Geometry*. Wiley Classics Library. Wiley, 2014. URL: `https://books.google.co.uk/books?id=-01YBAAAQBAJ`.

[104] H. Schenck. *Computational Algebraic Geometry*. London Mathematical Society Student Texts. Cambridge University Press, 2003. URL: `https://books.google.co.uk/books?id=RzWbgVRUzGoC`.

[105] Lara B. Anderson, Yang-Hui He, and Andre Lukas. "Heterotic Compactification, An Algorithmic Approach". In: *JHEP* 07 (2007), p. 049. arXiv: `hep-th/0702210`.

[106] Daniel R. Grayson and Michael E. Stillman. *Macaulay2, a software system for research in algebraic geometry*. Available at `http://www2.macaulay2.com`.

[107] Andrei Constantin and Andre Lukas. "Formulae for Line Bundle Cohomology on Calabi-Yau Threefolds". In: *Fortsch. Phys.* 67.12 (2019), p. 1900084. arXiv: `1808.09992 [hep-th]`.

[108] Magdalena Larfors and Robin Schneider. "Line bundle cohomologies on CICYs with Picard number two". In: *Fortsch. Phys.* 67.12 (2019), p. 1900083. arXiv: `1906.00392 [hep-th]`.

[109] Callum R. Brodie, Andrei Constantin, and Andre Lukas. "Flops, Gromov-Witten Invariants and Symmetries of Line Bundle Cohomology on Calabi-Yau Three-folds". In: (Oct. 2020). arXiv: `2010.06597 [hep-th]`.

[110] Callum R. Brodie and Andrei Constantin. "Cohomology Chambers on Complex Surfaces and Elliptically Fibered Calabi-Yau Three-folds". In: (Sept. 2020). arXiv: `2009.01275 [hep-th]`.

[111] Callum R. Brodie et al. "Topological Formulae for the Zeroth Cohomology of Line Bundles on del Pezzo and Hirzebruch Surfaces". In: *Compl. Manif.* 8.1 (2021), pp. 223–229. arXiv: `1906.08363 [math.AG]`.

[112] Callum R. Brodie et al. "Machine Learning Line Bundle Cohomology". In: *Fortsch. Phys.* 68.1 (2020), p. 1900087. arXiv: `1906.08730 [hep-th]`.

[113] Daniel Klaewer and Lorenz Schlechter. "Machine Learning Line Bundle Cohomologies of Hypersurfaces in Toric Varieties". In: *Phys. Lett. B* 789 (2019), pp. 438–443. arXiv: `1809.02547 [hep-th]`.

[114] Callum R. Brodie et al. "Index Formulae for Line Bundle Cohomology on Complex Surfaces". In: *Fortsch. Phys.* 68.2 (2020), p. 1900086. arXiv: `1906.08769 [hep-th]`.

[115] James Halverson, Brent Nelson, and Fabian Ruehle. "Branes with Brains: Exploring String Vacua with Deep Reinforcement Learning". In: *JHEP* 06 (2019), p. 003. arXiv: `1903.11616 [hep-th]`.

[116] Steven Abel and John Rizos. "Genetic Algorithms and the Search for Viable String Vacua". In: *JHEP* 08 (2014), p. 010. arXiv: `1404.7359 [hep-th]`.

[117] Alex Cole et al. "Probing the Structure of String Theory Vacua with Genetic Algorithms and Reinforcement Learning". In: *35th Conference on Neural Information Processing Systems*. Nov. 2021. arXiv: `2111.11466 [hep-th]`.

[118] Steve A. Abel et al. "Cosmic Inflation and Genetic Algorithms". In: *Fortsch. Phys.* 71.1 (2023), p. 2200161. arXiv: `2208.13804 [hep-th]`.

[119] Steven Abel, David G. Cerdeño, and Sandra Robles. “The Power of Genetic Algorithms: what remains of the pMSSM?” In: (May 2018). arXiv: `1805.03615 [hep-ph]`.

[120] B. C. Allanach, D. Grellscheid, and F. Quevedo. “Genetic algorithms and experimental discrimination of SUSY models”. In: *JHEP* 07 (2004), p. 069. arXiv: `hep-ph/0406277`.

[121] Sven Krippendorf, Rene Kroepsch, and Marc Syvaeri. “Revealing systematics in phenomenologically viable flux vacua with reinforcement learning”. In: (July 2021). arXiv: `2107.04039 [hep-th]`.

[122] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley series in artificial intelligence. Addison-Wesley, 1989. URL: `https://books.google.co.uk/books?id=3_RQAAAAMAAJ`.

[123] Fabian Ruehle. “Data science applications to string theory”. In: *Phys. Rept.* 839 (2020), pp. 1–117.

[124] Steven Abel et al. “Mathematica package for Genetic Algorithms”. In: (2022). URL: `"https://github.com/harveyThomas4692/GAMathematica"`.

[125] Steven Abel et al. “A realisation of genetic algorithms in C”. In: (2023). URL: `https://github.com/harveyThomas4692/GA-C`.

[126] Diederik Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *International Conference on Learning Representations* (Dec. 2014).

[127] Shengze Cai et al. “Physics-informed neural networks (PINNs) for fluid mechanics: A review”. In: *Acta Mechanica Sinica* 37.12 (2021), pp. 1727–1738.

[128] R.S. Sutton and A.G. Barto. *Reinforcement Learning, second edition: An Introduction*. Adaptive Computation and Machine Learning series. MIT Press, 2018. URL: `https://books.google.co.uk/books?id=5s-MEAAAQBAJ`.

[129] C. D. Froggatt and Holger Bech Nielsen. “Hierarchy of Quark Masses, Cabibbo Angles and CP Violation”. In: *Nucl. Phys. B* 147 (1979), pp. 277–298.

[130] Miriam Leurer, Yosef Nir, and Nathan Seiberg. “Mass matrix models”. In: *Nucl. Phys. B* 398 (1993), pp. 319–342. arXiv: `hep-ph/9212278`.

[131] Miriam Leurer, Yosef Nir, and Nathan Seiberg. “Mass matrix models: The Sequel”. In: *Nucl. Phys. B* 420 (1994), pp. 468–504. arXiv: `hep-ph/9310320`.

[132] Lara B. Anderson et al. “Heterotic standard models from smooth Calabi-Yau three-folds”. In: *PoS* CORFU2011 (2011). Ed. by K. Anagnostopoulos et al., p. 096.

[133] Lara B. Anderson et al. “A Comprehensive Scan for Heterotic SU(5) GUT models”. In: *JHEP* 01 (2014), p. 047. arXiv: `1307.4787 [hep-th]`.

[134] Andrei Constantin, Yang-Hui He, and Andre Lukas. “Counting String Theory Standard Models”. In: *Phys. Lett. B* 792 (2019), pp. 258–262. arXiv: `1810.00444 [hep-th]`.

[135] Steven Abel et al. *An environment for heterotic line bundle models, realised in C*. Tech. rep. 2023. URL: `https://github.com/harveyThomas4692/GA-LineBundles`.

[136] Magdalena Larfors and Robin Schneider. “Explore and Exploit with Heterotic Line Bundle Models”. In: *Fortsch. Phys.* 68.5 (2020), p. 2000034. arXiv: `2003.04817 [hep-th]`.

[137] Lara B. Anderson et al. “Fibrations in CICY Threefolds”. In: *JHEP* 10 (2017), p. 077. arXiv: `1708.07907 [hep-th]`.

[138] Andreas Braun, Andre Lukas, and Chuang Sun. “Discrete Symmetries of Calabi Yau Hypersurfaces in Toric Four-Folds”. In: *Commun. Math. Phys.* 360.3 (2018), pp. 935–984. arXiv: `1704.07812 [hep-th]`.

[139] Callum Brodie et al. “Flops for complete intersection Calabi-Yau threefolds”. In: *J. Geom. Phys.* 186 (2023), p. 104767. arXiv: `2112.12106 [hep-th]`.

[140] Callum R. Brodie et al. “Geodesics in the extended Kähler cone of Calabi-Yau threefolds”. In: *JHEP* 03 (2022), p. 024. arXiv: `2108.10323 [hep-th]`.

[141] Martin Bies et al. “Machine Learning and Algebraic Approaches towards Complete Matter Spectra in 4d F-theory”. In: *JHEP* 01 (2021), p. 196. arXiv: `2007.00009 [hep-th]`.

[142] Lara B. Anderson, Yang-Hui He, and Andre Lukas. “Monad Bundles in Heterotic String Compactifications”. In: *JHEP* 07 (2008), p. 104. arXiv: `0805.2875 [hep-th]`.

[143] Lara B. Anderson et al. “Exploring Positive Monad Bundles And A New Heterotic Standard Model”. In: *JHEP* 02 (2010), p. 054. arXiv: `0911.1569 [hep-th]`.

[144] Abhishek Dubey, Sven Krippendorf, and Andreas Schachner. “JAXVacua — a framework for sampling string vacua”. In: *JHEP* 12 (2023), p. 146. arXiv: `2306.06160 [hep-th]`.

[145] Simon Clark. “What’s it like studying string theory at Oxford?” In: (2022). https://www.youtube.com/watch?v=eoG4I5WZZ0w.