Technische Universität München

Machine Learning in String Theory

Dissertation

by

Erik Thomas Parr

Physik-Department T75, TUM

& SFB 1258

TECHNISCHE UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR PHYSIK

PHYSIK-DEPARTMENT T75

Machine Learning in String Theory

ERIK THOMAS PARR

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Stefan Schönert
Prüfer der Dissertation:	1.	Prof. Dr. Andreas Weiler
	2.	Prof. Dr. Ivo Sachs

Die Dissertation wurde am 25.05.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 10.09.2020 angenommen.

Abstract

In this thesis, we investigate the heterotic orbifold landscape under phenomenological considerations. The reason for a non-vanishing cosmological constant in non-supersymmetric heterotic orbifolds is studied. For supersymmetric orbifold compactifications, we apply machine learning methods to improve the search for MSSM-like string models. We use autoencoder neural networks to draw a chart of the heterotic orbifold landscape and identify islands where MSSM-like models accumulate. Furthermore, we utilize contrast mining to observe constraints that drastically reduce the landscape by excluding areas with a tiny probability of holding MSSM-like models. We show that the phenomenological characteristics of MSSM-like models depend on the specific orbifold compactification. Thus, we can utilize a predictive model to infer the orbifold origin of the MSSM.

Zusammenfassung

In dieser Arbeit untersuchen wir die heterotische Stringlandschaft unter phänomenologischen Aspekten. Zu Beginn analysieren wir den Grund für eine nicht verschwindende kosmologische Konstante in nicht supersymmetrischen heterotischen Orbifaltigkeiten. Für supersymmetrische Orbifaltigkeit-Kompaktifizierungen werden Methoden des maschinellen Lernens angewendet, um die Suche nach MSSM ähnlichen Modellen zu verbessern. Wir nutzen Autoencoder neuronale Netze, um eine Karte der heterotischen Orbifaltigkeits-Landschaft zu erstellen und identifizieren dabei Inseln, auf denen sich MSSM ähnliche Modelle ansammeln. Darüber hinaus entwickeln wir mittels Contrast-Mining Bedingungen, welche die Landschaft drastisch reduzieren, indem Bereiche mit einer geringen Wahrscheinlichkeit für MSSM ähnliche Modelle ausgeschlossen werden. Wir zeigen, dass die phänomenologischen Eigenschaften von MSSM ähnlichen Modellen entscheidend von der spezifischen Orbifaltigkeits-Kompaktifizierung abhängen. Daher können wir ein Vorhersagemodell entwickeln, um auf die Orbifaltigkeit des MSSM zu schließen.

Danksagung

Zuerst möchte ich mich bei Prof. Dr. Andreas Weiler bedanken, der es mir ermöglicht hat, an der TU München meine Forschung fortzuführen und zu promovieren. Ein großer Dank gebührt an dieser Stelle auch Dr. Patrick Vaudrevange für die fachliche Betreuung. Dank deiner gelungenen Mischung aus Forschungsimpulsen und Freiheit konnte ich meiner Forschung mit viel Freude und Eifer nachgehen.

Weiterhin gebührt Karin Ramm ein großer Dank für die Unterstützung beim Durchkämpfen der administrativen Fallstricke. Ein Teil dieser Arbeit benötigte zudem das T30 Rechencluster, für das ich Stefan Recksiegel danken möchte.

Besonders meinem Büro - Sebastian, Andreas, Kai, Anja, Jan, Andi - ein herzliches Danke! Die vielen Diskussionen über Physik sowie der wichtige Austausch von Klatsch und Tratsch haben jeden Tag kurzweilig erscheinen lassen und waren eine wesentliche Unterstützung in den letzten Jahren. Es war eine tolle Zeit, auf die ich gerne und wehmütig zurückblicken werde. Meine Tischtennis- und Schafkopffähigkeiten werden wohl nie wieder dieses Niveau erreichen.

Ganz besonders möchte ich an dieser Stelle noch meinen Freunden und fleißigen Korrekturlesern Moritz Platscher, Kai Urban, Alexander Baur, Sebastian Ingenhütt und Anja Brenner danken. Eure wertvollen Anregungen haben diese Arbeit erst zu dem gemacht, was sie ist. Vielen Dank für eure Zeit und Mühe!

Abschließend möchte ich noch den Menschen danken, die mich stets aus dem vermeintlichen "Hintergrund" gestärkt und gestützt haben. Ein großer Dank geht an meine Eltern Jürgen and Martina sowie an meine Schwester Angela. Eure bedingungslose Unterstützung und aufmunternden Worte haben mir in dieser Zeit viel Kraft gegeben.

Besonders hervorheben möchte ich meine bezaubernde Freundin Babsi, die mich unermüdlich angespornt, beraten und unterstützt hat. Ohne deine Rundumbetreuung, deinen starken Rückhalt und deine grenzenlose Liebe hätte ich diese Arbeit nicht schreiben können.

Danke.

Contents

Chapter 1

Introduction

In 2020 we are looking back at a memorable decade of scientific discoveries in fundamental physics. Two of them especially praise the synergy between theoretical and experimental physics. On July 4 in 2012, the ATLAS and CMS experiments at CERN's Large Hadron Collider (LHC) announced the observation of a new particle consistent with the Higgs boson [1, 2]. On February 11 in 2016, LIGO Scientific Collaboration and the Virgo Collaboration announced the first direct observation of gravitational waves [3–8].

These two dates are milestones of modern physics. Interestingly, their origins lie deep in the past. In the 1960s, physicists developed an understanding of the Brout-Englert-Higgs mechanism. This mechanism allows to describe the electroweak sector of particle physics in the Standard Model (SM) [9–13], by predicting a particle called the Higgs boson. The prediction of gravitational waves lies even further back in the past. Albert Einstein first described them in 1916 as an outcome of the theory of general relativity [14]. The relation of those experimental discoveries to theoretical considerations implies that those events not only mark essential steps on the journey to unveil the operations of nature. Indeed, they honor the immense predictive power of these theoretical models that goes far beyond the known phenomena they were built on to express the underlying laws of nature.

Especially the SM is in impressive agreement with experiment, and all ingredients, like an effective Higgs field, have been found. However, similar considerations, like in the past, motivate physicists to improve this model. The SM describes the strong and electroweak interactions of the known elementary particles. The simple fact that it requires as many as 28 parameters [15], which are exclusively determined by experiment, implies the need for a more fundamental understanding of nature. Additionally, some observations are not understood within the SM, e.g. the origin of dark matter or the number of generations of quarks and leptons. Unfortunately, the LHC was not able to produce significant signals of new physics so far. Equally, the theory of general relativity reveals shortcomings under certain conditions. Quantized theories describe high energy physics to astonishing precision. This quantization is a big hurdle for general relativity since canonical quantization leads to an unrenormalizable theory. This issue is the missing piece of quantum gravity.

In order to tackle the mysteries of physics beyond the SM, there are two approaches. One is the bottom-up construction. Here, the SM is modified in a modest way such that the effects of certain injections can be classified to evaluate their value in describing aspects of nature. The approach taken in this thesis is the top-down approach. There, we make use of a mathematical framework that incorporates the two most fundamental theories of our Universe: on the one side, the already mentioned SM as a realization of Quantum Field Theory (QFT), which currently dominates our understanding of physics, on the other side, general relativity. These two complementary descriptions of nature should ultimately be unified into one fundamental theory. In order to get a UV-complete theory of quantum gravity, there are a few different approaches known. For this thesis, the framework in favor is string theory, which unifies gravity and QFT in a higher dimensional quantum theory. The reason for having chosen string theory is that it does not only quantize gravity, like the asymptotic safety approach [16]. It is far more restrictive and unifies all fundamental particles to be excitation modes of a more fundamental object, i.e. a string. This concept is significantly more restrictive and leads to a framework with only one parameter: the string length l_s. Additionally, the value of string theory in this thesis is connected to the possibility of computing the low energy spectrum of a particular model, which describes the spectrum of a QFT. The resulting particle spectrum can then be compared to the SM. This possibility would not be given in other competitive approaches like loop quantum gravity [17].

The predictive power of string theory goes even further. In order to be a consistent theory, i.e. Weyl anomaly free, the framework demands certain values for the space-time strings can propagate in. Since a QFT can, a priori, be formulated in any dimension, experiments determine this value for the SM to three space and one time dimension. Considering that string theory also has to reproduce the observed 4D laws of physics, the extra dimensions have to be hidden from direct experimental observation. This can be achieved by the so-called compactification technique, which ‘wraps’ up these extra dimensions, thus making them inaccessible to today’s low energy range. This compactification comes with a huge drawback: it gives rise to the so-called string landscape. The number of possibilities to compactify the extra dimension is enormous and was estimated early on to $\sim 10^{1500}$ [18]. However, we will see that this large number of possible choices does not lead to phenomenological valid 4D theories in general. Actually, the amount of possible compactifications that come close to describing the known phenomena is a very tiny fraction of the possible compactifications that are consistent in the string theory framework. Once the compactification is selected, the physics is determined completely, i.e. the complete particle content as well as all couplings and remnant discrete symmetries that are possible flavor symmetries. This implies that each compactification choice leaves its particular imprint in the resulting 4D spectrum. Thus, the selection of such a compactification configuration is not comparable to fitting a parameter to experiment. Instead, the opportunity to select a compactification compares to deciding for the concrete model inside the string theory framework. This immense constraining power of the string theory framework could guide us towards QFTs that ultimately match to an UV-complete theory. More importantly, it might provide hints towards the origin and interplay of the parameters needed in the SM.

From the phenomenological side of ‘bottom-up’ constructions, a very promising extension of the SM is to introduce a new kind of symmetry, i.e. supersymmetry. In contrast to the known symmetries, this transformation has a spinorial generator and introduces a symmetry between fermions and bosons. The popular extension of the SM, the minimal supersymmetric standard model (MSSM), then predicts superpartners for the known particles. If those exist, their contributions cancel various divergent effects that would appear in the SM. However, space-time supersymmetry is not demanded by the string theory framework in general. There exist well-defined non-supersymmetric string theories. Moreover, even if we start with supersymmetric string theory, the concept of compactification does not preserve 4D supersymmetry in general. These string theories suffer from a non-vanishing cosmological constant at one-loop. The absence of supersymmetric observations at LHC reintroduces the interest to those non-supersymmetric theories even though they come along with problematic effects. As the purpose of this thesis is to link string theory and physics beyond the SM, both cases are considered, and we will study non-supersymmetric and supersymmetric string theory.

In chapter 3, the first topic we are going to investigate using non-supersymmetric string theory is the issue of the non-vanishing cosmological constant in the context of the so-called orbifold compactification method. We determine the exact terms of the partition function that do not vanish and contribute to the cosmological constant. Furthermore, we show how this is connected to a conjecture of finite group theory. For all relevant finite groups in heterotic orbifolds as well as for an even broader set of groups using the SmallGroups Library of GAP [19], the conjecture is explicitly verified.

Rather than generic properties, the search for an explicit realistic 4D model from string theory is an essential proof of concept. Most work on this issue is done in the string theory literature in the context of MSSM-like models. In chapters 4 to 8, we continue the thesis with a search for realistic MSSM-like models in the vast landscape of string compactifications. Concretely, we investigate the existence of MSSM-like models in the $\mathrm{E}_8 \times \mathrm{E}_8$ heterotic orbifold landscape with orbifold compactifications that lead to $\mathcal{N} = 1$ supersymmetric models. In the past, such searches have been conducted by a purely random sampling of the heterotic orbifold landscape. We are going to study the properties of the landscape by using techniques from data mining, more concretely machine learning, and deep learning. These methods are especially suitable to find patterns in large high dimensional datasets. Hence, the vast landscape of possible 4D string vacua, each specified by $\mathcal{O}(100)$ compactification parameters, represents a prime example of applying these techniques. In chapter 5, the starting point will be the application of an autoencoder neural network to the landscape, which will identify fertile patches. The success of the autoencoder approach shows that there are hidden patterns in the string landscape that have to be obeyed in order to connect string theory to the observable world.

Due to this, in chapter 6 we have chosen to incorporate the ‘bottom-up’ approach by informing the search algorithm about the phenomenological needs of the models during search time. Namely, we compute the intermediate physical properties of our string model while it is constructed and determine early on if the current subspace of the heterotic landscape can generate a model that fulfills the phenomenological needs. Furthermore, we study the symmetries of the string landscape and reduce the search space heavily by excluding the Weyl symmetry from the search process. This kind of study should ultimately promote the landscape to a practically finite search problem, such that all inequivalent physical models can be extracted from the landscape. On top of the previous improvements, we will use contrast mining. In this scenario, we make use of traditional machine learning algorithms that lead to a more direct knowledge gain in terms of interpretability. On the one hand, these contrast patterns allow us to find novel MSSM-like models in the landscape that were never observed nor believed to exist in the past. On the other hand, the contrast patterns are constraints on physical properties and lead to new insights on the particle spectrum of MSSM-like models from string theory.

Furthermore, the ‘bottom-up’ models can be tightly intertwined with the string landscape, as it is possible to infer the orbifold origin of ‘bottom-up’ models in the landscape (chapter 7) based on their phenomenology. It turns out that specific extensions of the MSSM particle spectrum can only be realized in particular subspaces of the string landscape. Hence, the origin of these extensions, as well as the exact MSSM (that has not been found in string theory compactifications so far), can be guided to specific regions of the string landscape.

Finally, before we conclude in chapter 9, the interplay of the different machine learning approaches is analyzed in chapter 8, where we also outline further improvements by synergy effects between the approaches.

As a closing remark, all procedures developed for the supersymmetric case can straightforwardly be transferred to the non-supersymmetric case by obtaining and studying those datasets. Notably, the analytic reduction constraints we derive are necessary conditions for the SM and not exclusively for the MSSM.

List of Publications

[20] E. Parr, P. K. Vaudrevange, and M. Wimmer, "Predicting the orbifold origin of the MSSM," *Fortsch. Phys.*, vol. 68, no. 5, p. 2 000 032, 2020. DOI: `10.1002/prop.202000032`. arXiv: `2003.01732 [hep-th]`

[21] E. Parr and P. K. S. Vaudrevange, "Contrast data mining for the MSSM from strings," *Nucl. Phys.*, vol. B952, p. 114 922, 2020. DOI: `10.1016/j.nuclphysb.2020.114922`. arXiv: `1910.13473 [hep-th]`

[22] S. Biermann, A. Mütter, E. Parr, M. Ratz, and P. K. S. Vaudrevange, "Discrete remnants of orbifolding," *Phys. Rev.*, vol. D100, no. 6, p. 066 030, 2019. DOI: `10.1103/PhysRevD.100.066030`. arXiv: `1906.10276 [hep-ph]`

[23] A. Mütter, E. Parr, and P. K. S. Vaudrevange, "Deep learning in the heterotic orbifold landscape," *Nucl. Phys.*, vol. B940, pp. 113–129, 2019. DOI: `10.1016/j.nuclphysb.2019.01.013`. arXiv: `1811.05993 [hep-th]`

[24] S. Groot Nibbelink, O. Loukas, A. Mütter, E. Parr, and P. K. S. Vaudrevange, "Tension Between a Vanishing Cosmological Constant and Non-Supersymmetric Heterotic Orbifolds," 2017. arXiv: `1710.09237 [hep-th]`, accepted by Fortschritte der Physik

Disclaimer

The results presented in this thesis represent original scientific work and have previously been published. To summarize:

- Chapter 3 and appendix A are based on [24] .
- The results of chapter 5 are adapted from [23] .
- Parts of chapter 6 were studied in [21] .
- Results from chapter 7 and appendix D.4 are based on [20] .

Chapter 2

The string theory framework

The upcoming chapter summarizes the most important aspects this thesis relies on and sets the notation. In this, we strongly rely on [25–28] for general string theory. Focused on the phenomenological implications the textbook [29] is recommended. Finally, especially related to the heterotic orbifold compactification process are [30–32]. The reader is referred to these for a more detailed discussion of the matter.

2.1 Heterotic string theory

String theory is an ambitious approach towards unifying the theories of fundamental physics into a single quantum-theoretical framework. The core observation that led to this idea is the ability to quantize gravity within string theory. From quantized strings, there emerges a spin-2 particle that can be identifed with the graviton [33, 34] as it provides the proper interactions at low energies [35]. Linking the string coupling and 4D gravitational couplings, this association demands that the string scale M_s is of the order of the Planck scale.

The UV-completeness of string theory arises from a simple but effective concept. Switching the description of the most fundamental objects from point-particles towards one-dimensional objects. These one-dimensional objects are so-called strings. The different vibrational modes of these strings then correspond to the particles of an effective quantum field theory. The main advantage of this idea is based on the world-sheet covered by a propagating string. This world-sheet can be parameterized by σ_0 and σ_1 for the propagation time and the spatial extension of the string, respectively. Hence, for one-loop amplitudes of closed strings, this world-sheet spreads out a two-torus. This two-torus is commonly parameterized by two parameters τ_1, τ_2 or in the complex plane by the modular parameter $\tau = \tau_1 + i\tau_2$. Equivalent tori are connected by modular transformations, i.e. the group $\mathrm{SL}(2,\mathbb{Z})$ [30], which is generated by,

$$\mathcal{T} : \tau \rightarrow \tau + 1, \qquad \mathcal{S} : \tau \rightarrow -\frac{1}{\tau} \ . \tag{2.1}$$

The string one-loop amplitude should not be affected by this re-parametrization of the world-sheet torus. Therefore, modular invariance is a necessary consistency condition of string theory [26–29]. For example the vacuum amplitude version of the one-loop diagram computes the Casimir energy density, i.e the 4D cosmological constant Λ. At the one-loop level Λ is proportional to

$$\Lambda \ \sim \ \int_{\mathcal{F}} \frac{\mathrm{d}^2\tau}{\tau_2^2}\, \mathcal{Z}_{\text{full}}(\tau,\bar{\tau}) \ . \tag{2.2}$$

The integrand $\mathcal{Z}_{\text{full}}$ and the measure $\frac{\mathrm{d}^2\tau}{\tau_2^2}$ are modular invariant by themselves. We will refer to the integrand as the full one-loop partition function. A particular consequence of these modular transformations is that they relate UV regimes to IR regimes of the dual channels. In detail, for the integration of the one-loop diagram over all inequivalent tori the modular transformations eq. (2.1) have to be taken into account, in order to avoid overcounting on equivalent tori. Hence, the modular integral is reduced to the fundamental domain of the world-sheet torus $\mathcal{F} = \{\tau = \tau_1 + \mathrm{i}\tau_2 \,|\, -\frac{1}{2} \leq \tau_1 < \frac{1}{2}, \tau_2 > 0, |\tau| > 1\}$. The modular parameter τ is now defined modulo modular transformations. This shows that the UV region $\tau \simeq 0$, which describes energetic strings propagation for a very short time, is excluded from the fundamental domain, i.e. it is mapped by modular transformations to $\tau \to -1/\tau$. Hence, the modular symmetry induced by the spatial extension of the string cuts off the potential UV divergence.

A controversial aspect of string theory is that in order to be a consistent theory it predicts additional space-like dimensions, i.e the space-time embedding functions $X^\mu(\sigma_0, \sigma_1)$ for the world-sheet to the target-space require more than the common four Minkowski dimensions. For details let us focus on the special type of string theory used in this thesis, i.e. *heterotic string theory*. Here, the bosonic and superstring theory are combined, i.e. the left-moving degrees of freedom are formalized by the bosonic string and hence have a 26D target-space, whereas the right-moving degrees of freedom relate to the supersymmetric string theory and propagate in 10D [25]. In order to make sense of this combination, the 16 extra dimensions of the bosonic string are compactified[1] on a torus $\mathbb{T}^{16}$. Invariance under the modular group generated by eq. (2.1) demands a self-dual lattice for the torus and the only lattices that can fulfill this condition and give rise to $\mathcal{N} = 1$ supersymmetry are the root lattices of $\mathrm{E}_8 \times \mathrm{E}_8$ and SO(32) [25]. This gives a very appealing 10D gauge theory as the starting point of the heterotic string theory approach. The combination of the bosonic and supersymmetric string is possible because heterotic string theory only contains closed strings. Their equation of motions are governed by wave equations. Thus the solutions are superpositions of left (L) and right (R) movers, i.e. it is possible to decompose the degrees of freedom to,

$$X^\mu(\sigma_0,\ \sigma_1) = X^\mu_{\mathrm{R}}(\sigma_0 - \sigma_1) + X^\mu_{\mathrm{L}}(\sigma_0 + \sigma_1)\ . \tag{2.3}$$

Moreover, closed strings are subject to boundary conditions along the spatial extension. Without compactification, the embedding function has to be (anti-)periodic with respect to one circulation along the string:

$$X^\mu(\sigma_0,\ \sigma_1 + \pi) = \quad X^\mu(\sigma_0,\ \sigma_1)\ , \tag{2.4}$$

$$\Psi^\mu_{\mathrm{R}}(\sigma_0 - (\sigma_1 + \pi)) = \pm\ \Psi^\mu_{\mathrm{R}}(\sigma_0 - \sigma_1)\ , \tag{2.5}$$

where world-sheet fermions Ψ^μ_{R} can have periodic (R: Ramond) or anti-periodic (NS: Neveu-Schwarz) boundary conditions. Hence, we can write physical states as tensor products of right and left-moving states. This is encoded in the bosonization formalism as,

$$|q\rangle_{\mathrm{R}} \otimes |p\rangle_{\mathrm{L}}\ , \tag{2.6}$$

which might be affected by oscillator excitations α_m, for the right, or $\tilde{\alpha}_m$ for the left-mover. They obey the creation and annihilation operator algebra. The subscript m labels the frequency of the vibrational mode. With the help of the GSO projection [36] the fermionic representations were combined to q, the right-moving momentum vector. Thus, $q \in$ SO(8) is in the vector (spinor) weight lattice for world-sheet fermions Ψ^μ_{R} in the NS (R) sector. This appears from the

---

[1] Technical details on torus compactification are given in section 2.2.1.

	$M_{3,1}$	M_6	$\mathrm{E}_8 \times \mathrm{E}_8$
$\vert q\rangle_{\mathrm{R}}$	X^μ_{R} Ψ^μ_{R}	X^μ_{R} Ψ^μ_{R}	
$\vert p\rangle_{\mathrm{L}}$	X^μ_{L}	X^μ_{L}	X^I_{L}
	$\mu = 0,\dots,3$	$\mu = 4,\dots,9$	$I = 1,\dots,16$

Table 2.1: *Left- and right-mover notation.*

light-cone gauge $X^\pm = \frac{1}{\sqrt{2}}(X^0 \pm X^1)$ which reduces the Lorentz SO(9, 1) symmetry to SO(8). Whereas p corresponds to the left-moving momenta and is quantized according to the $\mathrm{E}_8 \times \mathrm{E}_8$ or SO(32) lattice. For the quantized heterotic string in light-cone gauge one obtains mass equations for right- and left-movers,

$$\frac{M_{\mathrm{R}}^2}{4} = \frac{q^2}{2} + N + E_0^{\mathrm{R}} , \qquad \frac{M_{\mathrm{L}}^2}{4} = \frac{p^2}{2} + \widetilde{N} + E_0^{\mathrm{L}} . \tag{2.7}$$

The *zero point energy* E_0 for the 10D right and 26D left-movers, are given by $E_0^{\mathrm{R}} = -\frac{1}{2}$ and $E_0^{\mathrm{L}} = -1$, respectively. The oscillator number of the right-moving states is denoted by N whereas for the left-moving states $\widetilde{N}$ is used,

$$N = \sum_{n=1}^{\infty} \alpha_{-n} \cdot \alpha_n , \quad \widetilde{N} = \sum_{n=1}^{\infty} \widetilde{\alpha}_{-n} \cdot \widetilde{\alpha}_n , \tag{2.8}$$

counting the number of oscillators, i.e. $N \in \mathbb{N}$. Additionally, physical states have to obey the level matching condition $M_{\mathrm{R}}^2 = M_{\mathrm{L}}^2$ such that no point on the closed string is special.

Massive string states have masses of the order of the Planck mass and therefore are too heavy to contribute to the effective field theory of the (MS)SM. Hence, only massless string states are considered for the low energy spectrum. However, to describe the 4D world, 6D of the left and right moving strings require to be compactified. In the upcoming section, a specific technique for compactification, i.e. orbifold compactification, will be reviewed. Hereafter, the main purpose of this work is to improve the knowledge of how to obtain 4D phenomenologically promising models from a compactified heterotic string.

2.2 Heterotic orbifold compactification

In order to make contact to the observable world, it is necessary to hide the extra dimensions from experimental measurements at the current energy scales. In the case of the 10D $\mathrm{E}_8 \times \mathrm{E}_8$ heterotic string these extra dimensions aggregate to six spatial dimensions,

$$M_{10} = M_{3,1} \otimes M_6 . \tag{2.9}$$

In this thesis, the favoured method for compactification are 6D toroidal orbifolds[2] $\mathbb{O}$ [44, 45]. Those allow for a conformal field theory world-sheet formulation of string theory in the compact-

---

[2] As a remark let us mention that orbifolds can also be used in a field-theoretic context where they are used for GUT breaking [37–44] and can lead to discrete symmetries [22].

ified theory. Other scheme use supergravity approximations after compactification, see e.g. [46–56].

Starting from the definition of toroidal orbifolds and their geometrical action on the six spatial dimensions, it is crucial to extend the action to the gauge degrees of freedom in order to obey modular invariance under the modular transformations eq. (2.1). Hence, we will obtain a so-called gauge embedding (formalized in a gauge embedding matrix M) of the orbifold geometry $\mathbb{O}$ that acts on the sixteen gauge degrees of freedom X^I_L. Compared to the geometrical action of the orbifold $\mathbb{O}$, this gauge embedding comes with a vast ambiguity, and it appears that it is far less restricted than its geometrical companion. This gauge embedding will be the reason for the issue of the *heterotic orbifold landscape.*

2.2.1 Geometrical toroidal orbifolds

An orbifold is defined by identifying points of a manifold that differ by certain equivalence relations. In case of toroidal orbifolds, one starts by defining a linearly independent basis of vectors e_α that build a lattice Γ,

$$\Gamma = \left\{ \sum_\alpha^n n_\alpha e_\alpha ,\ n_\alpha \in \mathbb{Z} \right\} \ , \tag{2.10}$$

on the space $\mathbb{R}^n$. Points that differ by an integer combination of the lattice vectors are identified. Hence, any point in the plane can be mapped to a fundamental region. In case of modding out the lattice Γ, this fundamental domain describes a torus $\mathbb{T}^n = \mathbb{R}^n/\Gamma$ with metric

$$G_{\alpha\beta} = e_\alpha \cdot e_\beta \ . \tag{2.11}$$

Choices for the lattice Γ can be the root systems of simple Lie algebras, where one can utilize the simple roots as basis e_α. This first step is illustrated for the 2D case in fig. 2.1a, where the lattice was chosen to be spanned by the simple roots of SU(3) denoted by $\Gamma = \text{SU}(3)$.

For a more interesting structure, a new equivalence relation is introduced by the *point group* P, which is a discrete lattice automorphism of Γ, i.e. it maps the lattice to itself. Modding out P from the torus results in an orbifold $\mathbb{O}$,

$$\mathbb{O} \ = \ \mathbb{T}^n/P \ . \tag{2.12}$$

Combined, the action of the lattice translations and the point group rotations build S, the *space group* of the orbifold,

$$S \ = \ P \ltimes \Gamma \ . \tag{2.13}$$

An element of S is denoted by $g = (\theta, n_\alpha e_\alpha)$ with $\theta \in P$ and $n_\alpha e_\alpha \in \Gamma$ where summation over α is implied. This very illustrative simplification as semidirect product in eq. (2.13) holds, if there exist no roto-translations. A roto-translation combines the action of a point group element $\theta \in P$ with a simultaneous translation by a fractional lattice vector [57], i.e. $n_\alpha \in \mathbb{Q}$.

An element $g \in S$ acts on the a point $X \in \mathbb{R}^n$ as

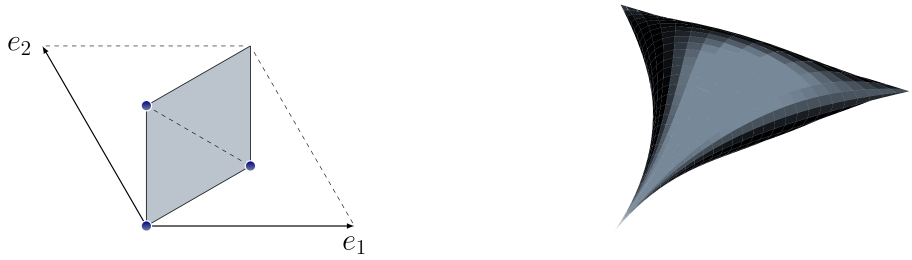
$$gX \ = \ \theta\, X + n_\alpha\, e_\alpha \ . \tag{2.14}$$

For space group elements $g', g \in S$ we find the following product rule,

$$g'\, g \ = \ \left(\theta', n'_\alpha e_\alpha\right)\, \left(\theta, n_\alpha e_\alpha\right) \ = \ \left(\theta'\, \theta, n'_\alpha e_\alpha + \theta'\, n_\alpha e_\alpha\right) \ . \tag{2.15}$$

Note, that the space group is in general non-Abelian, even if the point group is Abelian.

(a) *fundamental domain* $\mathbb{T}^2$ **(b)** *two-dimensional torus* $\mathbb{T}^2$



(c) *fundamental domain* $\mathbb{T}^2/\mathbb{Z}_3$ **(d)** *two-dimensional orbifold* $\mathbb{T}^2/\mathbb{Z}_3$

Figure 2.1: *Illustration of an orbifold compactification in 2D*

Crystallographic classification of 6D orbifolds

Fortunately, all point groups and all space groups relevant for toroidal orbifolds in 6D have been classified in [58, 59], as crystallographic actions in 6D. Given that $\theta \in P$ is an automorphism of the torus lattice one finds,

$$\theta\, e \;=\; e\,\widehat{\theta} \quad \text{with} \quad \widehat{\theta} \in \mathrm{GL}(6;\mathbb{Z})\ , \tag{2.16}$$

where for convenience we introduce the vielbein e as matrix with the columns given by the lattice vectors e_α. The hatted notation of $\widehat{\theta}$ indicates that it is an integral matrix, i.e. the so-called twist in the lattice basis. Consequently, the group in the lattice basis representation is denoted by $\widehat{P}$. In this classification, the geometrical space groups are sorted according to

- $\mathbb{Q}$-classes: The geometrical point groups P in terms of the lattice basis $\widehat{P}$, see eq. (2.16).
- $\mathbb{Z}$-classes: Possible lattices Γ for a given $\widehat{P}$.
- Affine-classes: The possible roto-translations of the lattices.

It turns out that in 6D, there are 7 103 $\mathbb{Q}$-classes, which are a subset of the 85 308 $\mathbb{Z}$-classes, and a total of 28 927 915 affine-classes, i.e. the affine-classes label all geometrically inequivalent toroidal orbifolds in 6D. The collection of this data is available in the CARAT package [58]. However, we will see later that only a tiny fraction of these possibilities can lead to meaningful string theory compactifications when physical considerations, e.g. $\mathcal{N} = 1$ target-space supersymmetry, are taken into account.

Details on the point group

In the physics literature of orbifold compactifications the automorphism on the lattice Γ is called point group P. However, the point group P is more precisely given as the six-dimensional representation of an abstract finite group H,

$$\begin{aligned} D_{\mathbf{v}} : H &\to P \subset \mathrm{SO}(6) \\ a &\mapsto D_{\mathbf{v}}(a) = \theta \ , \end{aligned} \tag{2.17}$$

where different representations, i.e. $\mathbb{Q}$-classes, for a single abstract group H can exist. The elements θ of P then are called rotations. The smallest positive integer,

$$N_\theta = \min\left(\{\ N \in \mathbb{N}\backslash\{0\} \mid \theta^N = \mathbb{1}\ \}\right) \ , \tag{2.18}$$

i.e. $\theta^{N_\theta} = \mathbb{1}$, is called the order of the point group element θ. For each element $\theta \in \mathrm{SO}(6)$ separately it is possible to block-diagonalize this element by a basis transformation. Then, this element acts as a rotation in the three orthogonal planes, that are given by (X^1, X^2), (X^3, X^4) and (X^5, X^6). Rewriting it in complex coordinates $Z^a = X^{2a-1} + \mathrm{i} X^{2a}$ for $a = 1, 2, 3$ the rotation is represented by

$$\theta = \begin{pmatrix} e^{2\pi\mathrm{i}\, v_g^1} & 0 & 0 \\ 0 & e^{2\pi\mathrm{i}\, v_g^2} & 0 \\ 0 & 0 & e^{2\pi\mathrm{i}\, v_g^3} \end{pmatrix} . \tag{2.19}$$

One obtains the *local twist vector* $v_g = \left(0, v_g^1, v_g^2, v_g^3\right)$, which parameterizes the representation. Note, we introduced an additional entry $v_g^0 = 0$ for later use. This parametrization will be frequently used in this thesis.

Even though some considerations on non-Abelian orbifolds will be discussed in the context of the cosmological constant in chapter 3, the explicit construction of non-Abelian orbifolds is not entirely understood (see [60] for an example). Hence, for explicit constructions we will use Abelian point groups. In the case of Abelian point groups, the abstract finite groups H are limited to the cyclic groups $\mathbb{Z}_{N_1}$ and $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$ [61, 62]. Thus, P consists of rotations generated by θ and ω, such that $\theta^{N_1} = \mathbb{1}$ and $\omega^{N_2} = \mathbb{1}$. For the Abelian case the block-diagonalization from eq. (2.19) can be performed on all elements P simultaneously and the local twist vector for the element $\theta^{k_1}\omega^{k_2}$ reads $v_g = k_1 v_1 + k_2 v_2$. Hence, it simplifies towards the *twist vectors* v_1 and v_2 associated to the generators of $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$, respectively. To label the different orbifold geometries, we use the concise notation of [57]. For example, the orbifold of particular interest in this thesis is $\mathbb{Z}_6$-II (1,1), where $\mathbb{Z}_6$ defines the cyclic group H and 'II' refers to the second type of embedding this action in 6D eq. (2.17). In detail, this part encodes the $\mathbb{Q}$-classes and hence the point group $P = \mathbb{Z}_6$-II is parametrized by the second twist vector, i.e. $v = (1/6, 1/3, 1/2)$. The first number of (1,1) enumerates the possible lattices ($\mathbb{Z}$-classes) and the second is the enumeration of possible roto-translations (affine-classes). A value of one in the second slot stands for no roto-translations.

Fixed points

The interesting structure of orbifolds arises most prominently on the special coordinate singularities, so-called fixed points. These special corners of the manifold still allow for a consistent definition of string propagation [44], and moreover introduce a particular behavior of the strings localized at these curvature singularities (see section 2.4.2). If the point group action P is said to be not freely acting on the manifold, there exist *fixed points* z_f defined by,

$$g z_f = \theta z_f + n_\alpha e_\alpha = z_f \ , \tag{2.20}$$

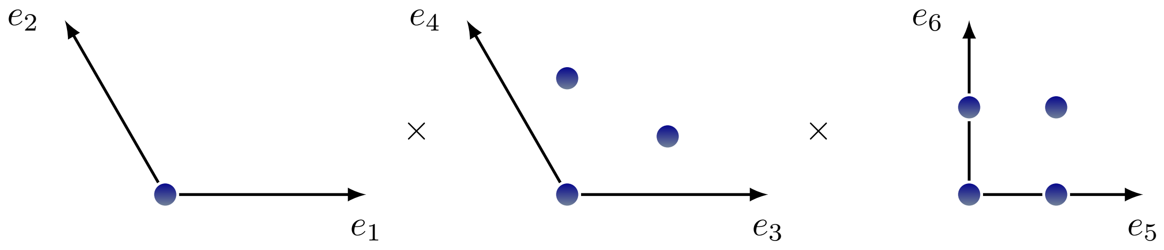


Figure 2.2: *First twisted sector* $T_{(1)}$ *of the* $\mathbb{Z}_6$-II $(1,1)$ *orbifold. The torus lattice is* $\Gamma = \mathrm{SU}(3) \times \mathrm{SU}(3) \times (\mathrm{SU}(2))^2$ *and the twist vector* $v = (1/6, 1/3, 1/2)$. *The dots in blue, indicate the position of the fixed points in the projection of each sub-torus, i.e.* $\mathbb{T}^2/\mathbb{Z}_6 \times \mathbb{T}^2/\mathbb{Z}_3 \times \mathbb{T}^2/\mathbb{Z}_2$. *One obtains* $1 \times 3 \times 4 = 12$ *fixed points in the first twisted sector.*

with $g \in S$. Hence, each fixed point is associated to the space group element g that maps the coordinate to itself. These special points are illustrated in fig. 2.1d as the corners of the orbifolded space. Concretely, this will become important in the context of the $\mathbb{Z}_6$-II (1,1) orbifold. Therefore, the fixed point structure is illustrated in fig. 2.2 in detail.

At this point, let us investigate under which condition two elements describe the same action on the orbifold. Assume that $g_1 z_{f_1} = z_{f_1}$ and $g_2 z_{f_2} = z_{f_2}$. These two fixed points are certainly equivalent if they are related by another element of the space group $h \in S$, i.e. $h z_{f_2} = z_{f_1}$. This situation is related to

$$g_1 z_{f_1} = h z_{f_2} = h g_2 h^{-1} z_{f_1} \ , \tag{2.21}$$

and hence for equivalent fixed points the associated elements lie in the same conjugacy class

$$[g] = \left\{ h\, g\, h^{-1} \;\middle|\; \forall h \in S \right\} \ . \tag{2.22}$$

Then g is a representative of the whole set $[g]$.

2.2.2 Gauge embedding

In the last section, we have considered the geometrical action of the orbifold, i.e. the action on the six spatial dimensions. However, remember from section 2.1 that there are also internal gauge degrees of freedom. It turns out that the orbifold action intertwines the modular invariance conditions of the space-time and gauge group $\mathrm{E}_8 \times \mathrm{E}_8$ part of the heterotic string [30]. Consequently, they are no longer modular invariant by themselves. In order to make the whole theory modular invariant again, the geometrical orbifold action has to be embedded into the gauge degrees of freedom (d.o.f.). An additional benefit of this embedding is that it allows for breaking the gauge group of $\mathrm{E}_8 \times \mathrm{E}_8$. From the geometrical action of the space group on the six spatial dimensions $\mu = 4, \ldots, 9$, one can infer conditions on the embedding equivalents of θ and e_α. Acting with a generic element $g = (\theta^k, n_\alpha e_\alpha)$ we obtain,

$$\text{geometrical:} \quad X^\mu \to (\theta^k X)^\mu + n_\alpha e_\alpha^\mu \ , \tag{2.23}$$

$$\text{gauge d.o.f.:} \quad X_{\mathrm{L}}^I \to X_{\mathrm{L}}^I + kV^I + n_\alpha W_\alpha^I \ , \tag{2.24}$$

where $I = 1, \ldots, 16$ and the embedding is realized by a shift embedding [44]. This is possible since in general the action of the space group on the gauge degrees of freedom corresponds to an inner automorphism. V is the so-called *shift vector* and encodes the action of the point group

vector	order N_k	additional constraint
V_1	6	
V_2	1	not present, i.e. $V_2 = (0^{16})$
W_1	1	$W_1 = (0^{16})$
W_2	1	$W_2 = (0^{16})$
W_3	3	$W_3 = W_4$
W_4	3	
W_5	2	
W_6	2	

Table 2.2: *Geometrical constraints for the gauge embedding, i.e shift vectors and Wilson lines, in the case of the* $\mathbb{Z}_6$-II (1,1) *orbifold geometry.*

generator θ, while W_α are the *Wilson lines* [63] which correspond to the lattice shifts e_α. From the action of the geometrical counterparts, we can infer four kinds of consistency conditions for the gauge embedding of the space group. For the case of a point group build on $\mathbb{Z}_{N_1}$ they can be summarized by,

$$\begin{array}{lll} \theta^{N_1} = \mathbb{1} & \Rightarrow & N_1 V_1 \in \Gamma_{\mathrm{E}_8\times\mathrm{E}_8}\ , \\ \theta^{N_\alpha-1}e_\alpha + \theta^{N_\alpha-2}e_\alpha \cdots + e_\alpha = 0 & \Rightarrow & N_\alpha \cdot W_\alpha \in \Gamma_{\mathrm{E}_8\times\mathrm{E}_8}\ , \\ \theta e_\alpha + e_\alpha & \Rightarrow & 2W_\alpha\ , \\ \theta(n_\alpha e_\alpha) = n_\beta e_\beta & \Rightarrow & n_\alpha W_\alpha = n_\beta W_\beta\ . \end{array} \tag{2.25}$$

These can be understood as follows: Acting N_1 times with $g = (\theta, 0)$ gives a trivial action in the geometrical embedding, hence the shift vector V_1 that corresponds to θ has to act trivially, i.e. as a lattice vector of $\Gamma_{\mathrm{E}_8\times\mathrm{E}_8}$. Similarly, the space group element $g = (\theta, e_\alpha)^{N_\alpha} = (\theta^{N_\alpha}, 0)$ implies the order N_α of the Wilson line W_α. Acting two times with $g = (\theta, e_\alpha)$ shows that θe_α corresponds to W_α as does e_α. This implication has consequences on the last line of eq. (2.25), in detail this correspondence forces certain Wilson lines to be identical. For an example that illustrates the intuitively given explanation above, see table 2.2, where the geometrical conditions for $\mathbb{Z}_6$-II (1,1) are summarized. Let us stress the fact that the Wilson lines $W_3 = W_4$ are related due to the fact that the $\mathbb{Z}_3$ action in the second tori relates $\theta e_3 = e_4$, see fig. 2.2.

Parametrization $\mathbb{Z}^{128}$ for $\mathrm{E}_8 \times \mathrm{E}_8$

For later convenience in the data science approach to the landscape, chapter 6, let us introduce a practical parametrization that was first developed in [21]. Therefore, we start by obtaining that in $\mathrm{E}_8 \times \mathrm{E}_8$ the sixteen-dimensional shift vector V_1 or Wilson line W_α consists of two eight-dimensional vectors $V_1^{(1)}$ and $V_1^{(2)}$ or $W_\alpha^{(1)}$ and $W_\alpha^{(2)}$, which act on the first and second E_8 factor, respectively. Then a gauge embedding is determined by a gauge embedding matrix M,

$$M \;=\; \begin{pmatrix} \boxed{\;\text{———}\; V_1^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; V_1^{(2)} \;\text{———}\;} \\ \boxed{\;\text{———}\; V_2^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; V_2^{(2)} \;\text{———}\;} \\[1ex] \boxed{\;\text{———}\; W_1^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; W_1^{(2)} \;\text{———}\;} \\ \boxed{\;\text{———}\; W_2^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; W_2^{(2)} \;\text{———}\;} \\ \boxed{\;\text{———}\; W_3^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; W_3^{(2)} \;\text{———}\;} \\ \boxed{\;\text{———}\; W_4^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; W_4^{(2)} \;\text{———}\;} \\ \boxed{\;\text{———}\; W_5^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; W_5^{(2)} \;\text{———}\;} \\ \boxed{\;\text{———}\; W_6^{(1)} \;\text{———}\;} & \boxed{\;\text{———}\; W_6^{(2)} \;\text{———}\;} \end{pmatrix} . \tag{2.26}$$

The sixteen-dimensional shift vectors and Wilson lines remain in $\Gamma_{\mathrm{E}_8 \times \mathrm{E}_8}$, which provides a basis in terms of the simple roots α_I of $\mathrm{E}_8 \times \mathrm{E}_8$. Then, one can expand the sixteen-dimensional vectors in terms of this basis as,

$$M_k \;=\; \frac{1}{N_k} \sum_{I=1}^{16} c_{k\,I}\; \alpha_I \;. \tag{2.27}$$

Here, N_k defines the order (see eq. (2.25) and eq. (2.18)) of the shift vector or Wilson line and $c_{k\,I} \in \mathbb{Z}$ for $k = 1, \ldots, 8$ and $I = 1, \ldots, 16$ are integers. While this would be the canonical approach, it will turn out in chapter 6 that some ambiguities, i.e. symmetries of the landscape, are challenging to track in this basis.[3] In fact, it will be beneficial to expand the sixteen-dimensional vectors in terms of the dual simple roots α_I^*, $I = 1, \ldots, 16$, which are defined as $\alpha_I^* \cdot \alpha_J = \delta_{IJ}$. We obtain,

$$M_k \;=\; \frac{1}{N_k} \sum_{I=1}^{16} d_{k\,I}\; \alpha_I^* \;. \tag{2.28}$$

Here $d_{k\,I} \in \mathbb{Z}$ for $k = 1, \ldots, 8$ and $I = 1, \ldots, 16$ are integers and parameterize the gauge embedding. However, the interpretation of these integers is very different from the c_{kI}, i.e. the $d_{k\,I}$ are proportional to the Dynkin labels. Nonetheless, both choices give a basis of the (self-dual) root lattice $\Gamma_{\mathrm{E}_8 \times \mathrm{E}_8}$ of $\mathrm{E}_8 \times \mathrm{E}_8$. A particular choice for the simple and dual roots is explicitly given in appendix D.1. Consequently, a gauge embedding matrix M corresponds to a point in $d \in \mathbb{Z}^{128}$ since M has $8 \times 16 = 128$ $d_{k\,I}$-coefficients. Note that the constructions, eq. (2.27) and eq. (2.28), inherently ensure the correct order of the vectors M_k derived in eq. (2.25). In detail, we have

$$N_k\, M_k \;\in\; \Gamma_{\mathrm{E}_8 \times \mathrm{E}_8} \quad \forall\, k = 1, \ldots, 8 \;. \tag{2.29}$$

---

[3] Note this basis is useful to have control over the addition of lattice vectors. However, this transformation is not an exact symmetry in general.

The other geometrical conditions are induced directly, i.e. for vanishing Wilson lines the corresponding d_{kI} are set to zero and for related Wilson lines the d_{kI}-coefficients are equal during model construction. All 138 orbifold geometries with Abelian point groups and $\mathcal{N} = 1$ supersymmetry [57] are stored in so-called geometry-files where the geometrical conditions are stated explicitly. These geometry-files can be found as ancillary files to [64].

2.3 Strings on orbifolds

After having defined the general set-up for orbifold compactifications, we can now get into the details of strings propagating on these special geometries in the extra dimensions. In heterotic string theory, the 6D toroidal orbifold geometry extends the closed string boundary conditions eq. (2.4) for the corresponding coordinate fields $X = (X^\mu)$, with $\mu = 4, \ldots, 9$. In detail, strings on toroidal orbifolds can close up to the action of an element $g = (\theta, n_\alpha e_\alpha) \in S$,

$$X(\sigma_0, \sigma_1 + \pi) \;=\; gX(\sigma_0, \sigma_1) \;=\; \theta\, X(\sigma_0, \sigma_1) + n_\alpha e_\alpha \;. \tag{2.30}$$

In such a case, the element g is associated to the string configuration as in the case of fixed points. Then, g is called the constructing element of the string. The specific form of the element g gives rise to so-called

- twisted strings: The corresponding boundary condition is connected to an element g with $\theta \neq \mathbb{1}$.
- untwisted strings: The corresponding boundary condition is connected to an element g with $\theta = \mathbb{1}$.

Untwisted strings are said to “live” in the untwisted sector U while the twisted strings “live” in the twisted sectors $T_{(k)}$ associated to θ^k. As they correspond to the fixed point structure they are organized in the different twisted sectors k for $v_g = kv$. For an illustration of the twisted sectors in case of $\mathbb{Z}_6$-II (1,1) see fig. D.1. These concepts directly extend to $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$ orbifold geometries, i.e. the twisted sectors are labeled according to the local shift vector $T_{(k_1,k_2)}$. However, as discussed in the gauge embedding section 2.2.2 the action of the orbifold also extends to the gauge degrees of freedom. Hence, associated to each constructing element $g = (\theta^{k_1}\omega^{k_2}, n_\alpha e_\alpha) \in S$ there exist so-called local shift vectors V_g associated to fixed points. This implies the *local shift*

$$V_g \;=\; k_1\, V_1 + k_2\, V_2 + n_\alpha\, W_\alpha \tag{2.31}$$

of the gauge degrees of freedom.

Finally, the states are in general not inequivalent for each element $h \in S$. Assume that $[g, h] = 0$ and g is the constructing element of $X(\sigma_0, \sigma_1)$,

$$X(\sigma_0, \sigma_1 + \pi) = gX(\sigma_0, \sigma_1) \;, \tag{2.32}$$

it then follows that h has to act trivially on the string state X since

$$h(gX(\sigma_0, \sigma_1)) = g(hX(\sigma_0, \sigma_1)) \;, \tag{2.33}$$

and hence X and hX are associated to the same constructing element. Thus, they describe the same physical state. This is related to the centralizer C_g of g,

$$\mathrm{C}_g = \{\; h \in S \;\big|\; [g, h] = 0 \;\}. \tag{2.34}$$

The trivial action under all elements of the centralizer C_g will lead to projection conditions for the string states.

2.3.1 Transformation under the space group

From the vertex operator representation one can infer that a string state acquires a phase on the orbifold. All the transformations can be decomposed into phases acquired under an element $h \in S$,

$$\begin{aligned} |q\rangle_{\mathrm{R}} &\mapsto e^{-2\pi\mathrm{i}\; q\cdot v_h}\, |q\rangle_{\mathrm{R}} && (2.35)\\ |p\rangle_{\mathrm{L}} &\mapsto e^{2\pi\mathrm{i}\; p\cdot V_h}\, |p\rangle_{\mathrm{L}} && (2.36)\\ \widetilde{\alpha}^{a}_{-1} &\mapsto e^{2\pi\mathrm{i} v^a_h}\, \widetilde{\alpha}^{a}_{-1} && (2.37)\\ \widetilde{\alpha}^{\bar{a}}_{-1} &\mapsto e^{-2\pi\mathrm{i} v^a_h}\, \widetilde{\alpha}^{\bar{a}}_{-1}\ . && (2.38) \end{aligned}$$

Note, the oscillator excitations $\widetilde{\alpha}^{\mu}_{-1}$ that act in the compactified directions $\mu = 4, \ldots, 9$ are here represented in the complex directions $a = 1, 2, 3$. Moreover, we introduced the holomorphic a and antiholomorphic $\bar{a}$ index where the antiholomorphic transforms in the complex conjugate representation. Combining these building blocks, we get the phase,

$$\Phi = e^{2\pi\mathrm{i}(p\cdot V_h - q\cdot v_h) + (\widetilde{N}_0 - \widetilde{N}^*_0)v_h}\quad \Phi_{\mathrm{vac}} \qquad \text{with} \qquad \Phi_{\mathrm{vac}} = e^{2\pi\mathrm{i}(V_g\cdot V_h - v_g\cdot v_h)}\ , \tag{2.39}$$

where the vacuum phase Φ_{vac} stems from geometrical considerations and details can be found in [65]. The entries of $\widetilde{N}_0$ and $\widetilde{N}^*_0$ are integer oscillator numbers, counting the number of excitations in the holomorphic and antiholomorphic directions, e.g. $\widetilde{N}^*_0 = (0, \widetilde{N}^{\bar{a}=1}_0, \widetilde{N}^{\bar{a}=2}_0, \widetilde{N}^{\bar{a}=3}_0)$. In the upcoming sections we will refer to these transformations and especially in the twisted sector some generalizations are introduced.

2.3.2 Physical constraints on the orbifold compactification

The orbifold compactification discussion so far did not include all consistency conditions or phenomenological considerations. In this subsection, supersymmetry and modular invariance will inject strong constraints on the possible orbifold compactifications in heterotic string theory.

Supersymmetry $\boldsymbol{\mathcal{N} = 1}$

The amount of 4D supersymmetry after orbifold compactification is determined by the number of spinors that are invariant under the action of P. Hence, the task is to embed $P \subset \mathrm{SO}(6)$ into spinor-space Spin(6). As $\mathrm{Spin}(6) \cong \mathrm{SU}(4)$ the representation acts as a four-dimensional representation **4** of SU(4). The spinor representations

$$\begin{aligned} D_{\mathbf{s}}:\quad P \subset \mathrm{SO}(6) &\to P_s \subset \mathrm{Spin}(6) \\ \theta \qquad &\mapsto D_{\mathbf{s}}(\theta)\ , \end{aligned} \tag{2.40}$$

can be constructed using the explicit spinor basis eq. (A.16). Then $D_{\mathbf{s}}(\theta)$, associated with the element g, is given in the following form,

$$D_{\mathbf{s}}(\theta) = e^{2\pi\mathrm{i}\, v^1_g\, \frac{1}{2}\sigma_3} \otimes e^{2\pi\mathrm{i}\, v^2_g\, \frac{1}{2}\sigma_3} \otimes e^{2\pi\mathrm{i}\, v^3_g\, \frac{1}{2}\sigma_3}\ . \tag{2.41}$$

We find that the spinor representation is parameterized by the local twist vector v_g as in case of the geometrical point group eq. (2.19). To determine the amount of 4D supersymmetry it is sufficient to compute the decomposition of **4**. In detail, the amount of invariant spinors corresponds to the number of trivial singlets contained in the decomposition. This relates the holonomy group of a 6D torus orbifold directly to the amount of supersymmetry. If $P \subset \mathrm{SU}(3)$, i.e. the holonomy group is SU(3), this ensures that the four-dimensional representation of SU(4) decomposes in a three dimensional, maybe reducible, and a one-dimensional representation:

$$\mathbf{4} = \mathbf{3} \oplus \mathbf{1}\ . \tag{2.42}$$

If $P = \mathrm{SU}(3)$ exactly[4] the only one-dimensional representation is the trivial representation, i.e. we have a covariantly constant spinor. However, since P is a discrete group embedded in SU(3) we have different choices for the one-dimensional representation, i.e. we can take non-trivial one-dimensional representations. An example will be discussed around table 3.3. Hence, one has some freedom and has to make a choice. To fix the condition such that a trivial singlet appears, one can take the viewpoint of the orbifold phase acting on the right-moving momentum q that is associated to the spinors, and finds

$$\theta : |q\rangle_{\mathrm{R}} \mapsto e^{-2\pi\mathrm{i}\ q\cdot v}\ |q\rangle_{\mathrm{R}}\ , \tag{2.43}$$

where we have $q = \left(\underline{\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}}\right)$ with even number of plus signs. The underline of the vector indicates that all permutations are considered. Hence, one obtains that the following condition

$$q \cdot v = 0 \mod 1\ , \tag{2.44}$$

summarizes the projection condition on the spinors associated to q. While, several equations relate to the above condition, for the notation in [57], the weight $q = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ is chosen to survive the projection. Together, with the simplified special solution to vanish exactly,

$$v_1 + v_2 + v_3 = 0\ , \tag{2.45}$$

instead of modulo integers. If the holonomy group is SU(3) this choice ensures a trivial phase of the chosen spinor and hence $\mathcal{N} = 1$. Let us stress that all 138 orbifold geometries with Abelian point groups and $\mathcal{N} = 1$ supersymmetry with the special twist vector choice eq. (2.45) have been classified in [57].

Modular invariance

The amount of supersymmetry can be directly related to the geometrical action of the orbifold, which comes from the fact that solely the right-movers are based on a supersymmetric string. Moreover, the right-movers are not involved in the gauge degrees of freedom. However, the orbifolding of the geometrical part breaks the modular invariance of the theory and demands that certain constraints on the gauge embedding M repair the situation. In order to obtain a consistent string compactification we have to impose conditions on the gauge embedding matrix M. These conditions, developed in [66], read

$$\begin{aligned}
N_1\ \left(V_1^2 - v_1^2\right) &= 0 \mod 2\ ,\\
N_2\ \left(V_2^2 - v_2^2\right) &= 0 \mod 2\ ,\\
\gcd(N_1, N_2)\ \left(V_1 \cdot V_2 - v_1 \cdot v_2\right) &= 0 \mod 2\ ,\\
\gcd(N_{\alpha+2}, N_1)\ \left(W_\alpha \cdot V_1\right) &= 0 \mod 2\ ,\\
\gcd(N_{\alpha+2}, N_2)\ \left(W_\alpha \cdot V_2\right) &= 0 \mod 2\ ,\\
N_{\alpha+2}\ \left(W_\alpha^2\right) &= 0 \mod 2\ ,\\
\gcd(N_{\alpha+2}, N_{\beta+2})\ \left(W_\alpha \cdot W_\beta\right) &= 0 \mod 2 \quad (\alpha \neq \beta)\ ,
\end{aligned} \tag{2.46}$$

---

[4] This is the case for Calabi-Yau constructions.

for $\alpha, \beta = 1, \ldots, 6$ and where v_1 and v_2 denote the twist vectors of $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$. In addition, gcd in eq. (2.46) denotes the greatest common divisor. Note that, included in these conditions are consistency constraints on the vacuum phase Φ_{vac} developed in [66]. These Diophantine conditions are very restrictive and already forbid a huge fraction of points in the space $\mathbb{Z}^{128}$. Illustrations and a detailed discussion of this behavior is given in chapter 4 where a statistical approach to the solutions of eq. (2.46) is discussed.

2.4 Spectrum computation

Massive strings have masses of the order of the Planck scale and therefore are too heavy to contribute to the effective field theory of the (MS)SM. Hence, only massless string states are considered for the low energy spectrum and limit the solutions of the mass equations eq. (2.7). Moreover, we have to consider the orbifold projections by the acquired phases from eq. (2.35). These conditions are investigated for the untwisted sector or 'bulk' and the twisted sector in the following subsections.

2.4.1 Untwisted sector

Remember that untwisted strings correspond to those strings that close already under a constructing element with $\theta = \mathbb{1}$. Hence, they have the same mass equations as the uncompactified string.

Mass equation

For the low energy spectrum only massless strings can contribute[5], i.e. $M_L = 0 = M_R$

$$0 \;=\; \frac{q^2}{2} + N - \frac{1}{2}\,, \qquad 0 \;=\; \frac{p^2}{2} + \widetilde{N} - 1\;. \tag{2.47}$$

The only solutions are given for $q^2 = 1$ and $N = 0$ in combination with the solutions for the left-mover that are $p^2 = 0$ and $\widetilde{N} = 1$ or $p^2 = 2$ and $\widetilde{N} = 0$.

Orbifold projection

Additional to the mass conditions, the states have to be invariant under the orbifold action,

$$|q\rangle_{\mathrm{R}} \otimes |p\rangle_{\mathrm{L}} \quad\to\quad e^{2\pi \mathrm{i}(p\cdot V_h - q\cdot v_h)}\; |q\rangle_{\mathrm{R}} \otimes |p\rangle_{\mathrm{L}}\;. \tag{2.48}$$

For the untwisted sector we have the trivial constructing element with $\theta = \mathbb{1}$. Since the centralizer of this trivial constructing element is the whole space group S, this transformation has to act trivially for all elements of S,

$$p \cdot V_h - q \cdot v_h \;=\; 0 \mod 1 \quad \forall h \in S\;. \tag{2.49}$$

Let us take a closer look on the case for $p^2 = 2$ and $\widetilde{N} = 0$ since this will become of importance during this thesis. This case corresponds to the breaking of the $\mathrm{E}_8 \times \mathrm{E}_8$ gauge group and results in the 4D charged fields.

---

[5] Kaluza-Klein and Winding numbers are zero for massless strings at a generic point in moduli space.

Charged gauge bosons

The first case is given for $q \cdot v_h = 0$. As for a special choice of $\mathcal{N} = 1$ condition we have a surviving spinor that is invariant under the orbifold eq. (2.43). The common choice is positive signs in the mod condition eq. (2.45). Hence, $q = (\pm 1, 0, 0, 0)$ or $q = \pm \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ lead to this case. Then, only states that are tensored to roots p of $\mathrm{E}_8 \times \mathrm{E}_8$ that fulfill,

$$p \cdot M_k \; = \; 0 \quad \text{mod } 1 \quad \forall k = 1, \ldots, 8 \tag{2.50}$$

can survive the orbifold action. Note that we have rewritten the condition eq. (2.49) for each V_h with $h \in S$ to its irreducible conditions.

Untwisted charged matter U-sector

Contrary to the case above, the second part of the orbifold phase is non-vanishing for the values $q = \left(0, \underline{\pm 1, 0, 0}\right)$ or $q = \pm\left(-\frac{1}{2}, \underline{-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}}\right)$ and we find different conditions for the particular q that correspond to the so called U-sectors. The conditions read,

$$p \cdot M_1 - q_{(a)} \cdot v_1 \; = \; 0 \quad \text{mod } 1 \tag{2.51}$$

$$p \cdot M_2 - q_{(a)} \cdot v_2 \; = \; 0 \quad \text{mod } 1 \tag{2.52}$$

$$p \cdot M_k \; = \; 0 \quad \text{mod } 1 \quad \text{for } k = 3, \ldots, 8 \tag{2.53}$$

for the unbroken roots p of E_8. The $q_{(1)} = (0, -1, 0, 0)$, $q_{(2)} = (0, 0, -1, 0)$ and $q_{(3)} = (0, 0, 0, -1)$ are then associated to the U_a-sectors with $a = 1, 2, 3$. They correspond to the three directions of the internal vector-boson index of the ten-dimensional $\mathrm{E}_8 \times \mathrm{E}_8$ gauge bosons and have different R-charges [31].

2.4.2 Twisted sector

For the case of a non-trivial rotation as part of the constructing element the situation becomes much more involved. As zero modes of twisted sectors $T_{(k_1,k_2)}$ are associated to fixed points and hence to the constructing element $g = (\theta^{k_1}\omega^{k_2}, n_\alpha e_\alpha) \in S$ related to them. In detail, this effects the mode expansion of the left and right-movers such that the center of mass of the twisted strings is localized at the fixed points. Additionally, the appearance of twisted oscillators with fractional frequency leads to a shift in the zero point energy [30] by

$$\Delta E_g \; = \; \frac{1}{4} \sum_{a=1}^{3} \omega_g^a \left(1 - \omega_g^a\right) , \tag{2.54}$$

where one can define $\omega_g^a \; = \; (v_g)^a \;$ mod 1 and $\bar{\omega}_g^a \; = \; -\,(v_g)^a \;$ mod 1. Note, that the shift in the zero point energy is dependent on the local shift vector and hence on the particular twisted sector. This indicates the subscript label g of the constructing element.

Mass equations

The twisted oscillators also lead to a shift of the right moving momenta q such that the mass equation reads together with eq. (2.54),

$$\frac{(q + v_g)^2}{2} - \frac{1}{2} + \Delta E_g \; = \; 0 \tag{2.55}$$

Similarly, the right handed momentum p is shifted by the action of the corresponding local shift vector V_g. Since the shift in the zero point energy eq. (2.54) is the same for the left-movers, one obtains

$$\frac{(p+V_g)^2}{2} + \widetilde{N}_{(g)} - 1 + \Delta E_g = 0 \ , \tag{2.56}$$

with the fractional number operator $\widetilde{N}_{(g)}$ according to the twisted sectors.

Orbifold projection

As the orbifold phase for commuting elements, i.e. centralizer elements, relate identical physical states, the combined phase from section 2.3.1 has to act trivially and vanish.[6] This results in the orbifold projection condition for twisted states as,

$$p_{sh} \cdot V_h \;-\; q_{sh} \cdot v_h + (\widetilde{N}_g - \widetilde{N}_g^*) \cdot v_h - \tfrac{1}{2}(V_g \cdot V_h \;-\; v_g \cdot v_h) \overset{!}{=} 0 \mod 1 \tag{2.57}$$

where we introduced the notation $q_{sh} = q + v_g$ and $p_{sh} = p + V_g$. Analogous to eq. (2.39) $\widetilde{N}_g^a$ and $\widetilde{N}_g^{*\,a}$, $a = 0, \ldots, 3$, are the integers counting the number of oscillators and are the entries of $\widetilde{N}_g$ and $\widetilde{N}_g^*$. They are given by splitting the eigenvalues of the fractional number operator $\widetilde{N}_{(g)}$ according to

$$\widetilde{N}_{(g)} \;=\; \sum_{a=1}^{3} \omega_g^a \widetilde{N}_g^a + \bar{\omega}_g^a \widetilde{N}_g^{*\,a} \ . \tag{2.58}$$

2.4.3 Local GUTs

So far, we directly considered the full orbifold action on the physical states in order to get the 4D perspective of our massless states. An approach that contains more information about the inner structure of the orbifold is given by a local inspection. Concretely, it is possible to consider the matter content on a single fixed point and inspect the local gauge group at this fixed point. This leads to the concept of so-called *local GUTs* [67–77], where the fixed points are equipped with a GUT (grand unified theory). If there is no further orbifold projection for the states at the fixed points, then by considering the complete orbifold, the gauge group of the local GUT branches group theoretically in representations of the 4D gauge group. This particular situation can be found in the first twisted sector of $\mathbb{Z}_N$ orbifolds. First of all, there are no roto-translations in the space group. Those can lead to a non-local breaking of the local GUTs. Moreover, in the case of the first twisted sector $T_{(1)}$ in $\mathbb{Z}_{N_1}$ orbifolds the special situation arises that the centralizer of g is trivial, i.e. only powers of $g^m \in C_g$ with $m \in \mathbb{N}$ are included. Hence they correspond to the same fixed point. This would not be the fact for higher-order twisted sectors as there exist fixed tori. They insert an ambiguity on the constructing element g, which is defined only up to lattice vectors in the fixed torus. Then, additional elements appear in the centralizer of g. However, the special situation for the centralizer in $T_{(1)}$ of $\mathbb{Z}_N$ orbifolds ensures the local GUT breaking and the invariance of the local matter states under orbifold projection phases [32].

For example there exist shift vectors such that a SO(10) gauge group with matter content in **16** survives on a fixed point [78]. For the desired SM gauge group SU(3) × SU(2) × U$(1)_Y$, this breaks into one generation of quarks and leptons, including a right-handed neutrino

$$\mathbf{16} = (\mathbf{3},\mathbf{2})_{1/6} + (\overline{\mathbf{3}},\mathbf{1})_{-2/3} + (\overline{\mathbf{3}},\mathbf{1})_{1/3} + (\mathbf{1},\mathbf{2})_{-1/2} + (\mathbf{1},\mathbf{1})_{1} + (\mathbf{1},\mathbf{1})_{0} \ . \tag{2.59}$$

---

[6] Note, in principle also the case of $[h, g] \neq 0$ has to be investigated. However, for the massless spectrum a so-called γ phase for a given state can be chosen such that these cases always remain in the spectrum [32].

It was possible to turn this concept into a guiding principle for MSSM-like searches in the heterotic orbifold landscape. In particular, the $\mathbb{Z}_6$-II (1,1) orbifold was studied with two kinds of shift vectors V_1 that ensure a SO(10) with **16**-plet GUT or an E_6 **27**-plet GUT, respectively. This led to a fertile area in the heterotic landscape with a significantly increased probability of obtaining MSSM-like models and was named Mini-Landscape [75, 79].

In chapter 5 we will build upon these findings and study the inner structure of local GUTs with neural networks.

2.5 The Orbifolder

After we have chosen an orbifold geometry $\mathbb{O}$ and found a consistent gauge embedding matrix M, it is possible to compute the low energy 4D orbifold model denoted by model(M). As the orbifold geometry $\mathbb{O}$ is mostly held fixed and we investigate the space of gauge embedding matrices M, we also name M orbifold model. However, some details of the models are not yet understood in a concise formalism, e.g. the Yukawa couplings. Nonetheless, already the spectrum computation seen in section 2.4 is rather involved and takes a considerable amount of time. Fortunately, this procedure was automated in the `orbifolder` [80] project. This program searches for consistent gauge embedding matrices M and returns the massless string spectrum, denoted by spectrum(M). This spectrum consists of the left-chiral matter with all non-Abelian gauge charges. Also, the allowed couplings according to the string selection rules as well as discrete symmetries, can be analyzed. Furthermore, the `orbifolder` can be used to identify equivalent orbifold models based on their massless particle spectra. Two orbifold models, model(M) and model(M'), are said to be equivalent if their massless spectra coincide on the level of the non-Abelian representations[7], i.e.

$$\text{spectrum}(M) \;=\; \text{spectrum}(M') \quad\Rightarrow\quad \text{model}(M) \;\sim\; \text{model}(M')\,. \tag{2.60}$$

In detail, the `orbifolder` can produce an ID label to identify different models. This ID label consists of,

- The number of generations of quark-like and lepton-like states, i.e. this value indicates for MSSM-like models the necessary three generations.
- Number of non-Abelian singlets charged under U(1).
- The number of representations of the left-chiral matter in the orbifold sectors, ordered in ascending order. This contains some information about the allowed couplings since the string selection rules allow only for certain couplings between the different sectors.
- The non-Abelian gauge groups in 4D.
- A value indicating if there is an anomalous U(1).
- The left chiral matter spectrum in terms of their representations under the non-Abelian gauge groups. For a comparison of the ID labels, this entry needs to be basis independent. Therefore, the spectrum is sorted according to its representation content.

---

[7] In case of MSSM-like models the $U(1)_Y$ hypercharge can also be used.

2.6 Definition of MSSM-like models

Using the `orbifolder` package it is possible to quickly search the $\mathbb{Z}^{128}$ space of d_{kI} coefficients for consistent gauge embedding matrices M as defined in section 2.2.2. However, the resulting particle spectra are in general very different from the (MS)SM. Hence, let us first state the MSSM spectrum in representations of the $\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_Y$ gauge symmetry and the left-chiral matter as,

$$\begin{aligned} &3 \times [(\mathbf{3},\mathbf{2})_{1/6} + (\overline{\mathbf{3}},\mathbf{1})_{-2/3} + (\overline{\mathbf{3}},\mathbf{1})_{1/3} + (\mathbf{1},\mathbf{2})_{-1/2} + (\mathbf{1},\mathbf{1})_{1}] + \\ &1 \times [(\mathbf{1},\mathbf{2})_{-1/2} + (\mathbf{1},\mathbf{2})_{1/2}]\ , \end{aligned} \tag{2.61}$$

where the subscript indicates the $\mathrm{U}(1)_Y$ hypercharge. Possible right-handed neutrinos $(\mathbf{1},\mathbf{1})_0$ are not included. In general a large number of additional fields are generated as remnants from the orbifolding procedure. Indeed, there is no known heterotic orbifold compactification that exactly reproduces the (MS)SM as low energy spectrum without additional particles. However, it is possible to define MSSM-like models which allow for additional particle content that is not a priori fatal for the model in the sense that it introduces phenomenological forbidden effects. In detail, it is possible to allow for exotic particles that are vector-like with respect to the SM. These exotic particles are matter that is charged under the SM. In order to be vector-like, each of these exotic particles has a mass-partner transforming in the complex conjugate representation with respect to the SM. The charges of the particles may be different from the ones known in the SM and hence are exotic particles. However, in this context an additional vector-like particle pair which is, e.g. charged like a down-type quark, will be accounted to the set of vector-like exotic particles. Interestingly, it turns out that all types of vector-like exotics that appear in the orbifold landscape amount to 43 different types. They are listed in table D.1 and table D.2.

As the MSSM is defined without any vector-like exotics, these particles eventually have to decouple from the effective field theory spectrum. We let X and $\bar{X}$ denote a pair of vector-like exotics. Usually, these exotics can become very massive through terms in the superpotential. They are of the form $\mathcal{W} \supset (M_\mathrm{Planck})^{1-p}\,(s^0)^p\,X\,\bar{X}$, where $p \in \mathbb{N}_+$ and s^0 is a SM singlet $(\mathbf{1},\mathbf{1})_0$. The SM singlet can acquire a large non-vanishing vacuum expectation value (VEV), which generates a mass term, that can be close to the Planck scale M_Planck, for the vector-like exotics X and $\bar{X}$. Thus, besides three (chiral) generations of quarks and leptons, a Higgs-pair and SM singlets a MSSM-like orbifold model can be equipped with vector-like exotics.

Finally, let us comment on the $\mathrm{U}(1)_Y$ that is normalized such that it results from an SU(5) GUT. In detail, the non-Abelian gauge factors $\mathrm{SU}(3) \times \mathrm{SU}(2)$ for which the three generations of quark-like states appear are embedded into a SU(5) space where the generator that defines the $\mathrm{U}(1)_Y$ is the orthogonal direction with respect to $\mathrm{SU}(3) \times \mathrm{SU}(2)$ in this SU(5) spaces. Hence, the $\mathrm{U}(1)_Y$ is normalized such that it is consistent with gauge coupling unification.

Chapter 3

Non-supersymmetric heterotic orbifolds with supersymmetric substructure

Supersymmetry is a promising extension of the SM that introduces a degeneracy of bosons and fermions. This theoretically appealing symmetry is so powerful that the question arises if a highly constrained framework, such as string theory, predicts its existence from quantum gravity. While supersymmetry is itself widely assumed in string theory, the setting of heterotic string theory provides a well defined non-supersymmetric SO(16)×SO(16) string [44, 81–83]. Compactifications of the SO(16)×SO(16) string [18, 50, 84–94] lead to low-energy spectra close to the SM [95–100]. However, in this non-supersymmetric theories the issue of the cosmological constant arises. As discussed below eq. (2.2) the 4D cosmological constant is proportional to the one-loop vacuum to vacuum amplitude in string theory. This amplitude is expressed in terms of the so-called partition function. In the case of supersymmetry, the partition function vanishes, and hence the cosmological constant vanishes, too. Thus a small breaking of supersymmetry would introduce a small cosmological constant as it is observed in the universe [101, 102]. However, missing signs of supersymmetry at the LHC [103], demand that also non-supersymmetric theories are explored and understood. Unfortunately, the partition function in non-supersymmetric string theories does not vanish at one loop by default and induces a considerably too high value for the cosmological constant. This issue is tackled in various attempts and will also be the main concern in this chapter. In the literature there already exist attempts for a exponentially suppressed cosmological constant at one loop [91, 99, 104–108] trying to reduce the problem. Moreover, there exist non-supersymmetric constructions in the type II string theory with a vanishing cosmological constant at one-loop, see [93, 109–112]. For the type II construction, it was crucial that the compactification was done on asymmetric orbifolds.[1] While it was argued in [99] that heterotic non-supersymmetric string theories induce a non-vanishing cosmological constant, the type II example is the motivation for a detailed understanding of this issue. The goal of this chapter is to investigate the inner structure of symmetric orbifold compactifications on the heterotic string that construct 4D non-supersymmetric models.

---

[1] Note that this cancellation may not persist towards higher loop orders as in supersymmetric models [113].

3.1 Conditions for a vanishing one-loop cosmological constant

In this chapter, we will connect the vanishing of the one-loop cosmological constant to a group theoretical question. Hence, we will focus on the group theory part here and refer the reader for details on the partition function to [24]. As the goal of this section is to find the minimal difference between the supersymmetric and non-supersymmetric case, let us start by investigating the overall structure of the full one-loop partition function that is the integrand of the modular integral eq. (2.2). The full partition function can be factorized into a 4D non-compact part $\mathcal{Z}_{3,1}$ and the orbifold $\mathbb{O}$ compactified internal part $\mathcal{Z}_{\text{int.}}$,

$$\mathcal{Z}_{\text{full}}(\tau,\bar{\tau}) \;=\; \mathcal{Z}_{3,1}(\tau,\bar{\tau})\;\mathcal{Z}_{\text{int.}}(\tau,\bar{\tau})\;. \tag{3.1}$$

Moreover, the world-sheet that describes the one-loop partition function is given by a torus. Hence, the closed string eq. (2.30) has to satisfy a second boundary condition from the element $h \in S$,

$$X(\sigma_0+\pi\tau_1,\sigma_1+\pi\tau_2) \;=\; hX(\sigma_0,\sigma_1)\;, \tag{3.2}$$

where $\tau=\tau_1+i\tau_2$ parameterizes the world-sheet torus [30]. In order to ensure that the boundary condition eq. (2.30) is compatible with eq. (3.2), the associated space group element $h \in S$ has to lie in the centralizer of C_g, i.e. $[h,g]=0$ with $g \in S$ the constructing element according to the boundary condition eq. (2.30). Then, the internal orbifold partition function can be formed as a sum over (g,h)-sectors,

$$\mathcal{Z}_{\text{int.}}(\tau,\bar{\tau}) \;=\; \frac{1}{|P|}\sum_{g}\sum_{h\in C_g}\mathcal{Z}_{\text{int.}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big](\tau,\bar{\tau})\;. \tag{3.3}$$

Finally, the individual terms of the sum in the internal partition function, i.e. each (g,h)-twisted sector, can be factorized in left- and right-moving parts

$$\mathcal{Z}_{\text{int.}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big](\tau,\bar{\tau}) \;=\; \overline{\mathcal{Z}_{\Psi_{\text{R}}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big](\tau)}\,\mathcal{Z}_{X_{\text{R}}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big](\tau,\bar{\tau})\,\mathcal{Z}_{X_{\text{L}}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big](\tau)\;, \tag{3.4}$$

related to the supporting world-sheet fields Ψ_{R}, X_{R} and X_{L}. The explicit form of the contribution from the gauge degrees of freedom $\mathcal{Z}_{X_{\text{L}}}$ in eq. (3.4) depends on the gauge embedding (see section 2.2.2). Hence, this part is conditioned on the specific model and needs detailed computations for each specific case.

The structure of the partition function reveals that the direct computation of the one-loop cosmological constant eq. (2.2) is, in general, connected to the particular compactification parameters and hence complicated and model-dependent. Analogous to the supersymmetric case where the full partition function vanishes, we ask here for model-independent situations with a vanishing cosmological constant. One possibility is to force the integral to be zero by a particular symmetry of the partition function. A suggestion in this context was the Atkin-Lehner symmetry [114–117]. However, this requires a detailed understanding of the modular properties of the full partition function which is not generically known. The next option would be a vanishing partition function. As the non-compact space-time part of the partition function is non-vanishing, the full partition function $\mathcal{Z}_{\text{full}}$ in eq. (3.1) only vanishes, if the internal partition function $\mathcal{Z}_{\text{int.}}$ is zero. From eq. (3.3) one can infer the trivial realization as $\mathcal{Z}_{\text{int.}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big]=0\;\forall\;(g,h)$-sectors. However, there exists also the possibility that some $\mathcal{Z}_{\text{int.}}\big[\begin{smallmatrix}g\\h\end{smallmatrix}\big]\neq 0$. This option involves a non-trivial cancellation between the contributions from the various twisted sectors in eq. (3.3) and hence needs further understanding of the model-dependent details. Since these options involve intricate cancellations, we focus on the model-independent option and demand that

$$\mathcal{Z}_{\text{int.}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right] = 0 \text{ for all pairs } [g,h] = 0\ . \tag{3.5}$$

This condition directly implies that we restrict ourselves to supersymmetric heterotic string theories, i.e. the $\mathrm{E}_8 \times \mathrm{E}_8$ or the SO(32) string. This is due to the one-loop (g,h)-sector corresponding to $(\mathbb{1},\mathbb{1})$. As this 10D sector is modular invariant, it has to vanish by itself. A fact that would not be given for the non-supersymmetric heterotic $\mathrm{SO}(16)\times\mathrm{SO}(16)$ string. In other words, the only way to have a model-independent vanishing integrand is to have supersymmetry at least at the string scale and breaking it by orbifolding to a 4D non-supersymmetric theory.

The internal partition function eq. (3.4), consists for $(g,h) \neq (\mathbb{1},\mathbb{1})$ of a product of three parts. Thus, it vanishes whenever one factor is zero. However, the part of the partition function associated with the right-moving fermions Ψ_{R} in eq. (3.4) encodes all the information about target-space supersymmetry. The (g,h)-twisted partition function $\mathcal{Z}_{X_{\mathrm{R}}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right]$ is associated with the internal orbifold geometry and has no relation to supersymmetry. Hence, it is in general non-zero. Then $\mathcal{Z}_{\Psi_{\mathrm{R}}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right]$ is the last factor that can yield a model-independent vanishing of the partition function. This factor only depends on the space group S, but not on the particular gauge embedding. Furthermore, $\mathcal{Z}_{\Psi_{\mathrm{R}}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right](\tau)$ depends only on the local twist vectors v_g and v_h corresponding to the commuting space group elements g and h, respectively, and not on the torus lattice Γ, see [24]. Thus, according to the classification of 6D orbifolds in section 2.2.1 it is the same for a vast collection of orbifold models.

3.1.1 Vanishing right-moving fermionic partition functions

The right-moving fermionic partition function $\mathcal{Z}_{\Psi_{\mathrm{R}}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right]$ in eq. (3.4) vanishes iff the corresponding (g,h)-sector admits at least one Killing spinor. This can be shown using Riemann identities (see Appendix A of [24]),

$$\mathcal{Z}_{\Psi_{\mathrm{R}}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right](\tau) \;=\; 0 \qquad \Leftrightarrow \qquad \begin{array}{l}\text{The space group elements } g,h \in S \text{ admit}\\ \text{at least one common Killing spinor.}\end{array} \tag{3.6}$$

Thus, in order to investigate a vanishing partition function we can study the spinor embedding of the geometrical twist θ of $g = (\theta, n_\alpha e_\alpha) \in S$. For the spinor representation eq. (2.41), we observe that the possible eigenvalues of $D_{\mathbf{s}}(\theta)$ are $\exp(\pm 2\pi i\, \widetilde{v}_g^a)$ with $a = 1,2,3,4$, where

$$\widetilde{v}_g \;=\; \frac{1}{2}\begin{pmatrix} v_g^1 + v_g^2 + v_g^3 \\ -v_g^1 + v_g^2 + v_g^3 \\ v_g^1 - v_g^2 + v_g^3 \\ v_g^1 + v_g^2 - v_g^3 \end{pmatrix}\ , \tag{3.7}$$

and similarly for $h = (\rho, m_\alpha e_\alpha) \in S$ with $h \in C_g$. Consequently, $D_{\mathbf{s}}(\theta)$ and $D_{\mathbf{s}}(\rho)$ have a common invariant eigenvector if the same components $\widetilde{v}_g^a$ and $\widetilde{v}_h^a$ are integer valued. This eigenvector corresponds to a preserved Killing spinor. Thus the (g,h)-twisted sector preserves some amount of supersymmetry in this substructure, such that the right-moving fermionic partition function $\mathcal{Z}_{\Psi_{\mathrm{R}}}\left[\begin{smallmatrix}g\\h\end{smallmatrix}\right]$ is zero. Hence, if a Killing spinor is invariant under a point group element $D_{\mathbf{s}}(\theta)$, the corresponding $(g,\mathbb{1})$-sector vanishes and we say that a Killing spinor exists *locally* for the $\mathbb{Z}_{N_\theta}$ Abelian subgroup generated by $\theta \in P$. These considerations make it evident that the partition function of any supersymmetric orbifold vanishes. Concretely, the supersymmetric case corresponds to one globally defined constant spinor, which implies a spinor that is invariant for all $D_{\mathbf{s}}(\theta)$ with $\theta \in P$. Such a Killing spinor will be called *global.* The term supersymmetry is avoided to point out that these concepts do not relate to local and global supersymmetry. The vanishing of the right-moving fermionic partition function $\mathcal{Z}_{\Psi_{\mathrm{R}}}\left[\begin{smallmatrix} g \\ h \end{smallmatrix}\right]$ for all (g,h)-sectors is in case of 4D supersymmetric orbifolds related to a global Killing spinor. However, there could exist non-supersymmetric toroidal orbifold compactifications with vanishing right-moving partition function in each (g,h)-twisted sector separately, due to the existence of different supersymmetries in each (g,h)-sector. It is sufficient to demand:

i. There exists a *local* Killing spinor in every (g,h)-sector. In other words, each orbifold sector preserves by itself at least $\mathcal{N}_{(g,h)} = 1$ space-time supersymmetry.

ii. The (g,h)-sectors preserve distinct Killing spinors, such that it is impossible to define a *globally* invariant spinor. Hence, target-space supersymmetry is broken, $\mathcal{N} = 0$. (3.8)

However, the no-go result in the upcoming sections will be that such a construction is not possible. Hence, non-supersymmetric toroidal orbifolds for each possible space group S will always break supersymmetry entirely in at least one (g,h)-sector.

3.1.2 Non-isomorphic embeddings

Another complication of eq. (3.5) is related to the double cover ambiguity for embedding the geometrical action into the spinor representation. As discussed in section 2.3.2, the action of the point group has to be embedded into the eight-dimensional spinor representation. The spin group Spin(6) is the double cover of the orthogonal group SO(6), hence any element θ of the geometrical point group P has two representatives in the Spin(6) group as $\pm D_{\mathbf{s}}(\theta)$, see eq. (A.8). Therefore, if the geometrical point group P has a set of K generators, there exist 2^K possibilities for the action of the point group on spinors. These different embeddings are in general inequivalent.

Note, that this ambiguity can lead to a non-isomorphic embedding. For example, if the order N_θ of a generator θ is odd, then the order of the corresponding spin generator depends on the choice of embedding sign. If $D_{\mathbf{s}}(\theta)$ has the same order N_θ as $\theta \in P$, then we find that $-D_{\mathbf{s}}(\theta)$ is of order $2N_\theta$. However, this non-isomorphic embedding, in the case of odd generators, can be generalized to the defining conditions of finite groups. In particular, this can happen for groups which have only even generators. Concretely, each finite group H is defined by a so-called presentation, i.e. by several defining relations between its generators. We can bring these to the form of a product of generators of H that equals the identity,

$$\theta_1 \cdots \theta_n \;=\; \mathbb{1}_6 \;. \tag{3.9}$$

Inserting eq. (3.9) into eq. (A.8) and using the relation to the charge conjugation matrix defined in eq. (A.7) we observe that the corresponding spinor element of eq. (3.9) has to commute with all gamma matrices Γ_i eq. (A.14) that build the basis of the Clifford algebra,

$$\big[D_{\mathbf{s}}(\theta_1)\cdots D_{\mathbf{s}}(\theta_n),\ \Gamma_i\big] \;=\; 0 \;, \tag{3.10}$$

for all $i = 1,\ldots,6$. However, the eight-dimensional Γ_i matrices generate the so-called extraspecial finite group of order 128, with (GAP ID [128, 2327]). This group has 64 one-dimensional irreducible representations and one eight-dimensional irreducible representation. The eight dimensional representation is the only one which can satisfy the Clifford algebra and is given by the gamma matrices Γ_i. Using this observation we can apply Schur's lemma for irreducible representations of a finite group. Hence, it follows that any element $D_{\mathbf{s}}(\theta_1 \ldots \theta_n)$ of $P_{\mathbf{s}}$ that has to commute with all Γ_i and is not the zero matrix, has to be proportional to the identity matrix $D_{\mathbf{s}}(\theta_1 \ldots \theta_n) = a\, \mathbb{1}_8$ with $a \in \mathbb{C}$. This condition in combination with eq. (A.8) results in $a^2 = 1$. Thus, we obtain that the defining relation for the corresponding group $P_{\mathbf{s}}$ in spinor space, can take two forms,

$$D_{\mathbf{s}}(\theta_1) \cdots D_{\mathbf{s}}(\theta_n) \;=\; \pm\, \mathbb{1}_8 \;. \tag{3.11}$$

Hence, any non-isomorphic embedding of P in $P_{\mathbf{s}}$ contains the element $-\mathbb{1}_8$. Otherwise, if all defining relations for $P_{\mathbf{s}}$ are the same as those of P, then $P_{\mathbf{s}}$ and P are isomorphic $P \simeq P_{\mathbf{s}}$. Thus, for non-isomorphic embeddings the space group element g with $D_{\mathrm{s}}(\theta) = -\mathbb{1}_8$ has no invariant spinor according to the condition on the eigenvalues, see eq. (3.7). Since the right-moving fermionic partition function has to vanish in each (g, h)-sector, we have to avoid this situation. Hence, we restrict ourselves to those cases where the point group P and $P_{\mathbf{s}}$ in spinor space are isomorphic,

$$P \cong P_{\mathbf{s}} \;. \tag{3.12}$$

However, even for the isomorphic case, several embeddings exist. Those correspond to inequivalent models that might have different cosmological constants. Thus, the various isomorphic embeddings are of interest for the following investigations.

3.2 Local Killing spinors in non-supersymmetric orbifolds

Recall that all geometrical point groups that act crystallographically (eq. (2.16)) on 6D tori are classified in CARAT and accordingly labeled from 1 to 7 103 for all $\mathbb{Q}$-classes. The generators for each $\mathbb{Q}$-class are given in some (unspecified) lattice basis e, as GL(6; $\mathbb{Z}$) matrices denoted by $\widehat{\theta} \in \widehat{P}$. The task is to find a lattice basis e and the corresponding SO(6) representation. We start with the relation between the generators in the lattice basis and the geometrical O(6) generators,

$$\theta \;=\; e\, \widehat{\theta}\, e^{-1} \;. \tag{3.13}$$

Using the orthogonality property $\theta^T \theta = \mathbb{1}_6$ and the definition of the torus metric eq. (2.11) this leads to the condition,

$$G \;=\; \widehat{\theta}^T G\, \widehat{\theta} \;, \tag{3.14}$$

for all $\widehat{\theta} \in \widehat{P}$. A solution to the equation above is given by

$$G \;=\; \sum_{\widehat{\rho} \in \widehat{P}} \widehat{\rho}^T \widehat{\rho} \;. \tag{3.15}$$

This solution can be verified by explicit insertion in eq. (3.14) and the simple observation that the action of $\widehat{\theta} \in \widehat{P}$ shuffles the individual terms of the sum. Next CARAT provides with the `FORM_SPACE` function a option to find a basis of all G's that are solutions to eq. (3.14). From there, one could choose a lattice basis, to get the best approximation to a diagonal representation for the generators. In this specific case, it is sufficient to utilize a `CholeskyDecomposition` of G in order to find a lattice basis e, according to eq. (2.11). This basis allows us to transform the $\widehat{\theta}$ by eq. (3.13). Checking the determinant, we find the θ that are actually in SO(6).

However, the geometrical point group P determined does not fully specify the action on the spinors because of the double cover ambiguities. Each orthogonal matrix θ can be written in terms of the Lie algebra as in eq. (A.2). Then, using eq. (A.7) we obtain both possible representations $\pm D_{\mathbf{s}}(\theta)$ in spinor space.

# $\mathbb{Q}$-classes	Restriction
7 103	All $\mathbb{Q}$-classes in six dimensions, i.e. all inequivalent point groups $P \subset \mathrm{O}(6)$.
1 616	Orientable geometrical point groups $P \subset \mathrm{SO}(6)$
106	No element from P rotates in a two-dimensional plane only
63	For each element $\theta \in P$ there is an $\mathrm{SU}(3)_\theta \subset \mathrm{SO}(6)$ in which θ lies, but not necessarily $P \subset \mathrm{SU}(3)$
60	All elements of the point group lie in the same SU(3), which allows for a choice of P_s with $\mathcal{N} \geq 1$ supersymmetry in four dimensions

Table 3.1: *$\mathbb{Q}$-classes that survive various requirements for local and global Killing spinors.*

3.2.1 Number of invariant spinors

The global and local Killing spinors correspond to the invariant Killing spinors of the whole group P_s and the $\mathbb{Z}_{N_\theta}$ subgroups generated by the individual elements $D_s(\theta) \in P_s$, respectively. Therefore, it is necessary to understand the number of Killing spinors preserved by some subgroup $A \subset P_s$. As Spin(6) is isomorphic to SU(4), the point group action in spinor space is fully specified in a four-dimensional irreducible representation **4** of SU(4). This so called Weyl representation is given by the chiral projection of $D_\mathbf{s}(\theta)$ to $D_\mathbf{4}(\theta)$ or its conjugate, see eq. (A.13). Thus, any A-invariant four-dimensional Weyl spinor $\Psi_{\text{inv.}}$ is left invariant under the elements of the subgroup $D_\mathbf{4}(\theta) \in A$,

$$D_\mathbf{4}(\theta)\, \Psi_{\text{inv.}} \;=\; \Psi_{\text{inv.}} \;, \tag{3.16}$$

To obtain the eigenspace of A-invariant spinor spinors, we can define the projection operator by $D_\mathbf{4}(\theta)\, \mathcal{P}^A = \mathcal{P}^A$ for all $D_\mathbf{4}(\theta) \in A$. This condition is satisfied by,

$$\mathcal{P}^A \;=\; \frac{1}{|A|} \sum_{D_\mathbf{4}(\theta)' \,\in\, A} D_\mathbf{4}(\theta') \;. \tag{3.17}$$

Since the projection operator eigenvalues are limited to 0 or 1, we can take the trace over the projection operator to get the dimension of the invariant subspace. Thus we can count the number of A-invariant Killing spinors,

$$\mathcal{N}^A \;=\; \mathrm{Tr}\big(\mathcal{P}^A\big) \;=\; \frac{1}{|A|} \sum_{D_\mathbf{4}(\theta') \,\in\, A} \mathrm{Tr}\left(D_\mathbf{4}(\theta')\right) \;. \tag{3.18}$$

For $A = \langle\theta\rangle$, $\mathcal{N}^{\langle\theta\rangle}$ computes the number of *local* Killing spinors according to the $\mathbb{Z}_{N_\theta}$ subgroup of P_s. Consequently, for $\mathcal{N} = \mathcal{N}^P$ we obtain the number of *global* Killing spinors which determines the amount of target-space supersymmetry.

3.2.2 Point groups admitting local Killing spinors

The definitions for the explicit construction of the spinor representation allows us to investigate the 7 103 $\mathbb{Q}$-classes associated with the geometrical point groups in 6D. However, a huge fraction of these possibilities will not survive the most basic physical considerations as we will discuss next. These considerations are summarized in table 3.1.

Beginning with the essential condition on the geometrical point group to preserve the orientation $P \subset \mathrm{SO}(6)$ in order to allow for a definition of spinors at all. This mere constraint

CARAT-Index	Group	Generator relations	Order	Local twist vectors
3375	$\mathbb{Z}_3 \rtimes \mathbb{Z}_4$	$\theta_1^4 = \theta_2^3 = \mathbb{1},\ \theta_2\,\theta_1\,\theta_2 = \theta_1$	12	$\left(\frac{1}{4}, \frac{1}{4}, -\frac{1}{2}\right)\,, \left(\frac{1}{3}, -\frac{1}{3}, 0\right)$
5751	Q_8	$\theta_1^4 = \mathbb{1},\ \theta_1^2 = \theta_2^2,\ \theta_1\,\theta_2\,\theta_1 = \theta_2$	8	$\left(\frac{1}{4}, \frac{1}{4}, -\frac{1}{2}\right)\,, \left(\frac{1}{4}, -\frac{1}{4}, 0\right)$
6737	SL(2, 3)	$\theta_1^3 = \theta_2^4 = \mathbb{1}, (\theta_2\,\theta_1)^2 = \theta_1^2\,\theta_2$	24	$\left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}\right)\,, \left(\frac{1}{4}, -\frac{1}{4}, 0\right)$

Table 3.2: *The three* $\mathbb{Q}$*-classes for which all elements have an individual* SU(3) *embedding but not simultaneously. Note that the two local twist vectors are obtained from two different bases since these groups are all non-Abelian.*

reduces the number of possible point groups to only 1 616. However, even more point groups are to be disregarded for our approach. From the remaining point groups, only 106 are able to get a trivial eigenvalue for a Killing spinor. The rest of the instances rotate only in one complex plane. In this case, the local twist vector v_g defined in eq. (2.19) is given by $v_g = (0, v_g^1, 0, 0)$. Hence, for a non-trivial rotation v_g^1 none of the eigenvalues in the spinorial representation $D_\mathbf{s}(\theta)$, eq. (2.41), is one and thus, $D_\mathbf{s}(\theta)$ has no invariant Killing spinors. These trivial considerations already reduce the possible set of situations we can investigate massively. However, there is more that we can demand at this point. In [57, 118] a check on the holonomy group of P was performed to find the $\mathbb{Q}$-classes that allow for $\mathcal{N} \geq 1$ supersymmetry compactifications. This check leaves only 60 point groups. However, in our case, we can ease this constraint on the holonomy group and modify the requirement such that each element fits into an individual SU(3). It turns out that only 3 additional geometrical point groups benefit from this non-trivial change in the SU(3)-check. Even though each point group element can be embedded individually into spinor space to preserve at least one Killing spinor, not all the required choices can be made at the same time for all elements in $P_\mathbf{s}$. Hence, those 3 geometrical point groups do not allow for supersymmetry. They are the natural first guess to be a possible solution to our investigation and are given in table 3.2.

However, it is not clear whether the 60 point groups with SU(3) holonomy also allow for another choice of the spin embedding $P_\mathbf{s}$ such that different Killing spinors are preserved in the various sectors. Thus they would break supersymmetry globally but keeping invariant Killing spinors locally. Eventually, it is necessary to construct, for each of the 63 geometrical point groups P, all possible spinor space embeddings. However, it turns out that in each embedding, there always exists at least one element that does not preserve a local Killing spinor if $P_\mathbf{s}$ has no global Killing spinor. Note that the relevant elements can change between the different embeddings. Consequently, this explicit check provides a nonexistence proof for non-supersymmetric orbifolds with local Killing spinors for each point group element. We exemplify this analysis for the Q_8 orbifolds in section 3.4. In the following section, we will connect this no-go result to the representation theory of finite groups.

3.3 Finite group theoretical non-existence proof

The main take away so far is that the amount of supersymmetry is not fixed by the condition $P \subset \mathrm{SU}(3)$ alone, but only the specific representation in spinor space determines the amount of supersymmetry. As this connects supersymmetry to the specific representation, explicit constructions are computationally intensive. Hence, we show that it is possible to use finite group theory instead. Let us state the necessary facts of group theory to follow the upcoming sections next.[2]

3.3.1 Elements of finite group theory

The representation theory of a finite group H is based on characters. For each representation $\mathbf{R}$ the corresponding character is given by the trace $\chi_{\mathbf{R}} : H \to \mathbb{C}$ for the elements $b \in H$

$$\chi_{\mathbf{R}}(b) \;=\; \mathrm{Tr}(D_{\mathbf{R}}(b)) \;. \tag{3.19}$$

Since the trace is invariant under conjugation the characters of the conjugacy classes[3] fully determine the representation. For finite groups the number of conjugacy classes is given by the number of irreducible representations, c. Hence, a vector of length c defines the representation $\mathbf{R}$

$$\chi_{\mathbf{R}} \;=\; \Big(\chi_{\mathbf{R}}([b_1]), \ldots, \chi_{\mathbf{R}}([b_c])\Big) \;. \tag{3.20}$$

For convenience the first entry of this vector is dedicated to the identity element, since $\chi_{\mathbf{R}}(\mathbb{1}) = |\mathbf{R}|$ determines the dimension of $\mathbf{R}$. For representations constructed by direct sums or tensor products, the characters are computed by normal sums and products of the corresponding characters

$$\chi_{\mathbf{R}\oplus\mathbf{R}'}(b) \;=\; \chi_{\mathbf{R}}(b) + \chi_{\mathbf{R}'}(b) \;, \qquad \chi_{\mathbf{R}\otimes\mathbf{R}'}(b) \;=\; \chi_{\mathbf{R}}(b)\,\chi_{\mathbf{R}'}(b) \;. \tag{3.21}$$

Furthermore, the k times anti-symmetrized tensor products, denoted by $[\mathbf{R}]_k$, can be expressed in terms of the characters of the representation $\mathbf{R}$. Using the expressions derived in [120, 121] we get

$$\chi_{[\mathbf{R}]_2}(b) = \frac{1}{2}\Big((\chi_{\mathbf{R}}(b))^2 - \chi_{\mathbf{R}}(b^2)\Big) \;, \tag{3.22}$$

$$\chi_{[\mathbf{R}]_4}(b) = \frac{1}{4!}\Big((\chi_{\mathbf{R}}(b))^4 - 6\,\chi_{\mathbf{R}}(b^2)\,(\chi_{\mathbf{R}}(b))^2 + 8\,\chi_{\mathbf{R}}(b^3)\,\chi_{\mathbf{R}}(b) + 3\,\left(\chi_{\mathbf{R}}(b^2)\right)^2 - 6\,\chi_{\mathbf{R}}(b^4)\Big) \;.$$

The power of this approach is that one can define the following inner product

$$\langle \chi_{\mathbf{R}}, \chi_{\mathbf{R}'} \rangle_H \;=\; \frac{1}{|H|}\sum_{b' \in H} \overline{\chi_{\mathbf{R}}(b')}\,\chi_{\mathbf{R}'}(b') \;, \tag{3.23}$$

for any two character vectors associated with representations $\mathbf{R}$ and $\mathbf{R}'$ of H. With respect to this inner product the characters of irreducible representations $\mathbf{r}_x$ are orthonormal,

$$\langle \chi_{\mathbf{r}_x}, \chi_{\mathbf{r}_y} \rangle_H \;=\; \delta_{xy} \;. \tag{3.24}$$

Thus, the irreducible representations build a simple basis of the space of representations. This inner product efficiently computes the coefficients n_x of the decomposition of any representation $\mathbf{R}$ in irreducible representations,

$$\mathbf{R} \;=\; \bigoplus_{x=1}^{c} (\mathbf{r}_x \oplus)^{n_x} \;, \tag{3.25}$$

---

[2] Details on finite group theory can be found for example in [119].

[3] Defined as in the case of inequivalent fixed points in eq. (2.22).

where $(\mathbf{r}_x\oplus)^{n_x} = \mathbf{r}_x \oplus \cdots \oplus \mathbf{r}_x$ and $n_x \in \mathbb{N}_0$. Using eq. (3.21) and the orthogonality of the irreducible characters eq. (3.24) we find that

$$\chi_{\mathbf{R}}(b) \;=\; \sum_{x=1}^{c} n_x\, \chi_{\mathbf{r}_x}(b)\;, \quad \text{with} \quad n_x \;=\; \langle \chi_{\mathbf{r}_x}, \chi_{\mathbf{R}} \rangle_H\;. \tag{3.26}$$

Furthermore, the characters can be used to compute the branching to $\mathbb{Z}_N$ representations. For the Abelian group $\mathbb{Z}_N$, all N group elements commute and thus lie in distinct conjugacy classes. Since the dimension squared of all irreducible representations equals the order of the finite group, $\sum_{x=1}^{c}(\chi_{\mathbf{r}_x}(\mathbb{1}))^2 = |H|$, there are N one-dimensional irreducible representations of $\mathbb{Z}_N$. They are denoted by $\mathbf{1}_q$ for $q = 0, \ldots, N-1$ and $\mathbf{1}_0$ labels the trivial representation. Since all representations are one dimensional the characters are simultaneously the representations, i.e. $\chi_{\mathbf{1}_q}(b) = D_{\mathbf{1}_q}(b) = \exp(2\pi\mathrm{i}q/N)$ for the $\mathbb{Z}_N$ generator b. Thus, for a $\mathbb{Z}_N$ subgroup that is generated by $b \in H$, we can give the branching of the representation $\mathbf{R}$ of H into the irreducible representations of the $\mathbb{Z}_N$-subgroup, using the character inner product eq. (3.23) for the subgroup $\mathbb{Z}_N$,

$$n_{\mathbf{R};q} \;=\; \langle \chi_{\mathbf{R}}, \chi_{\mathbf{1}_q} \rangle_{\mathbb{Z}_N} \;=\; \frac{1}{N} \sum_{b' \in \mathbb{Z}_N} \overline{\chi_{\mathbf{R}}(b')}\, \chi_{\mathbf{1}_q}(b')\;. \tag{3.27}$$

$n_{\mathbf{R};q}$ counts how often the irreducible representation $\mathbf{1}_q$ appears in the branching. Therefore, we had to use the characters with respect to two different groups, i.e. $\chi_{\mathbf{R}}$ are the characters of the full group H while $\chi_{\mathbf{1}_q}$ are the characters of the irreducible group from the $\mathbb{Z}_N$-subgroup.

3.3.2 Spinor representation and $\mathbb{Q}$-classes

From eq. (2.17), one can infer that the 7 103 $\mathbb{Q}$-classes of 6D orbifolds provided by CARAT are possible six-dimensional representations $\mathbf{6}$ of finite groups H. Those finite groups that make up all $\mathbb{Q}$-classes amount to only 1 594. In general, there exist several inequivalent realizations as integral 6×6-matrices with different eigenvalues for a given finite group H. For this reason, we may take the different abstract finite groups H, underlying the $\mathbb{Q}$-classes, as our starting point.

From section 3.2.1, it is clear that we are interested in representations $\mathbf{4}$ of these finite groups H, that are reducible in general. However, not all four-dimensional representation $\mathbf{4}$ of one of these finite groups corresponds to a spinor representation that can be connected with a 6D toroidal orbifold. The representation can originate from U(4) instead of SU(4). Moreover, the resulting six-dimensional representation,

$$\mathbf{6} \;=\; [\mathbf{4}]_2\;, \tag{3.28}$$

connected to the two-times anti-symmetrized tensor product of $\mathbf{4}$ (see section 3.3.1), not necessarily corresponds to one of the vector representations listed in the CARAT classification. Thus, two conditions need to be fulfilled:

1. $\det(D_{\mathbf{4}}(b)) = 1$ for all $b \in H$
2. The representation matrices associated with eq. (3.28) are isomorphic to a $\mathbb{Q}$-class. (3.29)

In order to circumvent the explicit construction of representation matrices, it is beneficial to observe that the determinant is the fully anti-symmetrized one-dimensional representation of a group [121]. Thus the determinant condition is linked to the the four-times anti-symmetrization of the **4**. In other words, if the resulting singlet representation from $[\mathbf{4}]_4$ is the trivial one-dimensional representation: $\chi_{[\mathbf{4}]_4} = (1, \ldots, 1)$ the determinant for all elements $b \in H$ is one. This can be checked easily using the four-times anti-symmetrized character formula eq. (3.22). Similarly, the isomorphism to a $\mathbb{Q}$-class can be checked using the character formula of two-times anti-symmetrized representations eq. (3.22) to compute the character $\chi_{\mathbf{6}} = \chi_{[\mathbf{4}]_2}$. This character values $\chi_{\mathbf{6}}$ have to match the characters $\chi_{\mathbf{v}}$ of a H related $\mathbb{Q}$-classes from the CARAT list. Hence, by utilizing finite group characters, we can check the conditions without the need to construct any representation matrix explicitly.

3.3.3 Killing spinors and singlet representations

The approach of counting the number of A-invariant spinors $\mathcal{N}^A$ by taking the trace of the corresponding projection operator eq. (3.18) (see section 3.2.1), connects now trivially to the concept of characters that are defined by the trace eq. (3.19). Thus we obtain,

$$\mathcal{N}^A \;=\; \frac{1}{|A|}\sum_{b'\,\in\,A} \chi_{\mathbf{4}}(b') \;=\; \langle \chi_{\mathbf{4}}, \chi_{\mathbf{1}} \rangle_A \;=\; n^A_{\mathbf{1}}\ , \tag{3.30}$$

where in the first step, the trivial representation **1** was inserted, to make use of the inner product eq. (3.23). According to eq. (3.27) this computes the coefficient in the branching of **4** for the trivial representation of A. Thus, the number of A-invariant spinors equals the number $n^A_{\mathbf{1}}$ of trivial singlet representations **1** from A in the branching of **4**.

Let us specify the general subgroup branching eq. (3.30) for *local* and *global* Killing spinors. First, for the case of a $\mathbb{Z}_{N_b}$ subgroup of H, which is generated by a single element $b \in H$. Hence, we identify $A = \langle b \rangle$, and recognize, that the number of trivial singlets in the branching equals the amount of preserved spinors by b. Thus, we determine the amount of supersymmetry in this substructure by,

$$\mathcal{N}^{\langle b \rangle} \;=\; n^{\langle b \rangle}_{\mathbf{1}}\ . \tag{3.31}$$

Secondly, if we identify $A = H$ we observe,

$$\mathcal{N} \;=\; n^H_{\mathbf{1}}\ . \tag{3.32}$$

Hence, the amount of target-space supersymmetry, i.e. the number of *global* Killing spinors, is given by the decomposition coefficient eq. (3.26) for the irreducible trivial singlets of H in **4**. In case the spinor representation **4** includes a trivial singlet **1** of H in its decomposition, this trivial singlet **1** will branch for any subgroup into the trivial singlet of that subgroup eq. (3.30), in particularly for the $\mathbb{Z}_{N_b}$ subgroups generated by each element $b \in H$. Hence, if an orbifold has a global Killing spinor, each (g, h)-sector has a local Killing spinor.

3.3.4 Nonexistence proof

With the above results, we can use finite group theory to prove that non-supersymmetric toroidal orbifolds only admit Killing spinors locally in all (g, h)-sectors, if there exists a global Killing spinor. Therefore, the procedure is:

1. Consider all faithful **4**'s of H, that do not contain a trivial singlet representation of H in order to avoid global Killing spinors by eq. (3.32).
2. Reject those representations **4** that can not satisfy the conditions in eq. (3.29). In particularly, the four-dimensional representation $D_{\mathbf{4}}$ should be a subgroup of SU(4) and $[\mathbf{4}]_2 = \mathbf{6}$ should result in a six-dimensional representation of the $\mathbb{Q}$-classes.

3. Construct, all $\mathbb{Z}_N \subset H$ subgroups, and compute the number of trivial $\mathbb{Z}_N$-singlet in the branching.

This method shows that none of the 1 594 different abstract groups H can fulfill all requirements. Equivalently, as in the explicit approach, for each **4** remaining after the first two steps, there is at least one cyclic subgroup, for which **4** does not branch to a trivial $\mathbb{Z}_N$-singlet. Hence, for all non-supersymmetric 6D toroidal orbifolds, there is always a (g, h)-sector without any local Killing spinor.

The power of the finite group theoretical proof lies especially in the clear and fast performance, i.e. no numerical difficulties arise with this approach. In the next section, the explicit construction and the finite group theoretical method will be illustrated in the case of the Q_8 finite group. However, in section 3.5 the group theoretical considerations will be extended far beyond the limits of 6D toroidal orbifolds.

3.4 Examples with Q_8 point group

In order to illustrate several of the abstractly discussed concepts above, we utilize in this section the example of $H = Q_8$. Note, this group connects to one of the three exceptional cases table 3.2 that allow for each element an SU(3) embedding but not all simultaneously. First, the group structure of Q_8 is reviewed in section 3.4.1. Afterwards, in section 3.4.2 the conditions of representation theory (section 3.3) are determined and illustrated. It turns out that only the representation of Q_8 orbifolds with CARAT-index 5750 can fulfill all conditions simultaneously. Hence, this $\mathbb{Q}$-class is used in the explicit construction to emphasize that there are four inequivalent embeddings of the same geometrical Q_8 action into spinor-space, see section 3.4.3.

3.4.1 Group Q_8 and its representations

For the quaternion group Q_8 with 8 elements (GAP-ID [8, 4]) the generators b_1 and b_2 fulfill the following defining conditions

$$Q_8 = \left\langle b_1, b_2 \;\middle|\; b_1^4 = \mathbb{1},\ b_2^2 = b_1^2,\ b_2\, b_1\, b_2^{-1} = b_1^{-1} \right\rangle \ . \tag{3.33}$$

It turns out that Q_8 has five conjugacy classes given by:

$$[\mathbb{1}] \ = \ \{\mathbb{1}\} \ , \quad [b_1] \ = \ \{b_1, b_1^3\} \ , \quad [b_2] \ = \ \{b_2, b_2^3\} \ , \quad [b_3] \ = \ \{b_3, b_3^3\} \ , \quad [b_1^2] \ = \ \{b_1^2\} \ , \tag{3.34}$$

where we used $b_3 = b_1 b_2$ as convenient notation. Q_8 has 5 irreducible representations which are denoted by $\mathbf{1}_{++}$ for the trivial one-dimensional representation, $\mathbf{1}_{+-}$, $\mathbf{1}_{-+}$, $\mathbf{1}_{--}$ for the three non-trivial one-dimensional representations, and a single faithful two-dimensional representation **2**. The characters for the conjugacy classes $[b]$ lead to the character table T of Q_8 given in table 3.3. The matrix representations of the elements b_a, $a = 1, 2, 3$ read

$$D_{\mathbf{1}_{\text{AB}}}(b_1) \ = \ \text{A}\,1 \ , \quad D_{\mathbf{1}_{\text{AB}}}(b_2) \ = \ \text{B}\,1 \ , \quad D_{\mathbf{1}_{\text{AB}}}(b_3) \ = \ \text{A}{\cdot}\text{B}\,1 \quad \text{and} \quad D_{\mathbf{2}}(b_a) \ = \ -\,\mathrm{i}\,\sigma_a \ , \tag{3.35}$$

for $\text{A}, \text{B} = \pm$ and σ_a are the Pauli matrices. The representation $\mathbf{R}$ of Q_8 can be decomposed in terms of the irreducible representations

$$\mathbf{R} \ = \ \bigoplus_{\text{A,B}=\pm} (\mathbf{1}_{\text{AB}} \oplus)^{n_{(\text{AB})}} \ \oplus \ (\mathbf{2}\,\oplus)^{n_{(\mathbf{2})}} \ . \tag{3.36}$$

$T = \chi_{\mathbf{r}}([b])$	$[b] = [\mathbb{1}]$	$[b_1]$	$[b_2]$	$[b_3]$	$[b_1^2]$
$\mathbf{r} = \mathbf{1}_{++}$	1	1	1	1	1
$\mathbf{1}_{+-}$	1	1	-1	-1	1
$\mathbf{1}_{-+}$	1	-1	1	-1	1
$\mathbf{1}_{--}$	1	-1	-1	1	1
$\mathbf{2}$	2	0	0	0	-2

Table 3.3: *Character table T of the quaternion group Q_8.*

From this general Ansatz of the decomposition we can compute the character of the representation $\mathbf{R}$ using table 3.3 and eq. (3.26) to obtain

$$\begin{aligned}\chi_{\mathbf{R}} = {} & \chi_{\mathbf{R}}\left([\mathbb{1}]\,,\ [b_1]\,,\ [b_2]\,,\ [b_3]\,,\ [b_1^2]\right) \qquad (3.37)\\ = {} & \left(n_{(\mathbf{1}_*)} + 2n_{(\mathbf{2})}\,,\ \sum_{\text{A,B}=\pm} \text{A}\, n_{(\text{AB})}\,,\ \sum_{\text{A,B}=\pm} \text{B}\, n_{(\text{AB})}\,,\ \sum_{\text{A,B}=\pm} \text{A}\cdot\text{B}\, n_{(\text{AB})}\,,\ n_{(\mathbf{1}_*)} - 2n_{(\mathbf{2})}\right)\,,\end{aligned}$$

where $n_{(\mathbf{1}_*)} = n_{(++)} + n_{(+-)} + n_{(-+)} + n_{(--)}$ is the total number of one-dimensional representations in the decomposition of $\mathbf{R}$. Rearranging the relations one can obtain the total number of singlets $n_{(\mathbf{1}_*)}$ and the number of doublet $n_{(\mathbf{2})}$ representations in terms of the characters $\chi_{\mathbf{R}}$ of $[\mathbb{1}]$ and $[b_1^2]$. This yields

$$n_{(\mathbf{1}_*)} \;=\; \frac{\chi_{\mathbf{R}}([\mathbb{1}]) + \chi_{\mathbf{R}}([b_1^2])}{2}\,, \qquad n_{(\mathbf{2})} \;=\; \frac{\chi_{\mathbf{R}}([\mathbb{1}]) - \chi_{\mathbf{R}}([b_1^2])}{4}\,. \qquad (3.38)$$

On the other hand, the assignment of the various types of singlets is encoded in the character $\chi_{\mathbf{R}}$ evaluated on $[b_1], [b_2]$ and $[b_3]$. For the five irreducible representations of Q_8, the rules of tensor products read:

$$\mathbf{1}_{\text{AB}} \otimes \mathbf{1}_{\text{CD}} \;=\; \mathbf{1}_{\text{A}\cdot\text{C B}\cdot\text{D}}\,, \qquad \mathbf{1}_{\text{AB}} \otimes \mathbf{2} \;=\; \mathbf{2}\,, \qquad \mathbf{2}\otimes\mathbf{2} \;=\; \bigoplus_{\text{A,B}=\pm} \mathbf{1}_{\text{AB}}\,, \qquad [\mathbf{2}]_2 \;=\; \mathbf{1}_{++}\,. \qquad (3.39)$$

These properties arise from the characters of table 3.3 and the tensor product rules of characters eq. (3.21). In the second relation the zero entries for the conjugacy classes $[b_a]$ in the character vector of $\mathbf{2}$ annihilate the characters of the specific singlet involved. Moreover, since $\chi_{\mathbf{2}\otimes\mathbf{2}} = (4,0,0,0,4)$, it follows from the generic equation for the number of doublets in $\mathbf{R}$, eq. (3.38), that $n_{(\mathbf{2})} = 0$ for $\mathbf{2}\otimes\mathbf{2}$. Consequently, $n_{(\mathbf{1}_*)} = 4$ and hence, shows that this tensor product decomposes into singlet representations. The characters show that only the combination of all four irreducible singlet representations can realize the zeros in the character vector of $\mathbf{2}$. Finally, the relation for $[\mathbf{2}]_2$ of eq. (3.39) can be confirmed by using the character formula for anti-symmetrized representations eq. (3.22).

The explicit branching of the irreducible representation of Q_8 for its cyclic subgroups is given in table 3.4. In total there exist three maximal subgroups isomorphic to $\mathbb{Z}_4$ generated by b_1, b_2, b_3, as well as a $\mathbb{Z}_2$ generated by the elements $b_1^2 = b_2^2 = b_3^2$ of the fourth non-trivial conjugacy class.

3.4.2 Spinorial interpretation of 4 of Q_8

The criteria on four-dimensional representations of a finite group H to ensure that a heterotic orbifold can be obtained, which admits Killing spinors locally, can now be applied to the possible

Subgroup		$\mathbf{1}_{++}$	$\mathbf{1}_{+-}$	$\mathbf{1}_{-+}$	$\mathbf{1}_{--}$	$\mathbf{2}$
$\mathbb{Z}_{\boldsymbol{N}}$	gen.	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$\mathbb{Z}_4$	b_1	$\mathbf{1}_0$	$\mathbf{1}_0$	$\mathbf{1}_2$	$\mathbf{1}_2$	$\mathbf{1}_1 \oplus \mathbf{1}_3$
$\mathbb{Z}_4$	b_2	$\mathbf{1}_0$	$\mathbf{1}_2$	$\mathbf{1}_0$	$\mathbf{1}_2$	$\mathbf{1}_1 \oplus \mathbf{1}_3$
$\mathbb{Z}_4$	b_3	$\mathbf{1}_0$	$\mathbf{1}_2$	$\mathbf{1}_2$	$\mathbf{1}_0$	$\mathbf{1}_1 \oplus \mathbf{1}_3$
$\mathbb{Z}_2$	b_1^2	$\mathbf{1}_0$	$\mathbf{1}_0$	$\mathbf{1}_0$	$\mathbf{1}_0$	$\mathbf{1}_1 \oplus \mathbf{1}_1$

Table 3.4: *Branching of the irreducible representations of Q_8 into the irreducible representations $\mathbf{1}_q$ of the three maximal subgroups that are isomorphic to $\mathbb{Z}_4$ as well as the $\mathbb{Z}_2$ generated from the non-trivial conjugacy class of Q_8, see eq.* (3.27)*.*

four-dimensional representation $\mathbf{4}$ of Q_8. For this particular abstract group exist four $\mathbb{Q}$-classes with CARAT-indices 5750, 5751, 6100, and 6101. The generators of the $\mathbb{Q}$-classes and some details are listed in table 3.5. We begin by analyzing if the described toroidal orbifold is orientable. For the $\mathbb{Q}$-classes with CARAT-indices 6100 and 6101 there exist generators with $\det\left(\widehat{\theta}_2\right) = -1$ and $\det\left(\widehat{\theta}_1\right) = -1$, respectively. Remember that this condition relates to the property of toroidal orbifold geometries not admitting spinors. Hence, only the $\mathbb{Q}$-classes with CARAT-indices 5750 and 5751 do. Next, the question of a faithful four-dimensional representation is investigated. As the highest irreducible representation of Q_8 is of dimension two, any four-dimensional representation is reducible. As the only irreducible and faithful representation of Q_8 has to be in the decomposition, i.e. $\mathbf{2}$, faithful decompositions for $\mathbf{4}$ read

$$n_{(\mathbf{1}_*)} = 2 \ ; \ n_{(\mathbf{2})} = 1 \ , \qquad n_{(\mathbf{1}_*)} = 0 \ ; \ n_{(\mathbf{2})} = 2 \ . \tag{3.40}$$

$\boldsymbol{D_4(Q_8)} \subset \mathbf{SU(4)}$

In order to interpret $\mathbf{4}$ as spinor representation, it is necessary that $D_{\mathbf{4}}(b) \in \mathrm{SU}(4)$ for all $b \in Q_8$. As discussed below eq. (3.29) this is the case if the character vector of $[\mathbf{4}]_4$ equals the trivial one-dimensional representation. Using the anti-symmetrized character relations eq. (3.22), for the realizations eq. (3.40), it turns out that $\chi_{[\mathbf{4}]_4}([\mathbb{1}]) = \chi_{[\mathbf{4}]_4}([b_1^2]) = 1$. From here, we proceed with a case differentiation. Therefore, let us start with

$$\mathbf{4} \ = \ \mathbf{2} \oplus \mathbf{2} \ . \tag{3.41}$$

Here, we have $\chi_{\mathbf{2}\oplus\mathbf{2}}([b_a]) = \chi_{\mathbf{2}}([b_a]) + \chi_{\mathbf{2}}([b_a]) = 0$ for $a = 1,2,3$. Hence, it follows from the anti-symmetric relations eq. (3.22), that also the other three conjugacy classes fulfill

$$\chi_{[\mathbf{2}\oplus\mathbf{2}]_4}([b_a]) = \frac{1}{4!}\left(3\left(\chi_{\mathbf{2}\oplus\mathbf{2}}\left(b_a^2\right)\right)^2 - 6\,\chi_{\mathbf{2}\oplus\mathbf{2}}\left(b_a^4\right)\right) = 1 \ , \tag{3.42}$$

as $\chi_{\mathbf{2}\oplus\mathbf{2}}\left(b_a^2\right) = -4$ and $\chi_{\mathbf{2}\oplus\mathbf{2}}\left(b_a^4\right) = \chi_{\mathbf{2}\oplus\mathbf{2}}(\mathbb{1}) = 4$, see table 3.3. Thus, the determinant criterion, i.e. $\chi_{[\mathbf{2}\oplus\mathbf{2}]_4} = \chi_{\mathbf{1}_{++}}$, is satisfied and eq. (3.41) results in a four-dimensional representation which admits a spinor interpretation.

For the other faithful realization, $\mathbf{4} = \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{1}_{\mathrm{CD}} \oplus \mathbf{2}$, requiring that $\chi_{[\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{1}_{\mathrm{CD}}\oplus\mathbf{2}]_4}([b_a]) = 1$ for all $a = 1,2,3$ leads to three equations

$$\chi_{\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{1}_{\mathrm{CD}}\oplus\mathbf{2}}([b_a])^4 + 8\,\chi_{\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{1}_{\mathrm{CD}}\oplus\mathbf{2}}([b_a])^2 - 48 \ = \ 0 \ . \tag{3.43}$$

$\mathbb{Q}$-class	Lattice basis generators		Det		Decomposition of $\mathbf{v}$
CARAT-ind.	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\lvert\widehat{\theta}_1\rvert$	$\lvert\widehat{\theta}_2\rvert$	and its character
5750	$\begin{pmatrix} -1 & -1 & -1 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 & 0 & 0 \\ -1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	1	1	$\mathbf{1}_{++} \oplus \mathbf{1}_{++} \oplus \mathbf{2} \oplus \mathbf{2}$ $\chi_{\mathbf{v}} = (6, 2, 2, 2, -2)$
5751	$\begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 & -1 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	1	1	$\mathbf{1}_{-+} \oplus \mathbf{1}_{-+} \oplus \mathbf{2} \oplus \mathbf{2}$ $\chi_{\mathbf{v}} = (6, -2, 2 - 2, -2)$
6100	$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$	1	−1	$\mathbf{1}_{++} \oplus \mathbf{1}_{+-} \oplus \mathbf{2} \oplus \mathbf{2}$ $\chi_{\mathbf{v}} = (6, 2, 0, 0, -2)$
6101	$\begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$	−1	1	$\mathbf{1}_{+-} \oplus \mathbf{1}_{--} \oplus \mathbf{2} \oplus \mathbf{2}$ $\chi_{\mathbf{v}} = (6, 0, -2, 0, -2)$

Table 3.5: *Table of the four $\mathbb{Q}$-classes based on the finite group Q_8. The first column gives their* CARAT*-index. The second column provides the* $\mathrm{GL}(6,\mathbb{Z})$ *matrices of the two generators. The third column gives the determinants. The last column states the characters $\chi_{\mathbf{v}}$ and the decomposition eq.* (3.36) *of the six-dimensional representations.*

From table 3.3 we observe that the finite group Q_8 contains no complex characters in its character table. Thus, the solution in $\mathbb{R}$ to these equations are given by $\chi_{\mathbf{1}_{\text{AB}}\oplus\mathbf{1}_{\text{CD}}\oplus\mathbf{2}}([b_a]) = \pm 2$ for $a = 1, 2, 3$. Inserting this in eq. (3.37) and solving for the number of singlet representation in $\chi_{\mathbf{1}_{\text{AB}}\oplus\mathbf{1}_{\text{CD}}\oplus\mathbf{2}}([b_a])$, we obtain that two identical singlet representations are necessary, i.e.

$$\mathbf{4} \ = \ \mathbf{1}_{\text{AB}} \oplus \mathbf{1}_{\text{AB}} \oplus \mathbf{2} \ , \tag{3.44}$$

for A, B $= \pm$. Hence, there survive four different realizations.

Resulting six-dimensional representation defines a $\mathbb{Q}$-class

In order to be a valid four-dimensional representation $\mathbf{4}$ of toroidal compactification, $\mathbf{4}$ has to fulfill the property of constructing $[\mathbf{4}]_2 = \mathbf{6}$ a six-dimensional representation that corresponds to a $\mathbb{Q}$-class, representation $\mathbf{v}$, in table 3.5. Hence, for each six-dimensional representation, we will check the condition, if it is possible to find a four-dimensional representation of Q_8, by using the anti-symmetric character relation eq. (3.22). The first cross-check is the dimension of the representation. This value is given by the character of $\mathbb{1}$: $\chi_{[\mathbf{4}]_2}([\mathbb{1}]) = (4^2 - 4)/2 = 6 = \chi_{\mathbf{6}}([\mathbb{1}])$. Next, we observe that $\chi_{\mathbf{6}}([b_1^2]) = -2$ for all Q_8 $\mathbb{Q}$-classes listed in table 3.5. Using eq. (3.37) together with $\chi_{\mathbf{4}}([b_1^4]) = \chi_{\mathbf{4}}([\mathbb{1}]) = 4$,

$$\chi_{\mathbf{6}}([b_1^2]) \;=\; \tfrac{1}{2}\,\chi_{\mathbf{4}}([b_1^2])^2 - 2 \;\overset{!}{=}\; -2\ , \tag{3.45}$$

which implies $\chi_{\mathbf{4}}([b_1^2]) = 0$. Consequently, we have for the other three conjugacy classes $[b_a]$, $a = 1, 2, 3$, of Q_8:

$$\chi_{\mathbf{6}}([b_a]) \;=\; \tfrac{1}{2}\,\chi_{\mathbf{4}}([b_a])^2\ , \tag{3.46}$$

using $b_a^2 = b_1^2$. The right-hand-side is non-negative. Thus, for negative $\chi_{\mathbf{6}}([b_a])$, there exists no related four-dimensional representation. From table 3.5, we observe that the vector representations defined by the $\mathbb{Q}$-classes 5751 and 6101 have negative characters, which rules out the connection to a four-dimensional representation of Q_8.

Resulting spinorial embedding

Altogether, a four-dimensional representation $\mathbf{4}$ of Q_8 admits a spinorial interpretation if either $\mathbf{4} = \mathbf{2} \oplus \mathbf{2}$ or $\mathbf{4} = \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{2}$:

- For $\mathbf{4} = \mathbf{2} \oplus \mathbf{2}$ we get $\chi_{\mathbf{2}\oplus\mathbf{2}} = (4, 0, 0, 0, -4)$. This directly implies $\chi_{\mathbf{2}\oplus\mathbf{2}}([b_1^2]) = -4 \neq 0$ and hence eq. (3.45) is violated and this can not be realized in any Q_8 toroidal orbifold geometry.
- For $\mathbf{4} = \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{2}$ we have $\chi_{\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{2}} = (4, \mathrm{A}\cdot 2, \mathrm{B}\,2, \mathrm{A}\cdot\mathrm{B}\cdot 2, 0)$. Thus, eq. (3.45) is satisfied by $\chi_{\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{1}_{\mathrm{AB}}\oplus\mathbf{2}}([b_1^2]) = 0$. Moreover, eq. (3.46) leads to $\chi_{\mathbf{6}}([b_a]) = 2$ for $a = 1, 2, 3$, which correspond to the $\mathbb{Q}$-class with CARAT-index 5750, see table 3.5.

Summarizing, the only combination that satisfies all conditions is the spinorial embedding $\mathbf{4} \;=\; \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{1}_{\mathrm{AB}} \oplus \mathbf{2}$ together with the $\mathbb{Q}$-class 5750 as geometrical point group.

3.4.3 Double cover embeddings

The analysis above resulted in the Q_8 $\mathbb{Q}$-class with CARAT-index 5750, as the only class that allows for four-dimensional spinorial representations, that are connected to toroidal orbifolds. These $\mathbf{4}$'s are stated in eq. (3.44). There exists still the freedom to choose the non-trivial one-dimensional representations by the signs of $\mathrm{A}, \mathrm{B} = \pm$. Since all different choices correspond to the same six-dimensional representation, this ambiguity can be understood as the different double cover embeddings of SO(6) into Spin(6): Q_8 has two generators, and thus, there are $2^2 = 4$ possible embeddings. We will show this freedom in the spinor embedding explicitly by applying the methods of section 3.2.

Explicit spinorial embeddings

We begin with the Q_8 group generators as GL(6, $\mathbb{Z}$) matrices $\widehat{\theta}_1, \widehat{\theta}_2$ in an unspecified lattice basis e. To derive the vielbein e, we make use of eq. (3.15) to find an instance of the metric G and then apply a `CholeskyDecomposition`. This results in,

$$G \;=\; \begin{pmatrix} 16 & 8 & 8 & -8 & 0 & 0 \\ 8 & 16 & 0 & -8 & 0 & 0 \\ 8 & 0 & 16 & -8 & 0 & 0 \\ -8 & -8 & -8 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 \end{pmatrix}\ , \qquad e \;=\; 2\begin{pmatrix} 2 & 1 & 1 & -1 & 0 & 0 \\ 0 & \sqrt{3} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & 2\sqrt{\frac{2}{3}} & -\sqrt{\frac{2}{3}} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{2} \end{pmatrix}\ . \tag{3.47}$$

Decomposition of the spinorial repr.	Number of Killing spinors locally $\mathcal{N}^{\langle \mathbb{1}\rangle}$	$\mathcal{N}^{\langle b_1\rangle}$	$\mathcal{N}^{\langle b_2\rangle}$	$\mathcal{N}^{\langle b_3\rangle}$	$\mathcal{N}^{\langle b_1^2\rangle}$	Globally $\mathcal{N}^{Q_8}$
$\mathbf{4}_{++} = \mathbf{2} \oplus \mathbf{1}_{++} \oplus \mathbf{1}_{++}$	4	2	2	2	2	2
$\mathbf{4}_{+-} = \mathbf{2} \oplus \mathbf{1}_{+-} \oplus \mathbf{1}_{+-}$	4	2	0	0	2	0
$\mathbf{4}_{-+} = \mathbf{2} \oplus \mathbf{1}_{-+} \oplus \mathbf{1}_{-+}$	4	0	2	0	2	0
$\mathbf{4}_{--} = \mathbf{2} \oplus \mathbf{1}_{--} \oplus \mathbf{1}_{--}$	4	0	0	2	2	0

Table 3.6: *Table of the the four spinorial representations* **4** *and their decompositions for the* $\mathbb{Q}$*-class 5750 associated with* Q_8*. Given are the locally preserved Killing spinors for each Abelian subgroup generated by the representatives of the five conjugacy classes* $[\mathbb{1}], [b_1], [b_2], [b_3]$*,* $[b_1^2]$ *and the globally preserved Killing spinors of* Q_8*.*

Then the SO(6)-representation matrices are obtained by the conjugation eq. (3.13),

$$\theta_1 = \begin{pmatrix} 0 & -\frac{1}{\sqrt{3}} & -\sqrt{\frac{2}{3}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & 0 & 0 & -\sqrt{\frac{2}{3}} & 0 & 0 \\ \sqrt{\frac{2}{3}} & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \theta_2 = \begin{pmatrix} 0 & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.48}$$

in the Euclidean basis. From this representation, we use the logarithmic map to the so(6) Lie algebra element for these generators. Here the basis can be used to compute the coefficients of the Lie algebra decomposition eq. (A.2) for both generators. These coefficients, together with the spin Lie algebra generators eq. (A.7) are used in the exponential map to observe the eight-dimensional spinor representation $D_{\mathbf{s}}$. At this point, it is convenient to use the positive chirality projector introduced in eq. (A.4) to express the obtained spinor representation in the Weyl-basis, as a four-dimensional matrix representation, namely

$$D_{\mathbf{4}_{\text{AB}}}(\theta_1) = \text{A} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{\mathrm{i}}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 \\ 0 & -\sqrt{\frac{2}{3}} & \frac{\mathrm{i}}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad D_{\mathbf{4}_{\text{AB}}}(\theta_2) = \text{B} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{\mathrm{i}}{\sqrt{3}} & -\frac{\mathrm{i}}{\sqrt{2}} - \frac{1}{\sqrt{6}} & 0 \\ 0 & -\frac{\mathrm{i}}{\sqrt{2}} + \frac{1}{\sqrt{6}} & \frac{\mathrm{i}}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.49}$$

The different double cover embeddings are parameterized by the prefactor $\text{A}, \text{B} = \pm$. The factors arise in the above procedure from the complex logarithm of the geometrical group that is not unique but leaves the freedom to add integer numbers to the coefficients. Adding those integers leaves the geometrical action invariant but results in a sign for the spinor representation. This corresponds to the four inequivalent decompositions eq. (3.44) of **4** of Q_8 that admit a spinorial interpretation. Thus, the abstract representation theory and the explicit construction yield the same result.

Number of local and global Killing spinors

To determine the number of Killing spinors, we can utilize either the explicit spinorial representation matrices from eq. (3.49) or the decompositions in eq. (3.44) of the four-dimensional representation **4** of Q_8 for the *local*, and *global* case. We use eq. (3.31) applied to any representative of each of the five conjugacy classes and eq. (3.32) for the whole Q_8 point group. Table 3.6 summarizes the results for the four inequivalent choices in the double cover.

The last column of this table reveals that the choice AB $= ++$ returns $\mathcal{N} = 2$ supersymmetry. Contrary, the other three situations produce $\mathcal{N} = 0$ in 4D, as these **4**'s do not contain any trivial singlet representation $\mathbf{1}_{++}$ of Q_8. For these cases, there always exist two conjugacy classes, which do not allow any Killing spinors. This illustrates the impossibility to build a toroidal Q_8 orbifold without global Killing spinors, but where each group element by itself admits invariant spinors.

3.5 A finite group conjecture

At this point, we have already excluded the idea of local Killing spinors that do not lead to a global Killing spinor for 6D toroidal orbifolds. However, condition 2. in eq. (3.29) is, in fact, obsolete. In other words, the no-go result decouples from the idea to be related to a crystallographic action associated with some 6D toroidal orbifold. Moreover, it appears that the conditions build up a pure group theoretical no-go. More specifically, even if no relation to the $\mathbb{Q}$-classes is assumed, it is impossible to find the desired branching into trivial singlets for $\mathbb{Z}_N$ subgroups. This allows us to strengthen the no-go result and connect it to a group theoretical conjecture:

Conjecture 1 *There does not exist any finite group H that has a four-dimensional representation $D_\mathbf{4}$ with trivial determinant, i.e.* $\det(D_\mathbf{4}(b)) = 1$ *for all $b \in H$, such that:*

i. $D_\mathbf{4}$ does not contain the trivial singlet representation of H.

ii. The branching of $D_\mathbf{4}$ to all $\mathbb{Z}_N \subset H$ subgroups always contains the trivial $\mathbb{Z}_N$-singlet representation.

Note already at this point how exceptional the constellation of this conjecture is. Below, we illustrate the importance of our assumptions for conjecture 1 and how sensitive it is to small variations. However, the conjecture is not only based on the finite groups from the CARAT set. Additional support comes from checks performed on the list of finite groups of order up to 500 from the SmallGroups Library of GAP [19]. In total, this amounts to $\mathcal{O}(100\,000)$ finite groups that were verified to fulfill the conjecture. The orders of 1 594 finite groups from the CARAT $\mathbb{Q}$-classes range from order one for the trivial group to order 103 680 for $\mathbb{Z}_2 \times (\mathrm{O}(5,3) \rtimes \mathbb{Z}_2)$ with CARAT-index 2804. Moreover, 443 groups of the set are of order 501 or higher, and extend the subset of the SmallGroups Library. We went through the whole list of finite groups and constructed for each group H the faithful four-dimensional representation **4**, such that the decomposition into irreducible representations of H does not contain the trivial singlet. Furthermore, the determinant constraint $D_\mathbf{4}(b) \in \mathrm{SU}(4)$ for every element $b \in H$ was checked by computing the four-times anti-symmetrized character eq. (3.22) and demand to obtain the trivial singlet of H. Then, the branching of **4** of H to every $\mathbb{Z}_N$ subgroup was considered with respect to the number of trivial $\mathbb{Z}_N$ singlets. It turned out that for all cases there exists at least one $\mathbb{Z}_N$ subgroup of H for which 4 does not branch into a trivial singlet. This provides strong evidence for the conjecture.

Exceptional constellation of the conjecture

The above inspections for a wide variety of finite groups increase the confidence for conjecture 1. However, let us illustrate how extraordinary the constellation of this conjecture is. For this, we will show two examples where the defining conditions of conjecture 1 are relaxed, i.e. the dimension of the representation and the unit determinant are allowed to deviate from the situation in string theory.

Five-dimensional representations

For the first example, we take the five-dimensional representation,

$$\mathbf{5} \;=\; \mathbf{1}_{+-} \oplus \mathbf{1}_{-+} \oplus \mathbf{1}_{--} \oplus \mathbf{2} \;, \tag{3.50}$$

of Q_8 that does not contain a trivial singlet in its decomposition. Furthermore, using the explicit matrix representation of the irreducible representations of Q_8 from eq. (3.35), we can express the generators of this representation as,

$$D_{\mathbf{5}}(b_1) \;=\; \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 & -\mathrm{i} & 0 \end{pmatrix} , \qquad D_{\mathbf{5}}(b_2) \;=\; \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} , \tag{3.51}$$

and check that the determinant $\det(D_{\mathbf{5}}(b)) = 1$ for all $b \in Q_8$. This five-dimensional representation eq. (3.50) has no global Killing spinor, i.e. does not contain the trivial singlet $\mathbf{1}_{++}$. Contrary to the situation for four-dimensional representations, we find that all three non-trivial singlets $\mathbf{1}_{\text{AB}}$, $\text{AB} \neq ++$, in this representation, ensure the trivial singlet $\mathbf{1}_0$ in each branching to $\mathbb{Z}_4$ subgroups, as table 3.4 shows. Hence, we have shown that this $\mathbf{5}$ of Q_8 does fulfill three conditions of conjecture 1. Consequently, the assumption of four-dimensional representations is crucial to our conjecture 1.

Representation without unit determinant

Similar to the above case, it is possible to construct an example where the determinant condition is violated, i.e. $\det(D_{\mathbf{4}}) \neq 1$, but the remaining conditions from the conjecture are fulfilled. We choose as suitable example the finite group T_7. This group has five irreducible representations: Namely, one three-dimensional representation $\mathbf{3}$ and its complex conjugate $\bar{\mathbf{3}}$, as well as the trivial singlet $\mathbf{1}_0$ and two non-trivial singlets $\mathbf{1}_1, \mathbf{1}_2$, see [122]. Additionally, the $\mathbb{Z}_N$ subgroups of T_7 are of order three and seven. The one-dimensional representation $\mathbf{1}_q$ $(q = 1, 2)$ of T_7 branches to $\mathbf{1}_q$ of $\mathbb{Z}_3$ and to $\mathbf{1}_0$ of $\mathbb{Z}_7$, respectively. Thus, to ensure the trivial singlet for the $\mathbb{Z}_7$ subgroup and obtain a faithful representation of T_7 we choose a four-dimensional representation $\mathbf{4}$ that decomposes into,

$$\mathbf{4} \;=\; \mathbf{3} \oplus \mathbf{1}_1 \;. \tag{3.52}$$

This representation is constructed such that it does not contain a trivial singlet. However, it turns out that the faithful $\mathbf{3}$ provides the trivial singlet for the $\mathbb{Z}_3$ subgroup, i.e.

$$\begin{array}{ccccl} \mathbf{4} \;= & \mathbf{3} & \oplus & \mathbf{1}_1 & \text{of } T_7 \;. \\ & \downarrow & & \downarrow & \\ & \mathbf{1}_0 \oplus \mathbf{1}_1 \oplus \mathbf{1}_2 & \oplus & \mathbf{1}_1 & \text{of } \mathbb{Z}_3 \\ \cline{2-5} & \mathbf{1}_1 \oplus \mathbf{1}_2 \oplus \mathbf{1}_4 & \oplus & \mathbf{1}_0 & \text{of } \mathbb{Z}_7 \end{array} \tag{3.53}$$

Therefore, by neglecting the condition on the determinant, it is possible to construct a representation that has exactly the branching behavior we were looking for, i.e. a trivial singlet of each $\mathbb{Z}_N$.

In combination with the Q_8 example, this shows how unfortunate the situation for the vanishing of the cosmological constant is connected to the representation theory of finite groups.

3.6 Chapter summary

In this chapter, we investigated the conditions for a vanishing cosmological constant at one-loop level, for symmetric heterotic orbifolds that lead to non-supersymmetric 4D theories. For a model-independent vanishing of the cosmological constant it turns out that the right-moving fermionic partition function has to preserve in each (g, h)-sector a so-called local Killing spinor. We proof the non-existence of these constructions as a local Killing spinor in each (g, h)-sector inevitably leads to a preserved global Killing spinor, which results in a $\mathcal{N} = 1$ 4D theory. We extend this non-existence result to a finite group theory conjecture and verify it for a wide variety of finite groups. This gives strong indications that the no-go result extends beyond symmetric toroidal orbifolds, e.g. to asymmetric orbifolds. Hence, the results seem to be in conflict with the vanishing one-loop cosmological constant for the non-supersymmetric type II string theory. For the type II string the left- and right-moving string is supersymmetric and hence they can preserve supersymmetry in an alternating pattern. Such a construction is not possible for the heterotic string. Finally, we developed a number of new techniques to study non-Abelian orbifolds systematically. These might prove useful to investigate heterotic model building on non-Abelian orbifolds more explicitly than in the past [118, 123].

Chapter 4

Statistics of the heterotic orbifold landscape

As we have seen in the previous section, supersymmetric compactifications of string theories avoid several complications. However, even if we restrict ourselves to the 138 orbifold geometries with Abelian point group (i.e. $\mathbb{Z}_{N_1}$ or $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$), that preserve the SU(3) holonomy and hence can be chosen to lead to $\mathcal{N} = 1$ supersymmetric 4D theories [57], a significant difficulty arises, the so-called string landscape; a name chosen to describe the huge number of possible compactifications. Early estimates on the lower bound of possible inequivalent constructions led to $\sim 10^{1500}$ [18, 124] possible 4D theories from strings. This huge but finite number seems at first sight to be a weak spot of string theory. However, once a compactification configuration is chosen, the resulting theory is completely fixed. Indeed, the complete particle content and all couplings are set. In the following it will become clear that the counting of these models takes into account the most basic consistency conditions, and includes all resulting 4D theories, even those that are not in accordance with the phenomenology of the SM. The analytic construction of realistic 4D models from the heterotic orbifold landscape is complicated and follows some rules that are yet to be discovered. In this thesis, we will focus on the construction of MSSM-like models (section 2.6) that are very appealing for beyond the SM physics. In the past many efforts were made to identify 4D string models close to the MSSM, see e.g. [46–54, 75, 79, 125–127]. These searches in the string landscape were based on random scans [126–128]. Over time physical intuition led to guiding principles in the search for realistic models, like the pattern of local GUTs [75, 79, 125]. Computer-aided techniques to improve the search are based on genetic algorithms [129], network science [130] and in general techniques from machine learning (see e.g. [131–142]).

4.1 Solving modular invariance consistency conditions

A particularly difficult problem in the context of heterotic orbifolds is to solve the Diophantine equations of the modular invariance conditions eq. (2.46). To sample a set of 128-compactification parameters for the gauge embedding matrix M eq. (2.26) at once, is doomed to fail. Testing the set against the modular invariance conditions will certainly lead to a violation of the conditions. The reason is the low probability of success, as we will show in this section. A more elaborated procedure to create a model is to do this step by step and draw one 16-vector M_k after another, always checking the relevant modular invariance conditions of the vectors drawn so far. This procedure is illustrated in fig. 4.1. Recall that we can expand the 16-vectors either in terms of the simple roots eq. (2.27), with coefficients $c_{kI} \in \mathbb{Z}$, or in the dual basis eq. (2.28), with coeffi-

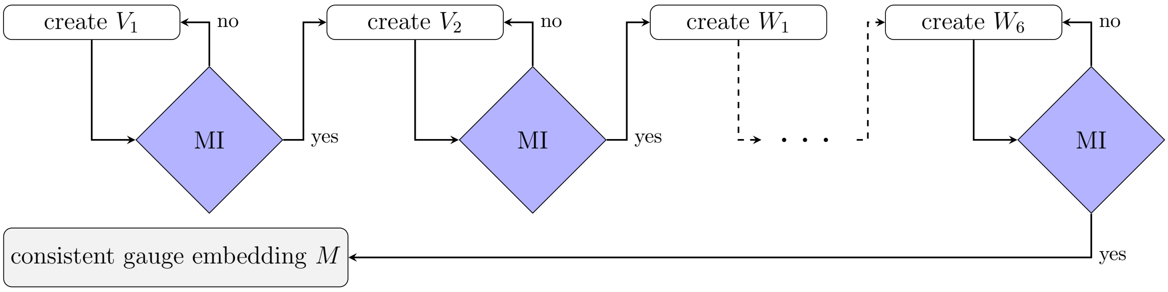


Figure 4.1: *Flowchart of the construction of shift vectors and Wilson lines, starting with the shift vector V_1 at step $n = 1$ and ending with the Wilson line W_6 at step $n = 8$. At each step $n = 1, \dots, 8$, the vector M_n is chosen randomly (eq. (4.1)) and the corresponding modular invariance (MI) conditions are tested.*

cients $d_{kI} \in \mathbb{Z}$. Since the upcoming chapters will make use of the d_{kI}-coefficients, we will focus on them. However, for certain investigations in this section the simple roots basis coefficients c_{kI} are more convenient. To sample the d_{kI}-coefficients a discrete uniform distribution, eq. (B.4), based on the order N_k of M_k is chosen,

$$d_{kI} \sim \mathcal{U}\{-\beta N_k,\ \beta N_k\} \qquad \text{i.i.d.} \qquad \text{for} \ \ I = 1, \dots, 16 \ , \tag{4.1}$$

where for a given k the d_{kI} are drawn independently from the same distribution and hence are said to be independent and identically distributed (i.i.d.). The parameter β scales the search space and is set to the standard value $\beta = 4$, if not stated differently.

Additionally, for non-Abelian symmetries between the matter on the fixed points, which could be used for flavor symmetry, it is beneficial to turn off Wilson lines [143]. Moreover, it turns out that a vanishing Wilson line has a positive effect on the probability of finding MSSM-like models [79], hence an adjustment of the joint probability seems to be appropriate. Therefore, we increase the chance for a zero vector by modifying the joint probability with an additional chance for drawing a zero vector,

$$\Pr(d_{kI} = 0 \ \forall \ I) \ = \ 2 \cdot 10^{-5} \ . \tag{4.2}$$

Sampling from this distribution ensures an even exploration of the compactification parameter space $\mathbb{Z}^{128}$ of the landscape. However, when processed by the successive search fig. 4.1, the resulting output probability distribution is a highly non-trivial joint probability distribution of the coefficients d_{kI}. Projecting the resulting distribution on the various coefficients (histograms) or combinations of two coefficients yields random uniform (bivariate) distributions. Consequently, the joint probability distribution that leads to modular invariant models is susceptible to the exact high dimensional combination of the coefficient. Simple patterns, like a different distribution for sampling the coefficients d_{kI}, can not be inferred from the data at this stage.[1]

However, for further studies of the landscape, and the application of various techniques, it is useful to estimate the probability $\Pr(M = \text{MI})$ that a gauge embedding M fulfills the modular invariance conditions (MI) from eq. (2.46). The probabilistic underlying principle for the success of the successive search strategy, fig. 4.1, is given by the product rule eq. (B.2). One can decompose $\Pr(M = \text{MI})$ into,

$$\Pr(M = \text{MI}) \ = \ \prod_{k=1}^{8} \Pr\left(M_k = \text{MI} \ \middle| \ \{M_i\}_{i=1}^{k-1} = \text{MI}\right) \ , \tag{4.3}$$

---

[1] For the projection, we use `pairplot` from seaborn [182].

where $\{M_i\}_{i=1}^{k-1} = \{M_1, \dots, M_{k-1}\}$ is the set of all gauge embedding 16-vectors M_i that were previously drawn and fulfill the corresponding modular invariance conditions. Each of these sub-probabilities is measured during model generation in the successive search. Therefore, the number of trials to generate a 16-vector is counted until a 16-vector is drawn that fulfills the related modular invariance conditions. This can be done at free cost as it is a side-product of performing the actual successive search to find modular invariant gauge embedding matrices M. This can be used to model each sub-probability with a binomial distribution, as is explained in detail in appendix B.4. In order to compute the margin of error for $\Pr(M = \mathrm{MI})$, the sub-probabilities with 95% confidence intervals for each step of the successive creation are estimated. The details of this statistical method are provided in appendix B.5. On a dataset of 9 981 245 successively constructed modular invariant models, we find for the $\mathbb{Z}_6$-II (1,1) orbifold geometry that,

$$\Pr(M = \mathrm{MI}) = 1.439 \cdot 10^{-8} \pm 3.395 \cdot 10^{-10} \tag{4.4}$$

with 95% confidence interval. This probability of a point in $\mathbb{Z}^{128}$ to be modular invariant, emphasizes the substantial restriction of the landscape given by the modular invariance conditions. However, this probability is more than just illustrative. For the application of explorative algorithms, such as reinforcement learning, it is crucial to keep this sparseness of the landscape in mind. As an example, the neighborhood of an MSSM-like model is examined in fig. 4.2. The definition of a reward function for the direct study of the compactification space will result in only very few positive responses, a so-called sparse reward signal problem [184]. The assumption for this argument is that the compactification space is tested on an euclidean path.

4.2 Symmetries of the landscape

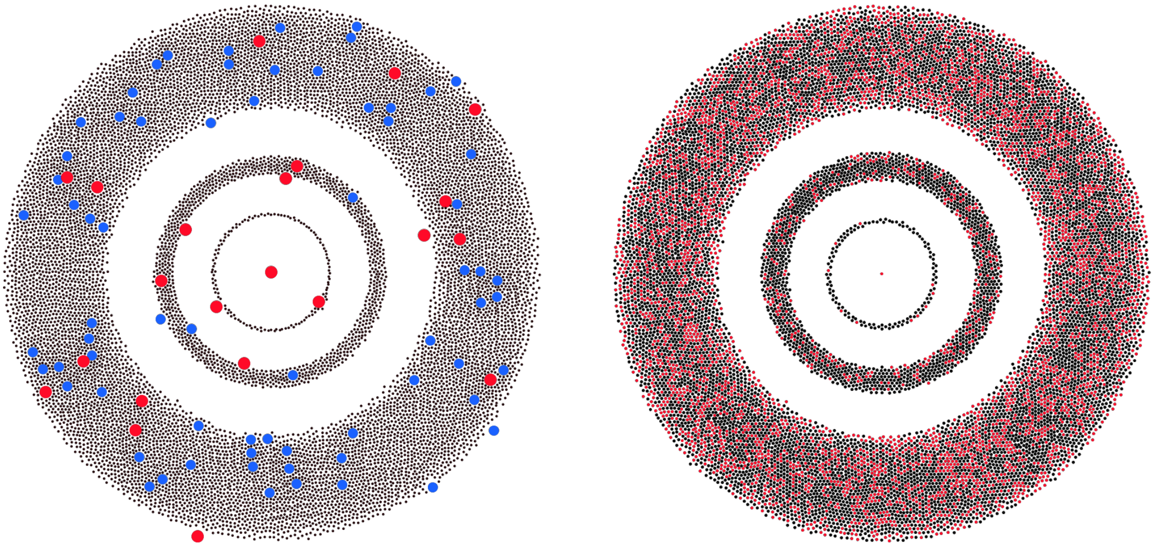
Subsequent to successfully creating a modular invariant gauge embedding M, one can compute the spectrum of massless particles spectrum(M), using the `orbifolder` (see section 2.5). Unfortunately, even though it is difficult to find such a solution, it is by no means guaranteed to find physically inequivalent solutions. This relates to the fact, that there is only direct access to the compactification parameters d_{kI} of $\mathbb{Z}^{128}$, but not to the space of inequivalent physical spectra.

Due to the symmetries of the landscape, the current search strategy produces several equivalent physical models, i.e., the function mapping from $\mathbb{Z}^{128}$ to the physical spectrum is not injective, and the complexity of finding novel physical models is a growing function of exploration time. An important issue that arises with this complication was studied in [144], where the term “floating correlations” is introduced. It refers to the fact, that the statistics of the landscape are non-trivially related to the inequivalent physical models. A technique to compute reliable statistics of the inequivalent physical models, taking the floating of correlations into account, is suggested in [144]. However, since this procedure itself requires a substantial amount of data, it is not practical in the context of this thesis. However, we take the floating of correlations serious, and statistics are estimated on the equivalent physical models. The statistics of equivalent models is not affected by this complication and still, provides guidance towards a refined search strategy.

The primary source behind the floating of correlations for inequivalent physical models are symmetries on the compactification parameters [145], which relate different numerical gauge embedding matrices M such that the same physics emerges. In the next subsections, the primary sources of such symmetries are described.

---

[2] Constructing the local neighborhood of all equivalent MSSM-like models found by the search of fig. 6.3 did not increase the total amount of inequivalent MSSM-like models stated in table 6.6.



(a) *Direct neighborhood of an MSSM-like model in $\mathbb{Z}^{64}$, i.e. generated by $c_{kI} \mapsto c_{kI} \pm 1$. 22 MSSM-like models which result in 4 inequivalent MSSM-like models.*[2] *68 models are with no generation.*

(b) *Lattice translation neighborhood of an MSSM-like model, i.e. $c_{kI} \mapsto c_{kI} \pm N_k$. 4 467 MSSM-like models, which result in 1 inequivalent MSSM-like model.*

Figure 4.2: *Neighborhood of an MSSM-like model in the $\mathbb{Z}_6$-II (1,1) orbifold geometry. In red, MSSM-like models. In blue, modular invariant models with no generation of quarks and leptons. In black, points in $\mathbb{Z}^{64}$ that are not modular invariant. For $\mathbb{Z}_6$-II (1,1) the geometrical conditions in table 2.2 show that 64 free parameters c_{kI} can be chosen, hence $\mathbb{Z}^{128}$ reduces to effectively $\mathbb{Z}^{64}$. From the MSSM-like model in the center each distinct radius represents the correspondent three L_1 distances (in fig. 4.2a in terms of the integers, in fig. 4.2b in terms of complete lattice vectors) of the models on these 64 integers. This corresponds to 357 889 points in $\mathbb{Z}^{64}$ that were created. The not modular invariant points are too dominant, and it is necessary to downsample those. Concretely, in the second radius, a black point corresponds to ~ 8 and in third (last) radius to ~ 34, not modular invariant parameter configurations. Note, the azimuth angle has no information about the position of models. The points randomly distribute along this direction. Created with `Gephi` [183].*

4.2.1 Translations

In the basis of the root lattice of $\mathrm{E}_8 \times \mathrm{E}_8$ with the coefficients c_{kI}, eq. (2.27), it is possible to see that the adding of an integer linear combination of lattice vectors

$$M_k \mapsto M_k + n_{kI}\alpha_I \qquad \text{with} \quad n_{kI} \in \mathbb{Z} \tag{4.5}$$

leaves most of the equations invariant. However, there are two substantial exceptions. In general the addition of lattice vectors does not fulfill the modular invariance conditions, see fig. 4.2b. Furthermore, translations can lead to brother models [66], since the projection conditions eq. (2.57) might get modified. While the second case does not apply to the orbifold geometry $\mathbb{Z}_6$-II $(1,1)$, i.e. brother models are not possible, this possibility of different physical models has to be kept in mind for the extension of any approach towards other orbifold geometries. Getting back to the first exception, the violation of modular invariance conditions can lead to the exclusion of several models. Particularly, if one restricts the search space such that no addition of lattice vectors is possible, such as searching in the fundamental domain of the torus $\mathbb{T}^{16}$,

$$c_{kI} \sim \mathcal{U}\{\,0,\; N_k - 1\} \quad \text{i.i.d.} \quad \forall\, I \tag{4.6}$$

this area does not guarantee to generate all physical models. There might exist a combination of compactification parameters that violates modular invariance and hence is not allowed in the fundamental torus domain. However, this model can get valid as soon as one adds special lattice vectors that do not lead to brother models. Even more symmetries might relate these combinations of points towards another combination of compactification parameters in the fundamental torus domain. However, the physically attractive MSSM-like models show a particular taste of finely balanced compactification parameters which makes them rare occasions in general. In a practical approach, a successive search in terms of fig. 4.1 is performed sampling the coefficients c_{kI} from the torus domain, i.e. from eq. (4.6). This procedure shows a good behavior lowering the number of inequivalent models and leads to an improvement in the number of MSSM-like models which is related to the fact that sampling from eq. (4.6) enhances the probability of drawing a zero 16-vector significantly.

Altogether, this approach leaves considerable uncertainty about the completeness of the total number of physical models while still being appealing for taking care of a significant amount of symmetry. Moreover, further study of the landscape results in a more sophisticated technique (section 6.1.1) that divides out another symmetry of the landscape, namely the Weyl symmetry. This technique outperforms the strategy of eq. (4.6).

Finally, let us stress that for constructed models, the points above have to be reconsidered. In the case of $\mathbb{Z}_6$-II $(1,1)$ constructing a set of gauge embedding matrices M results in special numerical values. For machine learning approaches, the symmetry of adding lattice vectors will significantly increase the complexity when they do not have access to the resulting physics of M. This is the case in the unsupervised learning approach in chapter 5, and hence we will find a representation of M such that lattice translations do not change the numerical representation.

4.2.2 Geometrical redefinitions

The fixed point structure of orbifolds allows for certain redefinitions that correspond to permutations of the local shift vectors V_h. This can be seen as a relabeling of the fixed points. However, the resulting physics is entirely equivalent. These redefinitions correspond to adding the fractional lattice vectors M_k that correspond to the Wilson lines, to the model M itself.

To make this concrete, let us investigate the fixed point structure of the first twisted sector $T_{(1)}$ of $\mathbb{Z}_6$-II (1, 1) from fig. 2.2. Here, the three fixed points of the $\mathbb{T}^2/\mathbb{Z}_3$ from model M are permuted in the orbifold M' that is shifted in the origin,

$$V_1 \;\mapsto\; V_1 + W_3 \;. \tag{4.7}$$

This new orbifold is physically equivalent. However, the local shift vectors V_h are permuted in an anti-clockwise direction. For example, for $h = (\theta, 0)$ we have for the fixed point at the origin $V_{(\theta,0)} \mapsto V_{(\theta,0)} + W_3 = V_{(\theta,e_3)}$. Hence, the same local shift vector acts in these two gauge embedding matrices in different fixed points. This changes the position of the local twisted matter but not the global spectrum of the orbifold compactification, neither the allowed interactions. This can be repeated, and hence such a transformation can be seen as the generator of a $\mathbb{Z}_3$ permutation symmetry of the $\mathbb{Z}_6$-II (1, 1) landscape. Together with the redefinition of the Wilson line,

$$W_3 \;\mapsto\; W_3 + W_3 \;, \tag{4.8}$$

which corresponds to a $\mathbb{Z}_2$, they generate the complete S_3 permutation group of the three fixed points. Finally, let us mention the special $\mathbb{Z}_2$ between the first and second E_8 factor. Interchanging the E_8 factors results in the same physics as it only interchanges the sorting of the gauge groups. All of these permutations lead to the same physics and together with the permutations of the other $\mathbb{T}^2$ sectors, we will develop in section 5.1.1 a method to chose a unique representative M of all equivalent gauge embeddings in the symmetry cycle of geometrical redefinitions.

4.2.3 Weyl symmetry

Another source of symmetry is given by the Weyl group [146]. It is a symmetry of the root system Φ of simple Lie algebras, in our case $\mathrm{E}_8 \times \mathrm{E}_8$. The Weyl group is generated by reflections on the hyperplanes orthogonal to the simple roots α_I and are accordingly named fundamental Weyl reflections w_I. Applied to the gauge embedding vector M_k, which can be expanded in terms of the $\mathrm{E}_8 \times \mathrm{E}_8$ simple roots, the transformation is defined as,

$$w_I(M_k) \;=\; M_k - 2\,\frac{(M_k \cdot \alpha_I)}{\alpha_I \cdot \alpha_I}\,\alpha_I\;, \tag{4.9}$$

for $I = 1, \ldots, 16$. Then, with $(\alpha_I)^2 = 2$ one obtains

$$w_I(M_k) \cdot w_I(M_\ell) \;=\; M_k \cdot M_\ell\;. \tag{4.10}$$

Hence, Weyl reflections leave the modular invariance conditions eq. (2.46) invariant if each M_k is simultaneously transformed by the same Weyl reflection w_I. Note that this allows constructing an element of the Weyl group such that the first and second E_8 transform differently. Furthermore, one can show that these are symmetries of the full string theory

$$M' \;=\; w_I(M) \quad\Rightarrow\quad \mathrm{model}\big(M\big) \;=\; \mathrm{model}\big(M'\big)\;, \tag{4.11}$$

where w_I acts simultaneously on all shift vectors and Wilson lines encoded in M. Hence, the gauge embedding matrices M and $M' = w_I(M)$ are equivalent for all Weyl reflections.

In more technical terms, the fundamental Weyl reflections w_I of the Weyl group divide the root space into different so-called Weyl chambers. These are open cones that can be transformed into each other by applying a series of fundamental Weyl reflections w_I. A special one among these is the *fundamental Weyl chamber*, which is framed by the fundamental Weyl reflections. A consequence of this is that a vector that lies in the fundamental Weyl chamber must have positive Dynkin labels. Dynkin labels are defined as the coefficients of the expansion in the dual basis eq. (2.28), and hence are given by d_{kI}/N_k in our case. To extract the values of the Dynkin labels, one can use the relation between the simple root basis and the dual basis $\alpha_I^* \cdot \alpha_J = \delta_{IJ}$. Hence, a 16-vector M_k lies in the fundamental Weyl chamber if

$$M_k \cdot \alpha_I \geq 0 \quad \forall\, I\;. \tag{4.12}$$

Each fundamental Weyl reflection maps a vector of the current Weyl chamber into another chamber if the vector does not lie on the boundary of these two chambers.

For the $\mathrm{E}_8 \times \mathrm{E}_8$ root lattice the generated group has $\mathcal{O}(10^{17})$ elements. Consequently, the Weyl group of $\mathrm{E}_8 \times \mathrm{E}_8$ yields an enormous redundancy between physically equivalent models in the heterotic orbifold landscape. In the upcoming analysis, this redundancy is dealt with in two different ways:

- In section 5.1.1: The issue is tackled by constructing an approximate representation of the gauge embedding M that is completely invariant under the symmetries of the landscape.
- In section 6.1.1: The search space is restricted to an extended definition of the fundamental Weyl chamber towards the gauge embedding matrix M, i.e. it is in general not possible to simultaneously map several 16-vectors M_k to the fundamental Weyl chamber. This technique allows searching directly in a search space that holds only one representative M of all Weyl reflected duplicates of M in the landscape.

Chapter 5

Clustering of MSSM-like models

In this section, autoencoder neural networks (see appendix C.7.2 for details) are applied to the string landscape in order to discover novel patterns. This is done in the $\mathbb{Z}_6$-II (1,1) orbifold geometry. In the past, physical intuition led to the so-called $\mathbb{Z}_6$-II (1,1) Mini-Landscape [75, 79]. In this spirit, the aim is to discover new patterns that lead to fertile islands in the landscape. Since the patterns observed by the neural network method are obtained by studying a dataset, the concept is highly transferable to other orbifold geometries. This generalization property is not given for the physical intuition approach, since the special property that in $\mathbb{Z}_6$-II (1,1) the matter at the fixed points of the first twisted sector is not subject to further orbifold projections is crucial (see section 2.4.3).

The basic working principle of autoencoder networks is to generate a non-linear map onto a lower-dimensional space, so-called latent space. This dimensional reduction forces the network to uncover correlations inside the dataset at hand, such that the projection into the lower dimensional space does keep as much information as possible. The arrangements made in order to keep the approach transferable to other orbifold geometries $\mathbb{O}$ is to start with a coarse sample that is not computationally intensive. From this dataset, patterns are obtained, which are orbifold specific. Hence to transfer the results to other orbifold geometries, an orbifold specific new dataset is needed. In order to evaluate the extrapolation properties of the found patterns, a large scan in the $\mathbb{Z}_6$-II (1,1) is performed and discussed.

5.1 Dataset of inequivalent models in the landscape

For the generation of data, we use the `orbifolder`. A sample of $\mathcal{O}(7\ 000\ 000)$ inequivalent models from the random sampling method depicted in fig. 4.1 is obtained. However, as discussed in section 4.2, several symmetries of the landscape exist. Unfortunately, these symmetries generate numerically different representations in $\mathbb{Z}^{128}$ of the same physics. This leads to a challenging task for the autoencoder.

To conjecture the effect of the symmetries, an autoencoder neural network is trained on the $\mathbb{Z}^{128}$ representation of the compactification parameters. After the training, we encode multiple Weyl reflected duplicates of a single model M into the latent space. The duplicates start to populate the whole latent space. The encoding of the autoencoder is not informed about this ambiguity and hence could not prepare for this case. This test verifies the assumption and shows that separate treatment of the landscape symmetries is necessary to get a physically meaningful lower-dimensional space.

However, since the number of possible duplicates is enormous, e.g., the Weyl group gives rise to $\mathcal{O}(10^{17})$ duplicates for each model, it is not practical to inform the neural network about symmetries by including transformed gauge embeddings M into the dataset.

Indeed, the strategy is to transform the compactification parameters into a representation that is invariant under the symmetries of the landscape. This representation is based on the local GUTs, section 2.4.3, of the first twisted sector of $\mathbb{Z}_6$-II (1, 1). The details are explained in the upcoming subsections.

5.1.1 Invariant feature representation of $\mathbb{Z}_6$-II (1,1) orbifold models

In this section, we will derive a 26 dimensional representation, i.e. $\mathbb{Z}^{26}$, that is invariant under the symmetry transformations of the landscape, described in section 4.2. This new representation will be connected to the local GUTs and is collected in a so-called feature vector F. This refers to features in machine learning that are useful quantities extracted from the raw data. The full $\mathbb{Z}^{128}$ gauge embedding M represents the raw data in our case (see appendix C). However, this leads to the drawback of an non-injective mapping D_{z_f},

$$\begin{aligned} D_{z_f} : \mathbb{Z}^{128} &\to \ \mathbb{Z}^{26} \\ M \quad &\mapsto D_{z_f}(M) = F\,, \end{aligned} \tag{5.1}$$

that takes care of the Weyl symmetry and lattice translations, i.e. $M' \mapsto D_{z_f}(M) = F$ if M and M' are related by Weyl symmetry of lattice translations. Furthermore, a second non-injective mapping D_{P},

$$\begin{aligned} D_{\mathrm{P}} : \mathbb{Z}^{26} &\to \ \mathbb{Z}^{26} \\ F \quad &\mapsto D_{\mathrm{P}}(F) = F_{\text{inv.}}\,, \end{aligned} \tag{5.2}$$

is necessary to find a unique representative under geometrical redefinition. In detail, two equivalent models under these symmetries will be mapped to the same feature vector $F_{\text{inv.}}$, while for some cases, two distinct $\mathbb{Z}_6$-II (1,1) models are mapped to the same feature vector $F_{\text{inv.}}$.

Invariance under lattice translations and Weyl reflections

To begin with, we define quantities that are invariant under the choice of $\mathrm{E}_8 \times \mathrm{E}_8$ basis vectors and the addition of arbitrary $\mathrm{E}_8 \times \mathrm{E}_8$ lattice vectors [145]. Hence, these quantities have to be manifestly invariant under Weyl transformations and the addition of lattice vectors. A quantity that is invariant under those transformations is the 4D gauge group G_{4D}. Furthermore, for $\mathbb{Z}_6$-II (1,1) orbifolds the 4D gauge group can be split into a 4D gauge group for each E_8 factor $G_{4D} = G_{4D}^{(1)} \times G_{4D}^{(2)}$. In order to have a numerical representation of this group we decide to compute the number of unbroken roots p that build the root systems of the surviving gauge groups that combine to G_{4D}. This is related to eq. (2.50) and we can compute the number of unbroken roots of G_{4D} as,

$$N_{\text{ur}}^{(\alpha)}(M) \;=\; \sum_{p\,\in\,\Phi_{\mathrm{E}_8}} \left(\prod_{k\,=\,1}^{n=8} \delta\left(p \cdot M_k^{(\alpha)}\right) \right) \;, \quad \delta(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Z} \\ 0 & \text{else} \end{cases} \tag{5.3}$$

for $\alpha = 1, 2$ corresponding to both E_8 factors and Φ_{E_8} is defined as the root system of E_8 with 240 roots p. Furthermore, the subindex "ur" in eq. (5.3) denotes *unbroken roots*.

As this mapping alone would destroy too much information about the inner structure of the orbifold we extend this concept towards the local GUTs. G_{4D} is given by the intersection of the local GUTs at the fixed points of the $T_{(1)}$ twisted sector of the $\mathbb{Z}_6$-II (1,1) orbifold [70]. Thus, we additionally consider these 12 local GUTs.

Therefore, we start with the local shift vectors V_h of these fixed points. From section 2.2.2 we recall the geometrical constraints for $\mathbb{Z}_6$-II (1,1) that are summarized in table 2.2. Due to these constraints, we have four independent 16-vectors in the $\mathbb{Z}_6$-II (1,1) orbifold, namely the shift V and the Wilson lines W_3, W_5 and W_6, which build the local shift vectors. In detail, for each constructing element $g_a = (\theta, n_\beta^{(a)} e_\beta)$ the corresponding local shift vector is given by,

$$V_{g_a} \; = \; V + (n_3^{(a)} + n_4^{(a)})\, W_3 + n_5^{(a)}\, W_5 + n_6^{(a)}\, W_6 \; . \tag{5.4}$$

This 16-vector is again split into two 8-vectors $V_{g_a} = (V_{g_a}^{(1)}, V_{g_a}^{(2)})$ corresponding to the first and second E_8 factor. For each of the 24 local GUTs $G_a^{(\alpha)}$ with $a = 1, \ldots, 12$ and $\alpha = 1, 2$, we count the number of unbroken roots p,

$$N_{\mathrm{ur}}^{g_a^{(\alpha)}}(M) \; = \; \sum_{p \,\in\, \Phi_{\mathrm{E}_8}} \delta\left(p \cdot V_{g_a}^{(\alpha)}\right) \; , \tag{5.5}$$

in analogy to eq. (5.3). As the centralizer of the $T_{(1)}$ sector is trivial, see section 2.4.3, the number of unbroken roots can be computed only by the local shift vector $V_{g_a}^{(\alpha)}$. Finally, we obtain a 26-dimensional feature vector of integers F that is invariant under the addition of $\mathrm{E}_8 \times \mathrm{E}_8$ lattice vectors and Weyl reflections,

$$F = \begin{pmatrix} \begin{pmatrix} (0,0,0,0)^{(1)} \\ (0,0,1,0)^{(1)} \\ (0,0,0,1)^{(1)} \\ (0,0,1,1)^{(1)} \end{pmatrix} & \begin{pmatrix} (0,0,0,0)^{(2)} \\ (0,0,1,0)^{(2)} \\ (0,0,0,1)^{(2)} \\ (0,0,1,1)^{(2)} \end{pmatrix} \\ & \\ \begin{pmatrix} (1,0,0,0)^{(1)} \\ (1,0,1,0)^{(1)} \\ (1,0,0,1)^{(1)} \\ (1,0,1,1)^{(1)} \end{pmatrix} & \begin{pmatrix} (1,0,0,0)^{(2)} \\ (1,0,1,0)^{(2)} \\ (1,0,0,1)^{(2)} \\ (1,0,1,1)^{(2)} \end{pmatrix} \\ & \\ \begin{pmatrix} (1,1,0,0)^{(1)} \\ (1,1,1,0)^{(1)} \\ (1,1,0,1)^{(1)} \\ (1,1,1,1)^{(1)} \end{pmatrix} & \begin{pmatrix} (1,1,0,0)^{(2)} \\ (1,1,1,0)^{(2)} \\ (1,1,0,1)^{(2)} \\ (1,1,1,1)^{(2)} \end{pmatrix} \\ & \\ N_{\mathrm{ur}}^{(1)}(M) & N_{\mathrm{ur}}^{(2)}(M) \end{pmatrix} . \tag{5.6}$$

For clarity the 12 fixed points are represented by the parameters of the constructing element. As an example $N_{\mathrm{ur}}^{g_a^{(\alpha)}}(M)$ with $g_a^{(\alpha)} = (\theta, n_\beta^{(a)} e_\beta)^{(\alpha)}$ is represented by $(n_3^{(a)}, n_4^{(a)}, n_5^{(a)}, n_6^{(a)})^{(\alpha)}$, i.e. $g_1^{(1)} = (\theta, 0)$ corresponds to $(0,0,0,0)^{(1)}$. The specific block structure encodes the number of fixed points in the second ($\mathbb{T}^2/\mathbb{Z}_3$ with three fixed points) and third ($\mathbb{T}^2/\mathbb{Z}_2$ with four fixed points) complex plane and is of relevance in the next subsection.

This first part of the mapping, i.e. D_{z_f} eq. (5.1), to the 26-dimensional feature vectors F is not injective. From all $\mathbb{Z}_6$-II (1,1) models under consideration, i.e. $\mathcal{O}(7\,000\,000)$ inequivalent models on the level of the spectrum, the transformation of models reduces this number by 0.5% for inequivalent vectors F. However, there is still some ambiguity due to geometrical redefinition in this representation. Hence, the second mapping D_{P} is intended to sort F to a unique representation.

Fixed point permutations

As described in section 4.2, geometrical redefinitions permute the local shift vectors V_h of the fixed points. However, these transformations lead to equivalent orbifold models in terms of the particle spectrum and the full theory. Hence, we have to take care of those ambiguities as well. In the case of the first twisted sector $T_{(1)}$ of the $\mathbb{Z}_6$-II (1,1) orbifold there exists the following three symmetries:

- $\mathbb{Z}_2$ between the first and second E_8 factor.
- S_3 permutation symmetry of the three fixed points in the $\mathbb{T}^2/\mathbb{Z}_3$ plane.
- D_8 permutation symmetry of the four fixed points in the $\mathbb{T}^2/\mathbb{Z}_2$ plane.

For the purpose of an efficient computational implementation, one can formulate the symmetry actions in a computational vectorized matrix multiplication scheme. Then the $\mathbb{Z}_2$ action that interchanges the E_8 factors, is given by,

$$F \mapsto F \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \tag{5.7}$$

where the multiplication from the right hand side enables the permutation of the columns of F. The S_3 fixed point interchange on the $\mathbb{T}^2/\mathbb{Z}_3$ plane is generated by $V_1 \mapsto V_1 + W_3$ and $W_3 \mapsto W_3 + W_3$. The corresponding $\mathbb{Z}_3$ permutation and $\mathbb{Z}_2$ reflection act on F as,

$$F \mapsto \begin{pmatrix} & \mathbb{1}_4 & & \\ & & \mathbb{1}_4 & \\ \mathbb{1}_4 & & & \\ & & & 1 \end{pmatrix} F \quad \text{and} \quad F \mapsto \begin{pmatrix} \mathbb{1}_4 & & & \\ & & \mathbb{1}_4 & \\ & \mathbb{1}_4 & & \\ & & & 1 \end{pmatrix} F , \tag{5.8}$$

respectively. Finally the transformations $V_1 \mapsto V_1 + W_5$, $V_1 \mapsto V_1 + W_6$ and the interchange $W_5 \leftrightarrow W_6$ give rise to three $\mathbb{Z}_2$ generators in the $\mathbb{T}^2/\mathbb{Z}_3$ plane that combine to the action of a D_8 permutation symmetry. In terms of the feature vector F this reads as,

$$F \mapsto \begin{pmatrix} R_x & & & \\ & R_x & & \\ & & R_x & \\ & & & 1 \end{pmatrix} F , \tag{5.9}$$

with R_x, $x = 1, 2, 3$, given by the permutations of the four local shift vectors:

$$R_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} , \quad R_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad R_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{5.10}$$

To conclude, for the symmetries described above each instance F is subject to a non-Abelian permutation group $\mathbb{Z}_2 \times S_3 \times D_8$ of order $2 \times 6 \times 8 = 96$. The above considerations were used in the following implementation of a sorting algorithm for F:

1. Order the values corresponding to the fixed points, i.e. F_{ij} with $i = 0, \ldots, 12$ and $j = 0, 1$, for each j independently in ascending order to obtain F^A.

2. Compare element-wise F^A_{i1} and F^A_{i2} from $i = 0$ to 12. For the smallest value of i where $F^A_{i1} \neq F^A_{i2}$ keep the present ordering if $F^A_{1j} < F^A_{2j}$. Otherwise act with the non-trivial $\mathbb{Z}_2$ on F and interchange the E_8 factors, e.g. like in eq. (5.7). This method is a computationally efficient approximation of dividing out the $\mathbb{Z}_2$ of the E_8 factors. Note that there exist some rare corner cases, i.e. $N_{\text{cases}} = 9\,326$ out of the $\mathcal{O}(7\ 000\ 000)$, where both ascended sorted E_8 vectors are equal. There remains the possibility that those cases have a definite order in their E_8 factors, if the constrained positioning of the F entries, i.e. $S_3 \times D_8$ symmetry, is considered in detail. However, for computational reasons we will exploit this speed up and accept these rare occasions that will be part of our dataset in both configurations.

3. Multiply out all 48 copies with $S_3 \times D_8$ acting on F. Than all transformed versions of F are compared and only the one F' that is smaller then all other transformed images $F^{''}$ of F, i.e. $F' < F''$ is kept. F' and F'' indicate two different transformation of F by nontrivial elements of $S_3 \times D_8$. We define the metric for $F' < F''$ according to $F'_{i1} < F''_{i1}$ for the smallest value of i. In case $F'_{i1} = F''_{i1}$ a tiebreak is taken into account, i.e. the second E_8 factor is considered $F'_{i2} < F''_{i2}$. If neither of these two decision conditions is positive, i.e. the values are equal, the next value of i is considered until the smaller F of this comparison of two images is found. This is done for the set of all 48 copies of F.

4. For the special cases where $F_{i1} = F_{i2}$ for $i = 0, \ldots, 12$ we act again with $\mathbb{Z}_2$ of E_8 such that the smaller 4D gauge group $F_{13\,j}$ is assigned to $F_{13\,1}$. Otherwise, the E_8 ordering of step 2 is kept.

This results in the feature vector $F_{\text{inv.}}$ which is a special choice to reduce the symmetries acting on the objects described above. The combined action of D_{z_f} and D_{P} on the $\mathcal{O}(7\ 000\ 000)$ models yield 84% distinct feature vectors $F_{\text{inv.}}$.

5.1.2 Encoding of integer valued data

The transformations eq. (5.1) and eq. (5.2) consider the symmetries and the feature vectors $F_{\text{inv.}} \in \mathbb{Z}^{26}$ are ready for use in the machine learning approach. However, it turns out that the autoencoder does not perform well on this representation. This issue may arise from the fact that the target set of values from the number of unbroken roots eq. (5.5) and eq. (5.3) is integer valued. This complication for the machine learning approach refers to integer programming or constrained optimization.

In constrained optimizations, the optimization procedure to find a solution is even more difficult. An auxiliary condition, here, the restriction of the solution to integers, has to be fulfilled too. A common way to improve the performance on such problems is given by a particular encoding procedure, so-called one-hot encoding `OHEnc` (see eq. (C.2)). This technique is commonly used in classification tasks to encode abstract types of classes numerically. Even though in the case of the $\mathbb{Z}^{26}$ representation, the different values are already numerical, one-hot encoding is a common strategy to express integer data (e.g., used in PixelCNN [185]). It turns out that this also increases the performance of the autoencoder in this unsupervised machine learning task.

To summarize, the 26 features are each separately one-hot encoded. Each entry of the feature vector $F_{\text{inv.}}$ transforms into a 37-dimensional vector, where the 37 entries refer to the 37 unique values that can appear in the whole dataset D regardless of their position in $F_{\text{inv.}}$. This strategy is chosen to keep the best generalization of this approach towards new data. In a traditional one-hot encoding approach, each feature is encoded separately into an n-dimensional vector where n is dependent on the number of categoricals that appear for a certain feature. As we encode each feature with the maximal number of inequivalent integers that appeared, we have a better generalization towards new models with different local GUTs, i.e. new values for $N_{\text{ur}}^{g_a^{(\alpha)}}$. This is necessary as the one-hot encoding, and the autoencoder can not dynamically adjust to new input values and dimensions, after training. This will get handy when we map models from other machine learning approaches (see chapter 8) into the same latent space, i.e. the latent space we train with the autoencoder in the next section. Hence, the input for the autoencoder is a $26 \times 37 = 962$ dimensional vector $F_{\text{inv.}}^{\texttt{OHEnc}}$.

5.2 Dimensional reduction via autoencoder

After the preparation in the previous section the data is ready to be used in a deep autoencoder neural network [186]. The data $|D| \sim \mathcal{O}(7\,000\,000)$ is first split into $|D^{\texttt{AE}}| \sim \mathcal{O}(700\,000)$ and $|D^{\texttt{test}}| \sim \mathcal{O}(6\,300\,000)$. Thus, $D^{\texttt{AE}}$ represents the coarse sample that contains only 10% of the whole dataset. Then the data for the autoencoder construction $|D^{\texttt{AE}}|$ is split into $|D^{\texttt{train}}| \sim \mathcal{O}(400\,000)$ and $|D^{\texttt{val}}| \sim \mathcal{O}(300\,000)$. The machine learning background of this data splitting is discussed in appendix C.

For a short introduction in the generic properties of neural networks see appendix C.7 and in particular for details on autoencoders appendix C.7.2. For the ongoing discussion, these basics are less important, and only some details on the concrete autoencoder model and training are part of the main text. The purely physics interested reader can take the viewpoint of an autoencoder as an algorithm to derive a deterministic highly non-linear projection from the 26-dimensional data space into a lower-dimensional space. In this particular application the lower dimensional latent space, is fixed to 2-dimensions. This constraint arises from the requirement to get an interpretable and visualizable, lower-dimensional representation of the data. Higher values of this latent space would demand the use of further machine learning techniques in order to handle the new lower-dimensional space. This additional part of abstraction through machine learning should be avoided. Also, our experiments showed that a higher dimension of the latent space does not significantly increase the information stored in the latent dimension.

The autoencoder is implemented using the TensorFlow [187] package. In order to obtain a maximally informative 2-dimensional representation, the architecture (layer size and number of hidden layers) is varied. The manual selection of the best architecture results in fully connected layers with seven hidden layers. The activation function of the hidden layers is the `SELU` activation function [188] that automatically accounts for batch normalization and hence makes the training process faster. For the latent layer, as well as the output layer, the identity activation function (id) is chosen. In summary, we have two mappings,

$$\texttt{AE}_{\text{encoder}}: \quad F_{(962)} \to \texttt{SELU}(F_{(100)}) \to \texttt{SELU}(F_{(26)}) \to \texttt{SELU}(F_{(13)}) \to \text{id}(F_{(2)}) \tag{5.11}$$

$$\texttt{AE}_{\text{decoder}}: \quad F_{(2)} \to \texttt{SELU}(F'_{(13)}) \to \texttt{SELU}(F'_{(26)}) \to \texttt{SELU}(F'_{(100)}) \to \text{id}(F'_{(962)}) \tag{5.12}$$

where '→' is a linear mapping and `SELU` (id) act element-wise. The subindex of F gives the dimension of the space. Note that $\texttt{AE}_{\text{encoder}}$ and $\texttt{AE}_{\text{decoder}}$ are independent functions with independent parameters, i.e. $F_{(x)}$ and $F'_{(x)}$ for $x \in \{13, 26, 100, 962\}$ have the same dimension but can describe different representations. The objective function for training is given by the L_2 loss,

$$L_2 = \frac{1}{|\texttt{B}|} \sum_{F_{\text{inv.}} \in \texttt{B}} ||F_{\text{inv.}}^{\texttt{OHEnc}} - \texttt{AE}_{\text{decoder}} \left(\texttt{AE}_{\text{encoder}} \left(F_{\text{inv.}}^{\texttt{OHEnc}}\right)\right)||_2 \,, \tag{5.13}$$

where $|| \cdot ||_2$ is the euclidean norm and `B` is a so-called batch, i.e. a partition of the whole dataset. Then, the autoencoder is trained iteratively on all batches `B` of $D^{\texttt{train}}$ to minimize the L_2 loss, until the loss value converges on the computational time scales. This convergence was also visually controlled by plotting the latent space, like in fig. 5.1, during training. Here, the emergence of clusters can be traced, and the freezing of those movements was obtained.[1] Also, through several re-trainings with different weight initializations, it was verified that the process is stable and gives very similar clustering maps. The reproducibility of very similar latent space structures refers to the fact that our model has a small variance but a significant bias and supports the interpretation of the 2-dimensional bottleneck as a strong regularization. During training, the autoencoder was applied to the $D^{\texttt{val}}$ set to keep track of overfitting of the training data $D^{\texttt{train}}$. However, the number of correctly reproduced features of the 26-dimensional representation did not decrease at any time of the training for $D^{\texttt{val}}$. The lack of capacity to overfit is most likely related to the strong regularization of the autoencoder, i.e. the small latent space dimension. An exact reproduction of any model, even for training data, seems to be impossible with such a strong constraint for the neural network. The final decoder can reproduce on average 16.3 out of 26 features $F_{\text{inv.}}$, which corresponds to $L_2 = 0.013$ on $D^{\texttt{val}}$. While this reproduction rate is astonishing, it is still far from a perfect reproduction and underlines the strong regularization of the latent space dimensions.

5.2.1 A chart of $\mathbb{Z}_6$-II (1,1) models and cluster selection

For visual analysis of the projection, all $\mathcal{O}(700\,000)$ models of the $D^{\texttt{train}}$ and $D^{\texttt{val}}$ set are transformed by the encoder mapping, $\texttt{AE}_{\text{encoder}}$, to obtain their two-dimensional representation from the latent layer. The results are illustrated in fig. 5.1. The projection of the landscape turns out to cluster in various isolated islands. The 18 MSSM-like models in $D^{\texttt{AE}}$ are emphasized by red triangles in fig. 5.1. Interestingly, the MSSM-like models accumulate in the lower region and do not separate over the entire chart. Note that the training was unsupervised. Still, the MSSM-like $\mathbb{Z}_6$-II (1,1) models indicate that physically similar models cluster. Hence, it seems that the autoencoder was able to identify common properties among the models in the 26-dimensional representation that correlate with physical properties.

Consequently, the next step is to select those islands in fig. 5.1 that contain at least one MSSM-like model. This selection is made by inspection by eye instead of applying any clustering algorithm. The qualitative criteria of the selection are to collect only models in a cluster that are close by and to keep the clusters as small as possible, in order to reduce the landscape significantly. The constraint to project into a 2-dimensional space allows for this simplification. The different shapes and distances between the clusters would imply a machine learning problem by itself. In total, this selection results in eleven fertile islands. In the following, the models cluster affiliation is used as a classification label, to find a refined search strategy for MSSM-like $\mathbb{Z}_6$-II (1,1) models.

5.2.2 Knowledge extraction using a decision tree

The projection towards the latent space $\texttt{AE}_{\text{encoder}}$ can be used as a filter for orbifold constructions, to predict, if they can lead to MSSM-like models or not. The primary goal of this section is to extract knowledge from the clustering. Therefore, we fit a decision tree to the 26-dimensional, i.e. not one-hot encoded, representation $F_{\text{inv.}}$ with the corresponding cluster labels R_i as the

---

[1] The video is available at `http://users.ph.tum.de/ga74vir/AutoencoderTrainingVid.html` or by contacting the author.

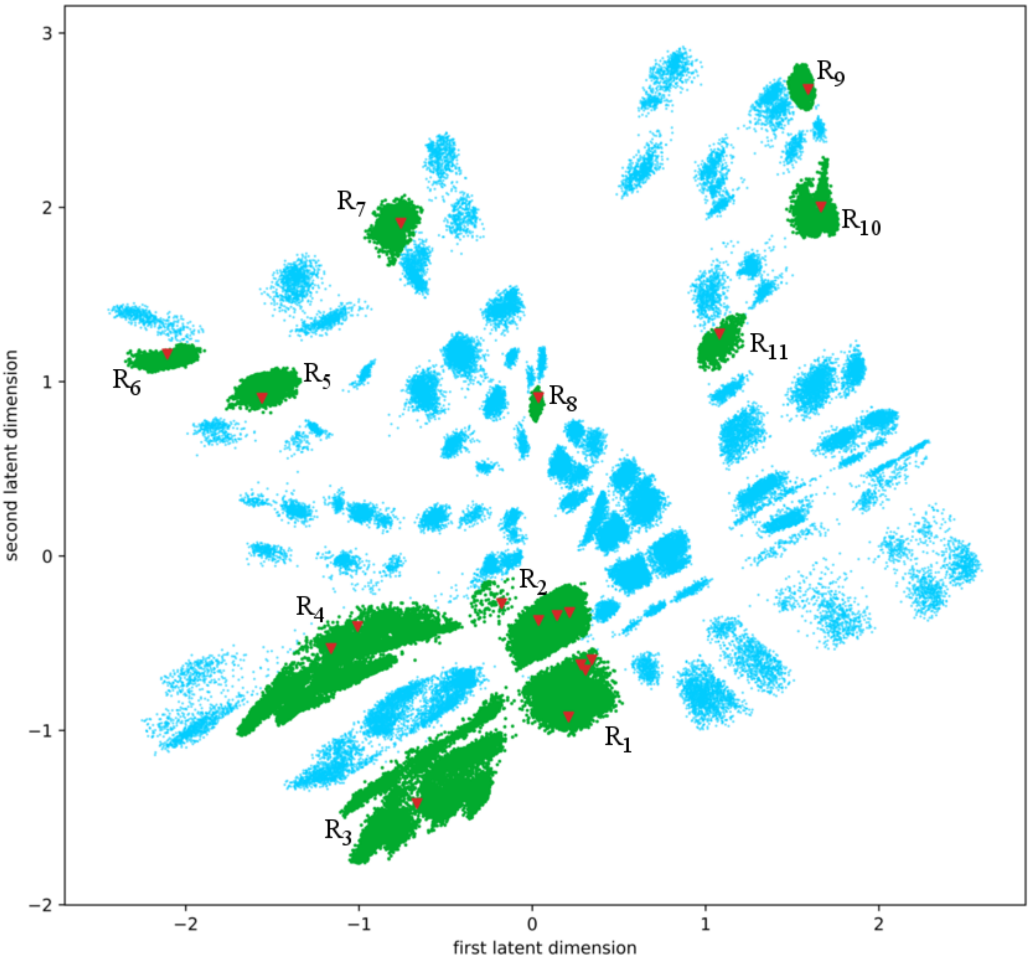


Figure 5.1: *Projection of the symmetry invariant representation* $F_{\text{inv.}}^{\texttt{OHEnc}}$ *to the autoencoder latent space. The* $\mathcal{O}(700\,000)$ *points correspond to* $\mathbb{Z}_6$-II (1,1) *models. The special MSSM-like models are highlighted as red triangles. It appears that the MSSM-like models populate eleven isolated islands. We color these islands in green and label them by* R_i *for* $i = 1, \ldots, 11$*. The* $\mathbb{Z}_6$-II (1,1) *models outside these islands are colored in blue and defined to reside in the region* R_0*. Figure adapted from [23].*

target values. In this context, the unsupervised clustering can be seen as a method to generate meaningful labels, i.e. less restrictive than the binary label MSSM-like or not MSSM-like (~~MSSM~~-like). It is possible to find rules based on these softened labels.

Therefore, the classical decision tree algorithm is chosen due to the high interpretability of this technique. For details on the working principle and further explanations on decision trees see appendix C.6. The trained decision tree $\texttt{DT}_{\texttt{Cluster}}$

$$\texttt{DT}_{\texttt{Cluster}} : \mathbb{Z}^{26} \to \{\mathrm{R}_i\}_{i=0}^{11} , \tag{5.14}$$

can predict the region R_i a given $\mathbb{Z}_6$-II (1,1) model belongs to. The tree only uses individually simple and interpretable threshold conditions on the 26 features, i.e. $(F_{\text{inv.}})_{ij} < t$ with $t \in \mathbb{Z}$. For the decision tree, the software scikit-learn [189] is used. As before the $\mathcal{O}(700\ 000)$ random models are split into train and validation set. This time 33% of the models are assigned to the validation set $D^{\texttt{val}}$. Additionally, the individual data points are weighted such that the

DT	depth	# nodes	f1-macro
A	6	119	0.503
B	10	599	0.693
C	15	1257	0.887
D	25	1787	0.993
E	50	1887	0.995

Table 5.1: *Table of trained decision trees (*`DT`*) with regularized depth of training. Note it was beneficial for the classification task to give individual labels to the fertile patches, i.e. for a binary classification 'fertile vs. R_0' the decision tree performed slightly worse for the same hyperparameter. For the definition of the* `f1-macro` *performance measure see appendix C.5.*

classification task is balanced according to the label of the regions R_i. Thus, the clear numerical superiority of the blue region R_0 does not bias the decision tree. Several hyperparameters, i.e. constraints on the complexity of the decision tree, were tried, and the resulting performance can be found in table 5.1. Unfortunately, it turned out that in order to get a very good prediction of the corresponding region R_i the tree needs a lot of splits. This resulted in a decision tree with 1 887 nodes. The performance on $D^{\texttt{val}}$ estimates how well the rules found by the tree generalize to the whole $\mathbb{Z}_6$-II (1,1) landscape. Details of the misclassifications can be extracted from the so-called confusion matrix given in table 5.2 (see appendix C.5 for details on this concept to analyze classification tasks). A scalar value to estimate the performance is provided by `f1-macro`=0.995, which is close to one and indicates a very good classification performance. Unfortunately, the simple decisions in such a huge tree are branched in a complex way and it is not possible to extract easy rules. However, one advantage of the very good performance is that $\texttt{DT}_{\text{Cluster}}$ acts directly on the 26-dimensional representation. This avoids the one-hot encoding and gives a very fast prediction.

5.3 Evaluation of the landscape projection

So far the machine learning workflow and the performance of the algorithm on the validation set $D^{\texttt{val}}$ were analyzed. Now the lab environment is left and the generalization to the whole $\mathbb{Z}_6$-II (1,1) landscape is the goal of this section. Especially the behavior of the MSSM-like models from the whole $\mathbb{Z}_6$-II (1,1) landscape is of interest. In detail, on a qualitative level, we ask if the MSSM-like models cluster on the fertile islands. On a quantitative level, it is interesting how many MSSM-like models within the whole $\mathbb{Z}_6$-II (1,1) landscape are projected onto the eleven fertile islands. This result implies the number of models that are neglected if the search restricts to fertile islands. For this evaluation, we analyze two kinds of test sets. First, the evaluation set $D^{\texttt{test}}$ containing $\mathcal{O}(6\,300\,000)$ $\mathbb{Z}_6$-II (1,1) models. Hence, it provides a much larger and refined brute force search of the landscape. These models are not only physically inequivalent towards each other but also physically inequivalent to the models used for training. Secondly, the dataset $D^{\texttt{test}}_{\text{MiniL}}$ of $\mathcal{O}(30\,000)$ $\mathbb{Z}_6$-II (1,1) models from the four patches of the Mini-Landscapes [75, 79].

The evaluation set $D^{\texttt{test}}$ contains 177 MSSM-like models in contrast to the 18 MSSM-like models that where used to classify the fertile islands and label the $\mathbb{Z}_6$-II (1,1) models accordingly. The mapping of these models into the 2-dimensional projection is plotted in fig. 5.2.

		predicted region											
		R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}
true region	R_0	198 994	10	39	10	24	1	7	17	3	16	4	13
	R_1	11	3 107	1	2	0	0	0	0	0	0	0	0
	R_2	19	3	9 667	2	1	0	0	0	0	0	0	0
	R_3	24	2	1	5 256	3	0	0	0	0	0	0	0
	R_4	31	2	4	1	6 430	0	0	0	0	0	0	0
	R_5	0	0	0	0	0	3 138	0	0	0	0	0	0
	R_6	3	0	0	0	0	0	994	0	0	0	0	0
	R_7	15	0	0	0	0	0	0	848	0	0	0	0
	R_8	0	0	0	0	0	0	0	0	1 139	0	0	0
	R_9	10	0	0	0	0	0	0	0	0	1 491	0	0
	R_{10}	2	0	0	0	0	0	0	0	0	0	3 333	0
	R_{11}	10	0	0	0	0	0	0	0	0	0	0	984

Table 5.2: *The confusion matrix* C *(appendix C.5) of our decision tree* $\texttt{DT}_{\texttt{Cluster}}$ *eq.* (5.14) *evaluated for the validation set* D^{val}. *The entries give the number of* $\mathbb{Z}_6$-II (1,1) *models that are predicted by* $\texttt{DT}_{\texttt{Cluster}}$ *to lie in the 'predicted region' but actually belong to the 'true region'. For example, there are 11 cases where* $\texttt{DT}_{\texttt{Cluster}}$ *predicted a model to be in region* R_0, *while the true region was* R_1.

	region	coarse sample $D^{\texttt{AE}}$	evaluation set $D^{\texttt{test}}$	total
	R_0	0	65	65
fertile islands	R_1	4	44	48
	R_2	4	17	21
	R_3	1	10	11
	R_4	2	16	18
	R_5	1	5	6
	R_6	1	2	3
	R_7	1	1	2
	R_8	1	1	2
	R_9	1	0	1
	R_{10}	1	11	12
	R_{11}	1	5	6
	total	18	177	195

Table 5.3: *Classification of MSSM-like* $\mathbb{Z}_6$-II (1,1) *models from either the coarse sample* D^{AE} *or from the evaluation set* D^{test} *as predicted by the decision tree* $\texttt{DT}_{\texttt{Cluster}}$ *eq.* (5.14).

Astonishingly, the majority of MSSM-like models are joining the known MSSM-like models inside the fertile islands. For a more quantitative statement the decision tree is applied to all $177 + 18 = 195$ MSSM-like models to obtain the cluster predictions given in table 5.3. The complementary group of MSSM-like models in the evaluation set, that are classified by the decision tree to region R_0, i.e. the blue region in the projection labeled as ~~MSSM~~-like , would be missed if the search focuses only on the defined fertile islands. However, overall the decision tree $\texttt{DT}_{\text{Cluster}}$ maps 130 of all 195 MSSM-like models to the fertile islands. Therefore, the predictive model based on a small set of only 18 MSSM-like models, can extrapolate, to reach 2/3 of the

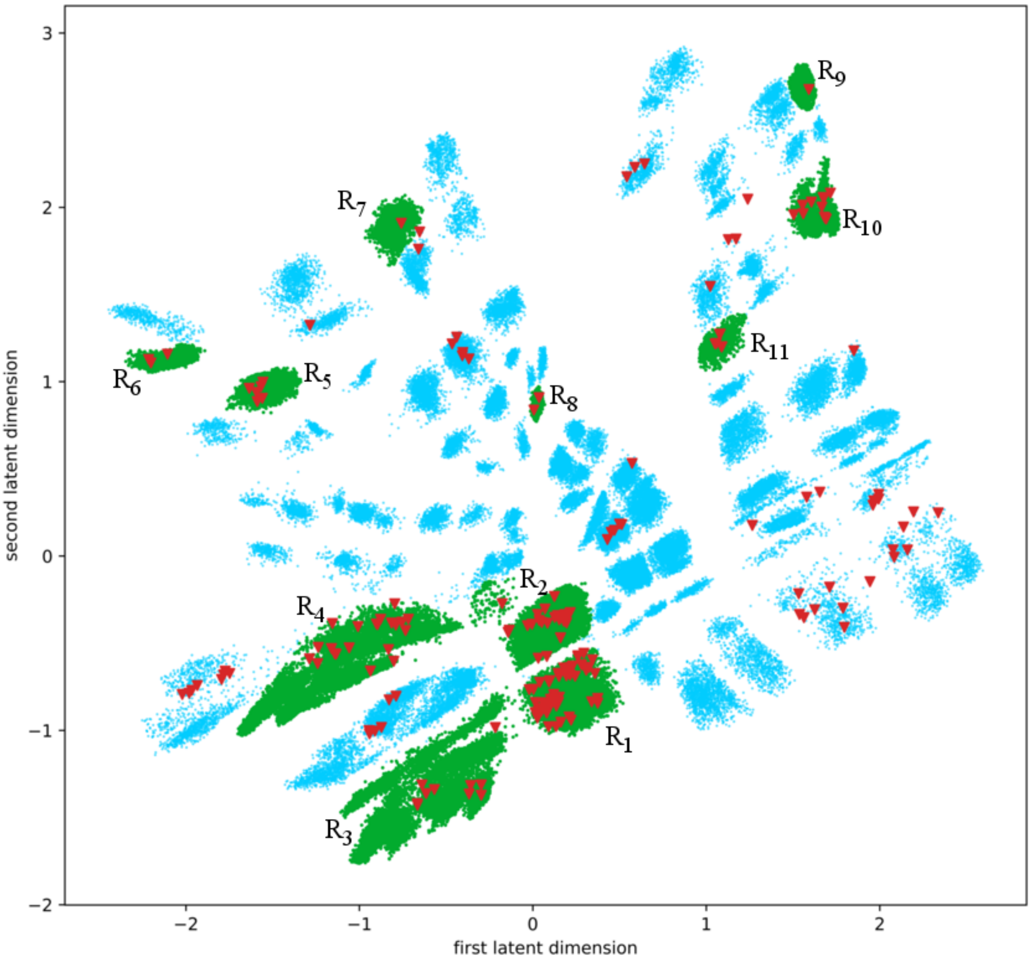


Figure 5.2: *Projection of all 195 MSSM-like models (red triangles) from the evaluation set* D^{test} *and the coarse sample* D^{AE} *to the latent space. We color in green the eleven fertile islands* R_i*, and in blue the whole* $\mathbb{Z}_6$-II (1,1) *landscape. It turns out, that the fertile islands generalize to the whole set of MSSM-like models. Figure adapted from [23].*

MSSM-like models. Mainly the fertile island R_1 contains in total 48 MSSM-like models, i.e. 25% of all MSSM-like models, yet this island contains only 1.3% of the whole $\mathbb{Z}_6$-II (1,1) landscape. Thus, the fertile islands provide a search strategy with a 20 times increased chance to find MSSM-like models.

Another critical question is how this data-driven approach connects to the Mini-Landscape found in [75, 79]. In fig. 5.3, one can observe that the MSSM-like $\mathbb{Z}_6$-II (1,1) models from all four different Mini-Landscapes do not distribute over the whole chart. Instead, they are concentrated on the machine learning fertile islands. Let us also analyze the performance of our decision tree $\texttt{DT}_{\text{Cluster}}$ on the MSSM-like $\mathbb{Z}_6$-II (1,1) models of the Mini-Landscape. Table 5.4 shows that nearly 2/3 of the MSSM-like $\mathbb{Z}_6$-II (1,1) models from the Mini-Landscape accumulate in the fertile islands. Intriguingly, the classification indicates a close connection between the two SO(10) patches of the Mini-Landscape and the R_1, R_2, and R_3 islands. In particular, models with shift vector $V^{\text{SO}(10),1}$ are most likely located on the island R_2, while the islands R_1 and R_3 hold most of the models with shift vector $V^{\text{SO}(10),2}$.

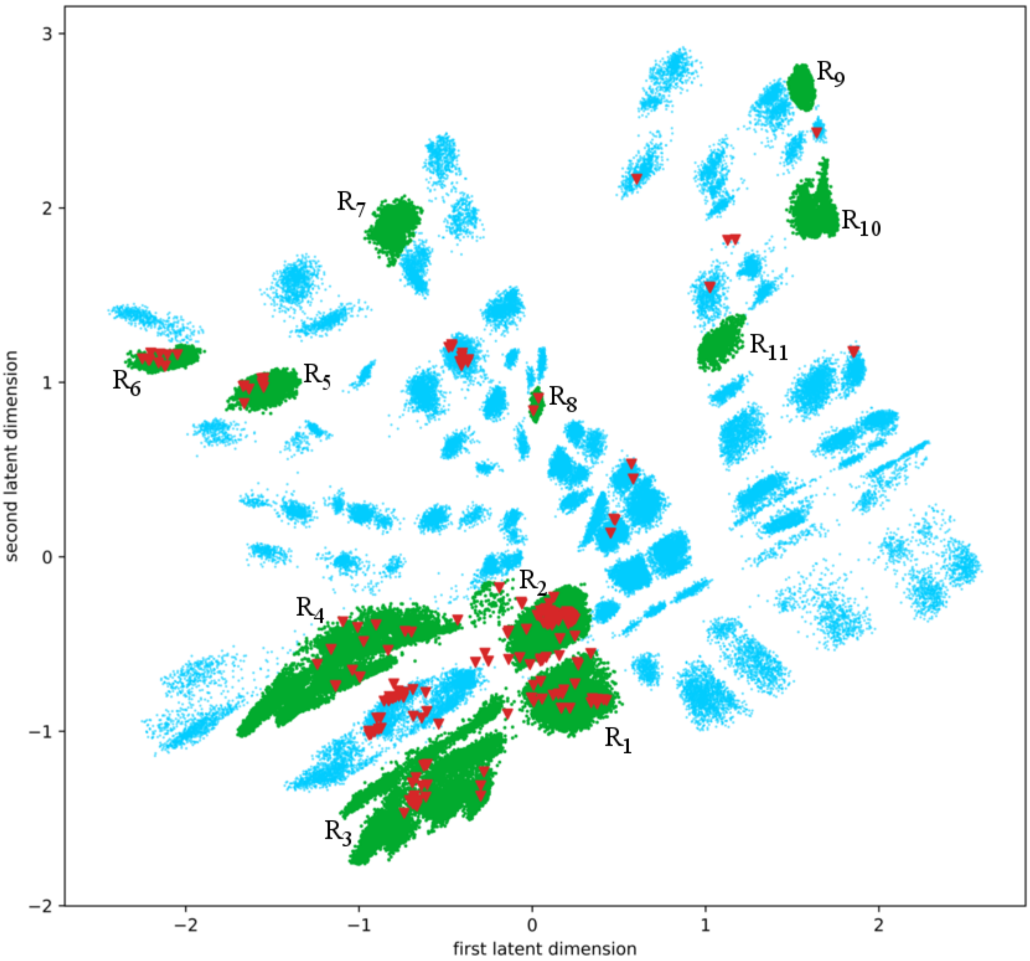


Figure 5.3: *Projection of the Mini-Landscape MSSM-like models (red triangles) to the latent space. We color in green the eleven fertile islands* R_i*, and in blue the whole* $\mathbb{Z}_6$-II (1,1) *landscape. Similar to fig. 5.2, the MSSM-like models from the Mini-Landscape accumulate in the fertile islands. In particular, the islands* R_1, R_2 *and* R_3 *contain many MSSM-like models from the Mini-Landscape. Figure adapted from [23].*

5.4 Chapter summary

In this chapter, we showed how patterns in the heterotic orbifold landscape are obtained from a general data level. The method is transferable to other orbifold geometries $\mathbb{O}$ and would lead to geometry specific fertile islands. This geometry dependence also holds for the transformations eqs. (5.1) and (5.2) towards the symmetry invariant fixed point local GUTs representation $F_{\text{inv.}}$. The representation then differs in the number of fixed points and the types of landscape symmetries. Note, there exist geometries that allow for brother models. These models differ in their particle spectrum but relate to each other by a lattice transformation. Thus, the mapping towards the local GUTs can not distinguish between brother models as these quantities are invariant under the addition of lattice vectors. The information loss, i.e. the non-injective property of the mapping, would increase for this kind of geometries. Finally, this part of the method is not data-driven and hence requires to be engineered by hand for most of the geometries.

Furthermore, one should be aware that a significant part of the clustering arises from the

	region	$V^{\mathrm{SO(10)},1}$	$V^{\mathrm{SO(10)},2}$	$V^{\mathrm{E}_6,1}$	$V^{\mathrm{E}_6,2}$
	R_0	50	37	2	1
fertile islands	R_1	12	16	0	0
	R_2	60	1	0	0
	R_3	2	24	2	0
	R_4	3	8	4	0
	R_5	10	0	0	0
	R_6	0	8	0	4
	R_7	0	0	0	0
	R_8	0	1	0	1
	R_9	0	0	0	0
	R_{10}	0	0	0	0
	R_{11}	0	0	0	0
	% found	64%	61%	75%	83%

Table 5.4: *Classification of Mini-Landscape MSSM-like* $\mathbb{Z}_6$-II (1,1) *models as predicted by the decision tree* $\mathtt{DT}_{\mathtt{Cluster}}$ *eq.* (5.14)*. They are grouped according to the patches of the Mini-Landscape, i.e. in terms of the local GUT shift vectors* $V^{\mathrm{SO(10)},1}$, $V^{\mathrm{SO(10)},2}$, $V^{\mathrm{E}_6,1}$ *and* $V^{\mathrm{E}_6,2}$, *see [75].*

particular 26-dimensional representation $F_{\text{inv.}}$. For the autoencoder eqs. (5.11) and (5.12), a projection into the 2-dimensional plane must carry information about the 26-dimensional numerical structure in order to reconstruct as many features as possible. Hence, since the MSSM-like models cluster, they should have a similar numerical structure in the 26-dimensional representation. Therefore, investigating the 26-dimensional representation with competitive projection algorithms, e.g. t-SNE [190], or direct exploration in the 26-dimensional space, can lead to different insights about the MSSM-like fixed point structure.

Connected to this, let us comment on the possibility of drawing parameters from the 2-dimensional plane and reconstructing the corresponding gauge embedding matrices M. While this procedure is very appealing, there are two main obstacles. First, the autoencoder has to be improved significantly, i.e. the reproduction rate has to be increased to decode the latent space into accurate representations of the 26-dimensional space. Secondly, the mapping from the 26-dimensional representation to the matrices M is not unique. Recall that the mapping is not injective. Thus, to find a corresponding matrix M is computationally costly as multiple realizations can correspond to the same $F_{\text{inv.}}$ representation.

Finally, we want to point out that it can be beneficial for the clustering of the MSSM-like models, to introduce a latent loss [191, 192]. This concept informs the neural network about the physical model behind the feature vector. Thus, the autoencoder could focus on clustering physically and numerically similar models. Caution is necessary for the definition of the latent loss metric. This measure should be generic in the distinction between particle spectra. A trivial latent loss, e.g. a metric based on the number of generations, will quickly be circumvented by the neural network, which directly memorizes the small set of MSSM-like models that are contained in the training set.

Chapter 6

The phenomenologically viable part of the landscape

Analyzing the decision tree $\mathtt{DT}_{\text{Cluster}}$ of chapter 5 reveals that the learned conditions are very advanced and tightly coupled. Further investigations, based on the spectrum of the data, lead to the insight that many models do not fulfill basic phenomenological properties. Concretely, they do not have a 4D gauge group G_{4D}, that can host the SM gauge group.

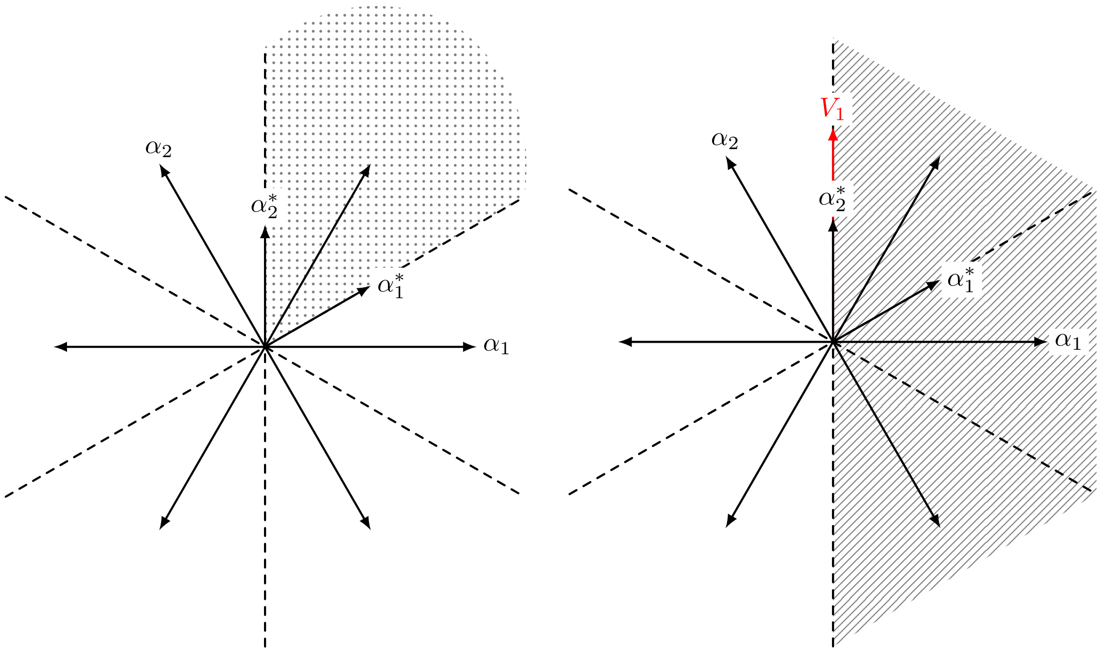
Those necessary phenomenological constraints can be implemented in the successive search fig. 4.1. Demanding the analytic constraints will reduce the variety of data and may have a data cleaning effect, such that the autoencoder clustering yields a down-scaled decision tree. However, before reconsidering this approach, the concept of the analytic constraint is extended towards a statistical analysis. The dataset is examined concerning so-called contrast patterns. Those patterns are based on statistical considerations and are used to distinguish data points from different classes. In this section, the two classes are the set of MSSM-like models and the complementary set of ~~MSSM~~-like models. It will turn out that several of those contrast patterns exist and that the successive search can be massively improved taking them into account during model construction.

6.1 Reduction of the search space

In this section, we investigate physical considerations to reduce the search space for gauge embeddings M (section 2.2.2). The first action is to divide out the Weyl symmetry from the search space, such that only one representative of all M's connected by the Weyl group remains in the minimized parameter space. In the second part of this section, we inform the search algorithm about basic phenomenological constraints, i.e. about the the necessity that a constructed physical model has to have the ability to host the gauge group of the SM.

6.1.1 Excluding redundant parameters

As described in section 4.2, the symmetries of the landscape interfere with the search goal to find many inequivalent physical models. A particularly large amount of symmetry is related to the Weyl group (section 4.2.3). Hence, a detailed understanding of the Weyl group action can be used to correlate a vast amount of parameters on the landscape. Therefore, a specialized search technique is developed, which accounts for this symmetry and restricts the search space towards a unique region of the orbifold landscape. This approach is more generic than the consideration of the Weyl group in section 5.1.1, since the restriction is directly on the search space $\mathbb{Z}^{128}$ and



(a) *Fundamental Weyl chamber of* SU(3). **(b)** *Domain for* W_1 *after a specific* V_1 *was chosen.*

Figure 6.1: *Illustration of the algorithm to divide out the Weyl symmetry from the search process. The root space of* SU(3) *serves as an example. In fig. 6.1a the fundamental Weyl reflections* w_1 *and* w_2 *define the the fundamental Weyl chamber of* SU(3) *(shaded area). Thus,* w_1 *and* w_2 *restrict the first vector* V_1 *to lie in the shaded region. Thus,* $d_{1\,I} \in \mathbb{N}_0$ *for* $I = 1, 2$. *In fig. 6.1b a specific vector* V_1 *along the direction of* α_2^* *was chosen. Consequently, the vector* V_1 *is invariant under the Weyl reflection* w_1*, i.e.* $w_1(V_1) = V_1$. *Hence, this choice for* V_1 *has broken* w_2 *of the Weyl symmetry and leaves* w_1 *as remnant. Thus, the Weyl reflection* w_1 *restricts the search space for* W_1 *to the broken Weyl chamber (shaded area) which is defined by* $W_1 \cdot \alpha_1 \geq 0$. *Hence, we can constrain the coefficient* $d_{2\,1} \in \mathbb{N}_0$ *of* W_1*, while* $d_{2\,2} \in \mathbb{Z}$. *Figures adapted from [21].*

completely divides out this symmetry. For this task, a new technique based on the concept of the *fundamental Weyl chamber* is developed. As discussed in section 4.2.3 this special chamber is uniquely related to a choice of simple roots. Once this choice is fixed there exists an algorithm (see [146]) based on the successive application of the *fundamental Weyl reflections* such that any vector in the root space is mapped to the fundamental Weyl chamber. Hence, this space is a reference parameter space for the whole root space. Remember that the defining property of the fundamental Weyl chamber is that all Dynkin labels are non-negative, i.e. $M_k \cdot \alpha_I \geq 0 \quad \forall \quad I$. Although the action of the Weyl group in combination with Wilson lines is more involved than for a single vector in the root space, the new search strategy is named after this consideration, the *fundamental Weyl chamber* search.

To motivate the idea behind the algorithm, let us start with the inverse problem. Starting from a gauge embedding matrix M, one can apply the algorithm of [146][1] such that the shift

---

[1] The algorithm computes the Dynkin labels of the root vector. It applies the fundamental Weyl reflection that corresponds to the first negative Dynkin label. The action switches the sign of this Dynkin label and has side effects on the non-orthogonal simple roots of the Dynkin diagram. This procedure is repeated for the transformed

vector V_1 is rotated to the fundamental Weyl chamber. In order to leave the resulting physics of the orbifold model unchanged, the fundamental Weyl reflections have to act simultaneously on all vectors M_k of the gauge embedding matrix. Hereafter, the shift vector V_1 may be invariant under certain fundamental Weyl reflections. Therefore, some Weyl symmetry of the complete model M is left. These unbroken Weyl reflections are those that leave V_1 invariant, i.e. $w_I(V_1) = V_1$ if and only if $V_1 \cdot \alpha_I = 0$, since

$$w_I(V_1) \;=\; V_1 - \underbrace{(V_1 \cdot \alpha_I)}_{=d_{1\,I}/N_1} \alpha_I \;, \tag{6.1}$$

i.e. $d_{1\,I} = 0$. These remnants of the Weyl group generate the symmetry group for the next 16-vector. In the case of $\mathbb{Z}_6$-II (1,1) the Wilson line W_3. Corresponding to the broken Weyl symmetry, W_3 can only be mapped to a *broken Weyl chamber* instead of the fundamental Weyl chamber. The reduced symmetry relates fewer parameters of the root space. Thus the broken Weyl chamber is defined regarding those fundamental Weyl reflections that leave V_1 invariant. As a result of mapping Wilson line W_3 to the broken Weyl chamber, the Dynkin labels $W_3 \cdot \alpha_I$ that correspond to the Weyl reflections w_I that leave the shift vector V_1 invariant are non-negative. This logic is reapplied to the next vectors until no Weyl reflection is left, which leaves all previous 16-vectors invariant.

The general procedure above leads us towards a method to directly choose gauge embedding matrices M that have the first vector in the fundamental Weyl chamber, and the subsequent vectors in the respective broken fundamental Weyl chamber, as illustrated in fig. 6.1. For this the dual basis expansion eq. (2.28) in terms of α_I^* is needed. In this basis the constraints on the Dynkin labels can directly be imposed to the coefficients $d_{k\,I}$ of the gauge embedding matrix M, as $M_k \cdot \alpha_I = \frac{1}{N_k} d_{k\,I}$. For the first vector, i.e. the shift vector V_1, we have the complete Weyl group and can draw this vector directly from the fundamental Weyl chamber

$$V_1 \cdot \alpha_I \;=\; \frac{d_{1\,I}}{N_1} \;\geq\; 0 \quad\Leftrightarrow\quad d_{1\,I} \;\in\; \mathbb{N}_0 \;, \tag{6.2}$$

for $I = 1, \ldots, 16$. Then, it is necessary to compute the unbroken Weyl symmetry that restricts the parameter space of the second vector. This unbroken Weyl symmetry is generated by those fundamental Weyl reflections that leave V_1 invariant. As V_1 lies in the fundamental Weyl chamber, it can only be invariant under a Weyl reflection if V_1 lies on the boundary of the fundamental Weyl chamber. The hyperplanes of the simple roots define the fundamental Weyl reflections w_I and yield this boundary [146]. Consequently, only those fundamental Weyl reflections w_I, which leave all previously chosen vectors M_k invariant, can still restrict the search space. Therefore, at step n in fig. 4.1 the coefficients $d_{n\,I}$ can be constrained for the vector M_n in eq. (2.28) as

$$d_{n\,I} \in \mathbb{N}_0 \quad \text{if} \quad d_{k\,I} \;=\; 0 \quad \text{for all } k = 1, \ldots, n-1 \;, \tag{6.3}$$

$$d_{n\,I} \in \mathbb{Z} \quad \text{if} \quad d_{k\,I} \;\neq\; 0 \quad \text{for any } k = 1, \ldots, n-1 \;. \tag{6.4}$$

Translating these constraints to the distribution, eq. (4.1), for sampling the coefficients leads to the distribution for V_1,

$$d_{1I} \sim \mathcal{U}\left\{0,\ \beta N_1\right\} \qquad \text{i.i.d.} \quad \forall\ I \tag{6.5}$$

---

root vector until all Dynkin labels are non-negative.

and to the case dependent sampling for any subsequently drawn M_k, where the d_{kI}-coefficients are drawn from,

$$d_{kI} \sim \mathcal{U}\{0,\ \beta N_k\} \qquad \text{in case (6.3)} \tag{6.6}$$

$$d_{kI} \sim \mathcal{U}\{-\beta N_k,\ \beta N_k\} \qquad \text{in case (6.4)}\ . \tag{6.7}$$

For the range coefficient β, the commonly used value for upcoming searches is $\beta = 4$, in order to consider an extensive area that does allow for a wide variety of physical models.

level	frequency of occurrence	first appearance of models with special properties
1	8 008	gauge group $\mathrm{U}(1)^{16}$ (with 218 matter fields)
19	1 915	non-Abelian gauge group ($\mathrm{SU}(2) \times \mathrm{U}(1)^{15}$)
635	114	$\mathrm{SU}(3) \times \mathrm{SU}(2)$
739	10	$\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_Y$ with 1 generation
747	2	$\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_Y$ with 2 generations
748	1	MSSM-like model

Table 6.1: *Special models from the* $\mathbb{Z}_6$-II (1,1) *orbifold geometry, ordered by their frequency of occurrence in a set of* 10^7 *random models. Note that the first column displays the label of the frequency of occurrence levels. Each level is degenerate by different physical models with the same frequency of occurrence. We list the first appearances of models with special properties that are given in the last column.*

6.1.2 Phenomenological constraints

Beyond the restriction of the landscape due to symmetries, there exists also gauge embedding matrices M that do not fulfill basic phenomenological properties. Of particular interest in this section are orbifold models with 4D gauge symmetry $G_{4D}(M)$ that do not contain the SM gauge group. Hence, these models are unable to match the SM with their effective QFT. The reason why this situation becomes important is that the main part of created heterotic orbifold landscape models is of the type described above. A search of 10^7 random models in the $\mathbb{Z}_6$-II (1,1) orbifold geometry, verified this. The models correspond to approximately $3.5 \cdot 10^6$ inequivalent massless spectra. The frequency of occurrence for the inequivalent spectra is visualized in fig. 6.2 and some special properties of the models are given in table 6.1. Astonishingly the phenomenologically excluded models turn out to have the highest repetition values, and the most interesting models are the rarest. To avoid these invalid models in our search for MSSM-like orbifold models seems to be necessary. Therefore, the upcoming sections show how the constraints can be taken into account during the search and focus on those areas of the heterotic orbifold landscape that at least satisfy the necessary condition of the SM gauge group.

The pseudo-GUT constraint $G_n(M) \geq G'_{\mathrm{SM}}$

Phenomenologically uninteresting models that have a gauge group smaller than the non-Abelian SM gauge group factors $G'_{\mathrm{SM}} = \mathrm{SU}(3) \times \mathrm{SU}(2)$ can be circumvented by checking the gauge group $G_n(M)$ at each step $n = 1, \ldots, 8$. This extends the algorithm fig. 4.1 naturally. Together with further constraints developed in the upcoming sections, this algorithm is illustrated in

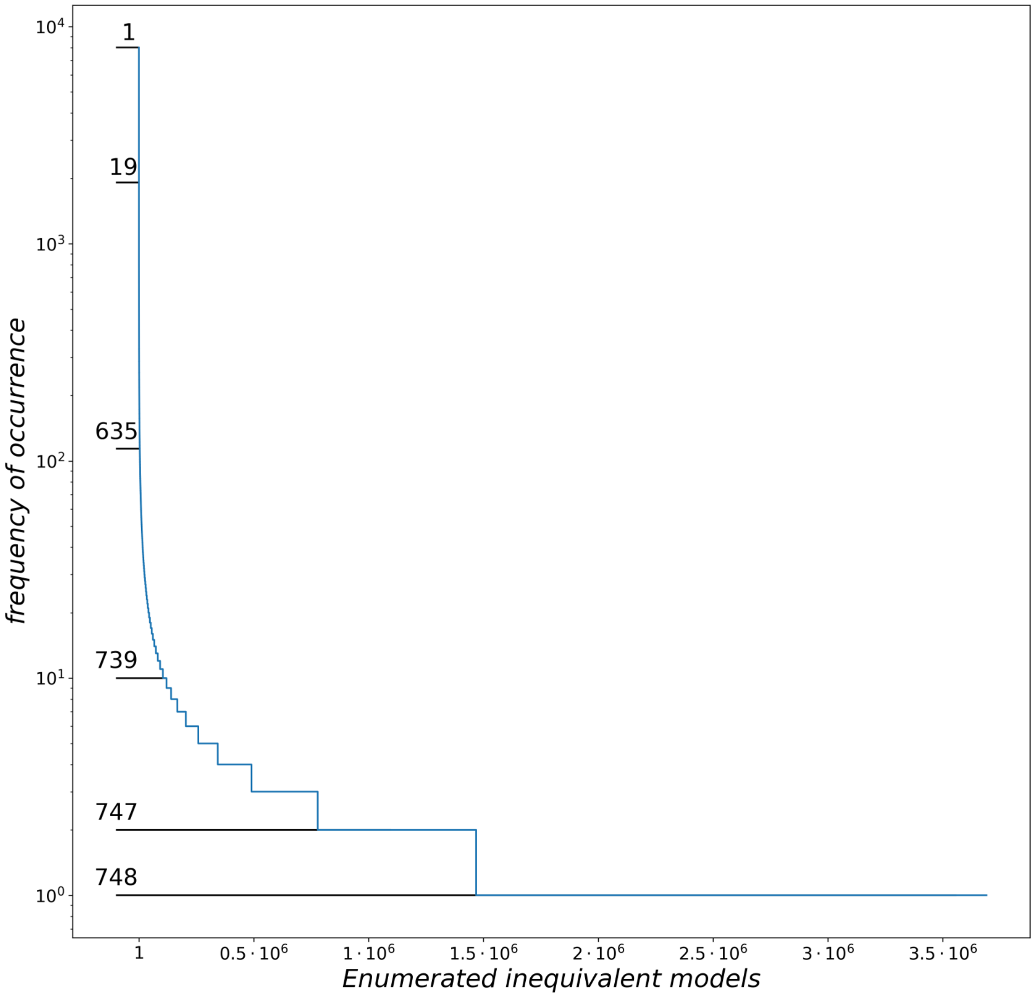


Figure 6.2: *Frequency of occurrence of inequivalent* $\mathbb{Z}_6$-II (1,1) *orbifold models. The inequivalent models are enumerated from* 1 *to* 3 690 513*, on the horizontal axis. On the vertical axis we see the corresponding frequency of occurrence, i.e. model # 1 has a frequency of* 8 008*. For more details on some of these models see table 6.1. Moreover, the different frequencies of occurrence levels from 1 to 748 are indicated. At level 748 the first MSSM-like model appears with a frequency of occurrence of 1. Figure adapted from [21].*

fig. 6.3. The reason why this extension is easy to implement, is that additional vectors, e.g. M_{n+1}, break the gauge group further or leave it equal, i.e. $G_{n+1}(M) \subseteq G_n(M)$. Hence, the SM gauge group provides a lower bound on the breaking pattern at each step n. A quick option to control the size of the remaining gauge group $G_n(M)$ is to compute the number of unbroken roots N_{ur}. In eq. (5.3) we computed N_{ur} for the full gauge embedding matrix M. As we here want to check this quantity at each step of constructing M we define,

$$N_{\text{ur}}^{(\alpha)}(M_n) \;=\; \sum_{p\,\in\,\Phi_{\mathrm{E}_8}} \left(\prod_{k\,=\,1}^{n} \delta\left(p \cdot M_k^{(\alpha)}\right)\right) \;, \quad \delta(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Z} \\ 0 & \text{else} \end{cases} \;. \tag{6.8}$$

Note, that the product runs from $k = 1, \ldots, n$. We introduced the shorthand notation that n labels the number of vectors that have been chosen to specify a consistent gauge embedding matrix M. For of the SM, the SU(3) holds six unbroken roots while we get plus two unbroken roots from the SU(2) factor, i.e. $N_{\mathrm{ur}}(G'_{\mathrm{SM}}) = 8$. This defines a lower bound at each step n in the production of a model M for the first E_8 factor, i.e.

$$N^{(1)}_{\mathrm{ur}}(M_n) \;\geq\; 8 \quad \text{at each step } n \;, \tag{6.9}$$

which is called the *pseudo-GUT* constraint. The second E_8 factor is unconstrained and free to produce any hidden gauge groups. Hence, we force the model to be able to host the MSSM spectrum in the first E_8. Different constraints on the E_8 factors have the additional benefit to divide out their $\mathbb{Z}_2$ symmetry from the search space, see section 4.2.

However, due to some special cases we have to extend this check. First of all SU(2) × SU(2) × SU(2) × SU(2) could fulfill the constraint eq. (6.9) but can not host the SM gauge group. Therefore, the gauge factors of $G_n(M)$ are forced to contain a root system Φ that allows for SU(3), i.e. there exists $|\Phi| \geq 6$. There are also the special cases SO(8) and SU(4), which can not be broken to G'_{SM}. Before we consider these situations we introduce another consistency constraint in the following subsection.

An important factor is the response of the algorithm if a newly chosen vector M_k results in a gauge group breaking below these lower bounds. Several strategies are reasonable: A straightforward strategy would be to replace the current vector M_k by the zero vector, which automatically satisfies the modular invariance conditions (eq. (2.46)) for Wilson lines and leaves the gauge group unbroken for this step. Another could be to take the coefficients d_{kI} of the insufficient vector M_k and set individual coefficients to $d_{kI} = 0 \mod N_k$ in order to ease the breaking of roots. However, these approaches deform the distribution massively and seem to be not conservative enough for an even exploration of the compactification parameter space. The strategy in this investigation is to neglect the vector M_k and draw the d_{kI} coefficients again from the distribution eq. (6.5) and eq. (6.6), until they fulfill the constraints.

The Standard Model gauge group constraint: $\mathbf{SU(3) \times SU(2) \subseteq \boldsymbol{G_{4D}(M)}}$

For a model to contain the non-Abelian gauge group factors G'_{SM}, the *pseudo-GUT* constraint is a necessary condition. However, the search focuses on MSSM-like models with $\mathrm{SU}(3)_{\mathrm{C}} \times \mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_Y$ gauge symmetry in 4D. Hence, after we have chosen the last vector M_k, we have to check that the model M has a 4D gauge symmetry

$$G_{4D}(M) \;=\; \mathrm{SU}(3) \times \mathrm{SU}(2) \times G_{\mathrm{hidden}} \;. \tag{6.10}$$

We denote this constraint by,

$$\mathrm{SU}(3) \times \mathrm{SU}(2) \subseteq G_{4D}(M) \;. \tag{6.11}$$

The geometrical conditions eq. (2.25) make this condition geometry specific. In particular, the initial step is to identify the last shift vector or Wilson line that can be chosen independently, i.e. which is not of order one $N_k = 1$ and not related to a former 16-vector. For the $\mathbb{Z}_6$-II (1,1) orbifold geometry, this results in the Wilson line W_6, see table 6.1. However, for other orbifold geometries, e.g. $\mathbb{Z}_3 \times \mathbb{Z}_6$ $(2,2)$ the geometry fixes all Wilson lines and the second shift vector V_2 has to enable the constraint eq. (6.11). The constraint is checked by calculating the unbroken roots from the first E_8 factor and the sizes of the orthogonal root systems. Concretely, this implies that in order to contain SU(3) × SU(2) at least two root systems $\Phi_{\mathrm{SU}(3)}, \Phi_{\mathrm{SU}(2)}$, one of size six $|\Phi_{\mathrm{SU}(3)}| = 6$ and another of size two $|\Phi_{\mathrm{SU}(2)}| = 2$, are included in the set of root systems of $G_{4D}(M)$. Note, extra gauge group factors are allowed.

Now we can come back to the SO(8) and SU(4) gauge groups, which can not be broken to G'_{SM}. They occur on rare occasions, and in order to keep the search fast, a separate check is only applied previously to the last freely selectable vector M_k. In the case of $\mathbb{Z}_6$-II (1,1) this is before the Wilson line W_6. If it appears that the gauge group $G_7^{(1)}(M)$ has only one root system and this root system is of size $N_{\text{ur}}^{(1)}\,(\text{SO}(8)) = 12$ or $N_{\text{ur}}^{(1)}\,(\text{SU}(4)) = 24$ the construction of the model is discarded and starts from scratch. This method avoids trying to find the impossible last vector M_k that breaks towards G'_{SM}.

The phenomenological constraints from eq. (6.9) and eq. (6.11) implemented into the search algorithm get applied to the test case of $\mathbb{Z}_6$-II (1,1) orbifold models. It turns out that the proportion of MSSM-like models increases from $\frac{1}{10\,000\,000} = 10^{-7}$ in the case without the phenomenological constraints to $\frac{3}{2\,665\,463} \approx 10^{-6}$ in the case with phenomenological constraints. The proportions are computed based on the equivalent models in order to avoid "floating correlations"(see [144]). A cross-check on the increase of MSSM-like models is the fact that $\approx 90\%$ of all models of the *fundamental Weyl chamber* search do not fulfill the essential gauge group constraint. Hence, this approach reduces the landscape parameter space to a tenth. Let us stress that such direct implications on the proportions are entirely trustworthy for the set of equivalent models. However, this makes no direct implications on the probability of finding more inequivalent MSSM-like spectra. In upcoming sections, this issue arises, and specific analysis is needed to find patterns for inequivalent models. In addition, the empirical observation that MSSM-like models are often related to a vanishing Wilson line [79] is the reason for a second search with $W_5 = (0^{16})$. The results are summarized in table 6.2 (the corresponding dataset is called *phenomenology*).

6.2 Contrast patterns for $\mathbb{Z}_6$-II (1,1) orbifolds

In the previous section, phenomenological constraints that can be checked easily during the search for MSSM-like orbifold models were discussed. Importantly, these conditions are necessary for a model to be MSSM-like while not sufficient. An extension of this procedure is introduced to include new constraints for MSSM-like models by exploiting methods from contrast data mining [193]. This specific type of data mining focuses on finding so-called contrast patterns that allow distinguishing between different types of classes. A statistical approach will determine these new constraints. Thus, demanding them can potentially miss MSSM-like models even though the probability for this to happen is tiny. In other words, the new constraints have not yet an analytical origin, but they have strong statistical evidence and remarkably enhance the probability for a constructed model to be MSSM-like. Utilizing the contrast patterns will further reduce the heterotic orbifold landscape, to the subspaces of MSSM-like models. The conventional search algorithm fig. 4.1 rarely reaches some of these subspaces since the distribution for sampling has to be strongly deformed. The additional constraints from the contrast patterns will allow accessing these subspaces with a much higher probability by deforming the probability mass function according to those subspaces.

A contrast pattern c can be defined as a pattern with significantly different supports for distinct datasets [193]. The support is defined as

$$\text{supp}(c, D) \;=\; \frac{|\{M \in D \mid M \text{ satisfies } c\}|}{|D|}\,, \tag{6.12}$$

where D is a set of data points, i.e. orbifold models, and c is a set of specified constraints that have to be fulfilled. Particularly, the two datasets are $D_{\text{MSSM-like}}$ and $D_{\cancel{\text{MSSM}}\text{-like}}$, which is the set of MSSM-like models and the complementary set, respectively. We seek for constraints c

	added constraint	dataset	condition	# models	# MSSM-like				
					equiv.	inequiv.			
traditional	eq. (6.3)	*fundamental Weyl chamber*		10 000 000	1	1			
	eq. (6.13)	*phenomenology*		2 665 463	3	3	130		
			$W_5=(0^{16})$	2 551 272	509	129			
contrast patterns	eq. (6.24)	*hidden E_8*		2 543 415	12	11	136	415	468
			$W_5=(0^{16})$	2 609 872	863	135			
	eq. (6.31)	*dynamic hidden E_8*		1 876 273	3 299	245	395		
			$W_5=(0^{16})$	1 231 608	8 455	321			
			$W_3=(0^{16})$	378 604	7	2			
	eq. (6.35)	*U-sector*		4 793 146	4 953	357	459		
			$W_5=(0^{16})$	3 046 262	17 406	358			

Table 6.2: *Table of all datasets constructed in the* $\mathbb{Z}_6$-II (1,1) *orbifold landscape. In terms of the applied constraints the datasets are sorted in ascending order, i.e. a new dataset incorporates all of the previous constraints. Note that in the dynamic search* $N_{\text{ur}}^{(2)} \geq X$ *from section 6.2.3 the case* $X = 6$ *was disregarded since it was already sampled in the* hidden E_8 *dataset. We also made use of the additional conditions* $W_5 = (0^{16})$ *or* $W_3 = W_4 = (0^{16})$*, where* $W_5 = (0^{16})$ *is known to be beneficial for finding MSSM-like models [79].*

that are satisfied for (almost) all MSSM-like models while a considerable fraction of ~~MSSM~~-like models violate them.

In the ideal case, one can identify contrast patterns c with $\text{supp}(c, D_{\text{MSSM-like}}) = 1$ and $\text{supp}(c, D_{\cancel{\text{MSSM}}\text{-like}}) = 0$. As an example, one can reformulate the phenomenological constraints from section 6.1.2 as perfect contrast patterns,

$$c_{G_n(M) \geq G'_{\text{SM}}} \;=\; \left\{ N_{\text{ur}}^{(1)}(M) \;\geq\; 8,\; \exists\, |\Phi| \geq 6 \right\} , \tag{6.13}$$

$$c_{G'_{\text{SM}} \subseteq G_{4D}(M)} \;=\; \left\{ \exists\, \Phi_1, \Phi_2 \;\subseteq \Phi_{G_{4D}(M)} \;\middle|\; |\Phi_1| = 6 \text{ and } |\Phi_2| = 2 \right\} , \tag{6.14}$$

$$c_{\text{phenomenology}} \;=\; \left\{ c_{G_n(M) \geq G'_{\text{SM}}},\; c_{G'_{\text{SM}} \subseteq G_{4D}(M)} \right\} , \tag{6.15}$$

i.e. they have no uncertainty since they are derived from analytical conditions rather than statistics. Furthermore, from the previous section we can infer that all MSSM-like models but only a tenth of the ~~MSSM~~-like models can survive the constraint, i.e.

$$\text{supp}(c_{\text{phenomenology}}, D_{\text{MSSM-like}}) \;=\; 1 \;, \tag{6.16}$$

$$\text{supp}(c_{\text{phenomenology}}, D_{\cancel{\text{MSSM}}\text{-like}}) \;\approx\; \frac{1}{10} \;. \tag{6.17}$$

However, in the case of contrast patterns that are derived from statistics one should be careful if the found patterns generalize to all (or at least other) MSSM-like physical spectra, or if the contrast pattern restricts the whole landscape only to the type of MSSM-like models that are

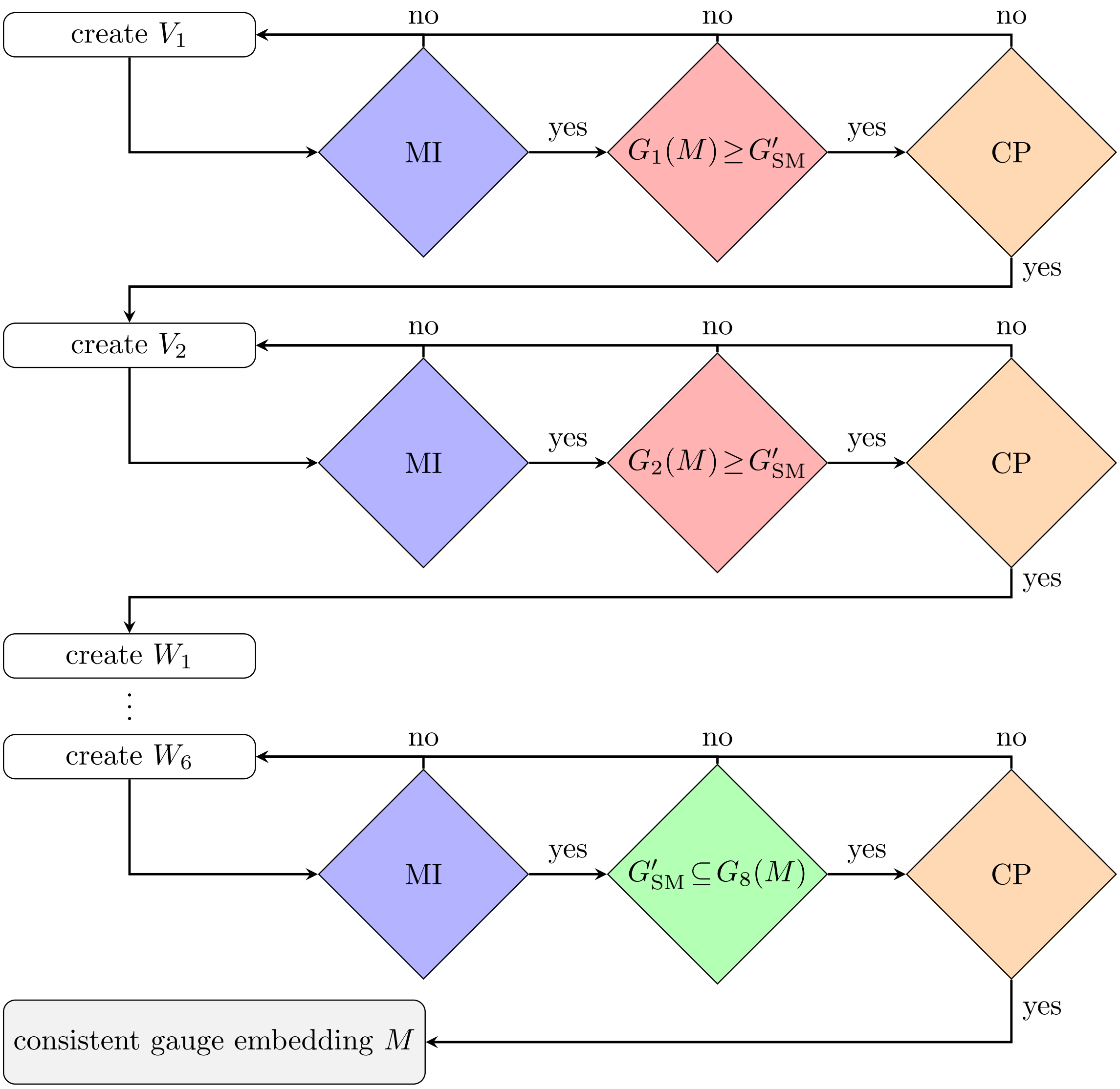


Figure 6.3: *Successive creation of a gauge embedding matrix M. The flowchart illustrates the new construction algorithm that is an extension of the search strategy fig. 4.1. Additionally, to the modular invariance (MI) conditions, two extra types of conditions are imposed at each step n of the construction:*
(i) As discussed in section 6.1.2 the gauge group $G_n(M)$ in terms of the root system eq. (6.8) is computed using the already chosen vectors M_k for $k = 1, \ldots, n$. If $G_n(M)$ satisfies the pseudo-GUT *condition eq. (6.9) to allow for the non-Abelian gauge group factors $G'_{\mathrm{SM}} = \mathrm{SU}(3) \times \mathrm{SU}(2)$ of the SM, the construction process continues with the next vector M_{n+1}.*
(ii) The new contrast mining constraints from section 6.2, the contrast patterns (CP) are imposed.
Finally, for the last Wilson line that is not constrained by geometrical conditions, e.g. table 2.2, the 4D gauge group $G_{4D}(M)$ must contain G'_{SM}, see section 6.1.2. Here illustrated for the last Wilson line $M_8 = W_6$. Figure adapted from [21].

included in the statistical analysis. For example, imagine the extreme case where $D_{\text{MSSM-like}}$ contains only one model and the contrast mining algorithm is allowed to use any feature, e.g.

the individual coefficients d_{kI} of the gauge embedding matrix M. Indeed, this can lead to an overfitted pattern $c_{d_{kI}}$ that enforces exactly the set of coefficients d_{kI} of the one MSSM-like model to the search algorithm. For the support we obtain $\mathrm{supp}(c_{d_{kI}}, D_{\text{MSSM-like}}) = 1$ and $\mathrm{supp}(c_{d_{kI}}, D_{\cancel{\text{MSSM}}\text{-like}}) = 0$. However, this contrast pattern would certainly show poor generalization performance, applied to the heterotic orbifold landscape. Hence, too promising constraints should raise suspicion and demand an analytical explanation. Due to this, a conservative approach should be taken when mining for contrast patterns.

The generic goal of contrast mining can be formalized by defining the growth rate

$$\mathrm{gr}(c, D_{\text{MSSM-like}}, D_{\cancel{\text{MSSM}}\text{-like}}) \;=\; \frac{\mathrm{supp}(c, D_{\text{MSSM-like}})}{\mathrm{supp}(c, D_{\cancel{\text{MSSM}}\text{-like}})}\,, \tag{6.18}$$

which has to be maximized. In the following, $\mathrm{gr}(c)$ is used for convenience if the datasets $D_{\text{MSSM-like}}$ and $D_{\cancel{\text{MSSM}}\text{-like}}$ are clear from the context. To get some intuition for the growth rate, we rewrite it in terms of the probability $\hat{p}$. Here, the hat indicates that the probability is estimated by the sample proportion $\hat{p}(Y) = \frac{N_Y}{N}$, where the label $Y \in \{\text{MSSM-like}, \cancel{\text{MSSM}}\text{-like}\}$ and the total sample size is given by $N = N_{\text{MSSM-like}} + N_{\cancel{\text{MSSM}}\text{-like}}$. For a reminder on the connection of sample proportions and probabilities see appendix B.1. In detail $\hat{p}^c(Y) = \frac{N_Y^c}{N^c}$ with $N_Y^c = |\{M \in D_Y \mid M \text{ satisfies } c\}|$ is the probability of a model being $Y = $ MSSM-like or $Y = \cancel{\text{MSSM}}$-like given the constraints c. The corresponding probability without imposing the constraints c is denoted by $\hat{p}(Y)$. It follows for $\hat{p}^c(\text{MSSM-like})$ as a function of $\hat{p}(\text{MSSM-like})$,

$$\hat{p}^c(\text{MSSM-like}) \;=\; \frac{\mathrm{gr(c)}\,\hat{p}(\text{MSSM-like})}{1 + (\mathrm{gr(c)} - 1)\hat{p}(\text{MSSM-like})}\,. \tag{6.19}$$

As the probability for MSSM-like models is typically very small, one can observe:

$$\hat{p}(\text{MSSM-like}) \;\ll\; 1 \;:\; \hat{p}^c(\text{MSSM-like}) \;=\; \mathrm{gr(c)}\hat{p}(\text{MSSM-like}) + \mathcal{O}\big(\hat{p}(\text{MSSM-like})^2\big)\;, \tag{6.20}$$

where the Taylor expansion in eq. (6.20) converges if $\hat{p}(\text{MSSM-like}) < \frac{1}{|\mathrm{gr}(c)-1|}$. Thus, for $\mathrm{gr}(c) < 1$ the proportion of MSSM-like models in the constrained space is expected to decrease. For $\mathrm{gr}(c) = 1$ the effects on both classes cancel each other and for $\mathrm{gr}(c) > 1$ one expects a higher probability to find MSSM-like models in the subspace defined by the contrast patterns c. However, before the search for contrast patterns c can start, one has to define some (physical) quantities that possibly can lead to such patterns. This is known as feature engineering, see appendix C.2 for a conceptual introduction.

For further convenience, only the impactful features are stated. Furthermore, it is essential to realize that the features have to be controllable in the sense that they are testable during the successive construction. Thus, physical quantities related to the bulk (see section 2.4.1) turn out to be useful. Content from the twisted sector is strongly related to the exact combination of shift vectors and Wilson lines by depending on the local shift vectors V_h, see section 2.4.2. Hence, these quantities change non-linearly the projection conditions on the physical states, which makes them challenging to implement even though patterns in the final orbifold model might exist. From section 6.1.2 it is known that the 4D gauge group has a significant impact and it can be checked easily at every step in the production of a model, by estimation of the number of unbroken roots $N_{\text{ur}}^{(\alpha)}(M)$ see eq. (6.8). Hence, we expect that further features can be obtained from the 4D gauge group. Additionally, the numbers of orbifold-invariant bulk matter fields, eq. (2.51), will be used as features. They are calculated similar to the number of unbroken roots of the gauge group in eq. (6.8), with the difference of the action of the geometrical twist vector v (see section 2.4.1). Thus, at each step n of our search algorithm displayed in fig. 6.3, we compute

$$N^{(\alpha)}_{\mathrm{U}_a}(M_n) \;=\; \sum_{p\,\in\,\Phi_{\mathrm{E}_8}} \prod_{k\,=\,1}^{n} \delta\left(p \cdot M^{(\alpha)}_k - \Theta(2-k)\, q_{(a)} \cdot v_{(k)}\right) \,, \tag{6.21}$$

for $\alpha = 1, 2$ and $a = 1, 2, 3$. Note that the term $q_{(a)} \cdot v_{(k)}$ in eq. (6.21) vanishes for the Wilson lines $M^{(\alpha)}_k$, $k = 3, \ldots, 8$, due to

$$\Theta(x) \;=\; \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x \geq 0 \end{cases} \,. \tag{6.22}$$

Furthermore, the twist vectors in eq. (6.21) are geometry dependent and are given for $\mathbb{Z}_6$-II (1,1) orbifold geometry by $v_{(1)} = (0, \frac{1}{6}, \frac{1}{3}, -\frac{1}{2})$ and $v_{(2)} = (0^4)$. Finally, the setup is ready for the mining of contrast patterns from the features listed in table 6.3. A decision tree (appendix C.6) is applied to the defined feature space of our datasets, in order to identify threshold conditions on the features that correlate with the property of a model being MSSM-like. If such a correlation exists, the corresponding threshold condition can be used as a contrast pattern in our search algorithm for MSSM-like orbifold models.

	first E_8				hidden E_8			
feature	$N^{(1)}_{\mathrm{ur}}$	$N^{(1)}_{\mathrm{U}_1}$	$N^{(1)}_{\mathrm{U}_2}$	$N^{(1)}_{\mathrm{U}_3}$	$N^{(2)}_{\mathrm{ur}}$	$N^{(2)}_{\mathrm{U}_1}$	$N^{(2)}_{\mathrm{U}_2}$	$N^{(2)}_{\mathrm{U}_3}$

Table 6.3: *Table of the features for contrast mining.*

However, a decision tree is not able to model any function precisely, especially with the small feature space we provide here. Thus, we can distinguish between two types of errors, i.e. cases where $Y_{\mathrm{correct}}(M) \neq Y_{\mathrm{predicted}}(M)$. They are called:

- *false positives*: $Y_{\mathrm{correct}}(M) = \cancel{\text{MSSM}}$-like but $Y_{\mathrm{predicted}}(M) =$ MSSM-like
- *false negatives*: $Y_{\mathrm{correct}}(M) =$ MSSM-like but $Y_{\mathrm{predicted}}(M) = \cancel{\text{MSSM}}$-like

Every classification task (appendix C.5) attempts to minimize the number of false predictions. Nonetheless, at a certain point it always comes to a trade-off between *false positives* and *false negatives*. Thus, the decision is unavoidable whether one wants to suppress one of them for the inconvenience of raising the other one. In our case, a *false positive* classification by the decision tree is less problematic, since afterwards we can check each of these orbifold models explicitly, using the `orbifolder`. A single verification takes only a very short time. Certainly, the purpose of this investigation aims towards excluding vast regions, i.e. huge combinatory possibilities. As long as the reliably excluded region is substantial, the explicit checks do not lower the effectiveness of this procedure.

However, in the case of a *false negative* classification, the consequence would be that we neglect areas in which MSSM-like orbifold models exist. Because these MSSM-like orbifold models are the primary goal, we want to minimize the number of *false negatives* by all means, while we want to keep the number of *false positives* as low as possible. Therefore, a loss matrix l teaches the machine learning algorithm about the different rank of specific models [194]. This reads,

$$l = \left(\begin{array}{cc} 0 & 10^6 \\ 1 & 0 \end{array} \right) \,, \tag{6.23}$$

where the values $Y_{\text{predicted}}(M)$ being either MSSM-like or not correspond to the two columns and the two rows correspond to the correct value $Y_{\text{correct}}(M)$ being either MSSM-like or not. Then, l_{12} corresponds to the *false negative* cases of MSSM-like orbifold models M that have been classified by the decision tree as $Y_{\text{predicted}}(M) =$ ~~MSSM~~-like. As this should be avoided, the arrangement is penalized with a high loss value of $l_{12} = 10^6$. The other potential error of a *false positive* classification is not as problematic. Consequently, we set $l_{21} = 1$. This will supervise the decision tree algorithm towards suppressing the *false negative* cases such that (almost) all MSSM-like orbifold models are correctly classified.

The concrete realization of this loss matrix is given by a python dictionary handed to the `class_weight` argument of the `DecisionTreeClassifier` from the scikit-learn package [189], i.e. `class_weight`={ MSSM-like : 10^6, ~~MSSM~~-like : 1 }. Note, that the value of l_{12} is not fixed for all upcoming investigations but can be adjusted as a hyperparameter in order to balance the trade-off between the *false negatives*, as well as, to give the decision tree enough flexibility to cut off larger areas of the landscape. A concrete example of this will appear in section 6.2.5.

To quantify the quality of the predictions we use the so-called `Recall`(MSSM-like) score. It is defined in detail in appendix C.5 and is given as the ratio between the number of correct predictions of the MSSM-like class and the total number of MSSM-like orbifold models. Thus, if the number of *false negatives* for MSSM-like orbifold models is zero, we find `Recall`(MSSM-like) = 1.00. This is analyzed on the validation set $D^{\texttt{val}}$ and all MSSM-like orbifold models are expected to be assigned with the correct value $Y =$ MSSM-like. In upcoming sections, we will transform our data to the defined feature space and apply decision trees to this representation in order to obtain contrast patterns.

6.2.1 The *hidden* E_8 contrast pattern

For the machine learning setup the dataset *phenomenology* from table 6.2 is split randomly in $D^{\texttt{train}}$ and $D^{\texttt{val}}$ with a validation size of 33%. However, a necessary modification is added to the dataset to avoid *data leakage*. *Data leakage* refers to a corrupted data mining process, through a confusion in the dataset arrangement (appendix C). Such a mistake can emerge if there exists an MSSM-like spectrum that completely dominates $D_{\text{MSSM-like}}$ with all of its equivalent copies. Then, this model would appear most likely in both, $D^{\texttt{train}}$ and $D^{\texttt{val}}$. Thus, the algorithm can already use the characteristics of this model during training. Moreover, the same model would dominate the classification result of this algorithm on the $D^{\texttt{val}}$ and pretend that the learned predictions generalize to generic MSSM-like models. (Recall that the performance of any machine learning approach is estimated on the validation set $D^{\texttt{val}}$.) However, it is sufficient that this particular MSSM-like model is predicted correctly by the method. Therefore, the dataset has to be adjusted to be in accordance with the true target of this analysis. Especially, the patterns we are aiming for should generalize to inequivalent MSSM-like models. Hence, only inequivalent MSSM-like models from the dataset *phenomenology* are used and split to $D^{\texttt{train}}$ and $D^{\texttt{val}}$. Then the performance on $D^{\texttt{val}}$ will evaluate how the approach generalizes towards MSSM-like models with distinct matter spectra from the MSSM-like models used for training.

Contrary, for the ~~MSSM~~-like models, it is required to keep the equivalent models. The frequency of occurrence gives a notion on the size that these models occupy in the landscape, i.e. in the space of compactification parameters $\mathbb{Z}^{128}$. A split of the decision tree is thus especially valuable for reduction if it excludes a considerable region in $\mathbb{Z}^{128}$. As each equivalent ~~MSSM~~-like model is a particular point in $\mathbb{Z}^{128}$, the reduction in the search space correlates with the exclusion of equivalent ~~MSSM~~-like. Hence, in terms of the space of compactification parameters, our goal is to exclude as many points in $\mathbb{Z}^{128}$ as possible, regardless if they are related to many different inequivalent physical models or not.

As the defined features directly depend on spectrum(M) and not on the coefficients d_{kI} of $\mathbb{Z}^{128}$ the decision tree performs its splits based on these features. Here an additional advantage of this approach emerges. A specific split in the decision tree will exclude all points $\mathbb{Z}^{128}$ that give rise to the same forbidden feature values. Due to this, a small restriction in feature space can give rise to a massive effect in the space $\mathbb{Z}^{128}$ of gauge embedding matrices M. In cases of many equivalent models, the restriction of the feature space can be kept minimal. The advantage is that enough room for the discovery of novel MSSM-like models is left available in the feature space. In opposition, if only inequivalent ~~MSSM~~-like models are used, the decision tree would not consider excluding a single physical model, e.g. $\mathrm{U}(1)^{16}$, even though it might correspond to a vast area in $\mathbb{Z}^{128}$, i.e. a lot of equivalent copies are produced during the search of the landscape, remember fig. 6.2.

The decision tree is trained on $D^{\texttt{train}}$, such that the hyperparameters are tuned to realize `Recall`(MSSM-like) = 1.00, eq. (C.9), on $D^{\texttt{val}}$. This corresponds to the case of no *false negatives*, hence all MSSM-like models are assigned to the correct class. The reason for this strong restriction is that the focus lies on contrast patterns that are satisfied by all MSSM-like models, and the performance on $D^{\texttt{val}}$ gives this estimate. For example lowering the loss value l_{12} to 10^5 already leads to undesirable *false negatives*.

During training, the decision tree identifies areas in the feature space and assigns the two classes MSSM-like or ~~MSSM~~-like to them. However, the goal is that the decision tree assigns the class ~~MSSM~~-like only to those areas that are highly populated with ~~MSSM~~-like models. Thus, it can be ensured that the probability of an MSSM-like model is extremely small in these areas of ~~MSSM~~-like models. Therefore, only the very first cuts of the tree where a lot of data is available are considered. From fig. 6.4 we can infer the pattern for a lower bound on the number of unbroken roots from the hidden E_8 factor,

$$c_{\text{hidden E}_8} \ = \ \left\{ N_{\text{ur}}^{(2)}(M) \ \geq \ 6 \right\} , \tag{6.24}$$

for a model M to be MSSM-like. According to its origin, this constraint is called *hidden E_8* contrast pattern. In principle, the decision tree keeps splitting the feature space to separate the data further. However, these separations are based on fewer data and hence have a larger margin of error. In order to only consider splits with very high confidence, we stop the tree after the first and most crucial division. Later, we collect new data and revisit the subsequent splits of the tree. Before we implement the new contrast pattern *hidden E_8*, we can estimate the growth rate on the training set $D^{\texttt{train}}$ from our *phenomenology* dataset,

$$\mathrm{gr}\left(c_{\text{hidden E}_8}, D_{\text{MSSM-like}}, D_{\cancel{\text{MSSM}}\text{-like}}\right) \ = \ \frac{1}{\frac{896\,169+551\,319}{3\,494\,867}} \ \approx \ 2.4 \ , \tag{6.25}$$

using eq. (6.18) with $\mathrm{supp}(c_{\text{hidden E}_8}, D_{\text{MSSM-like}}) = 1$ for our *phenomenology* dataset and the numbers for $\mathrm{supp}(c_{\text{hidden E}_8}, D_{\cancel{\text{MSSM}}\text{-like}})$ can be read off from figure fig. 6.4. Thus, the contrast pattern eq. (6.24) provides a method to exclude a vast area of the $\mathbb{Z}_6$-II (1,1) orbifold landscape. Statistically, it is highly unlikely for this area to lead to MSSM-like models. The computational resources can be focused on searching in areas where the proportion of a model to be MSSM-like is nonzero.

The *hidden E_8* contrast pattern can easily be implemented into the search algorithm displayed in fig. 6.3. The collected dataset is called *hidden E_8* and summarized in table 6.2. One recognizes

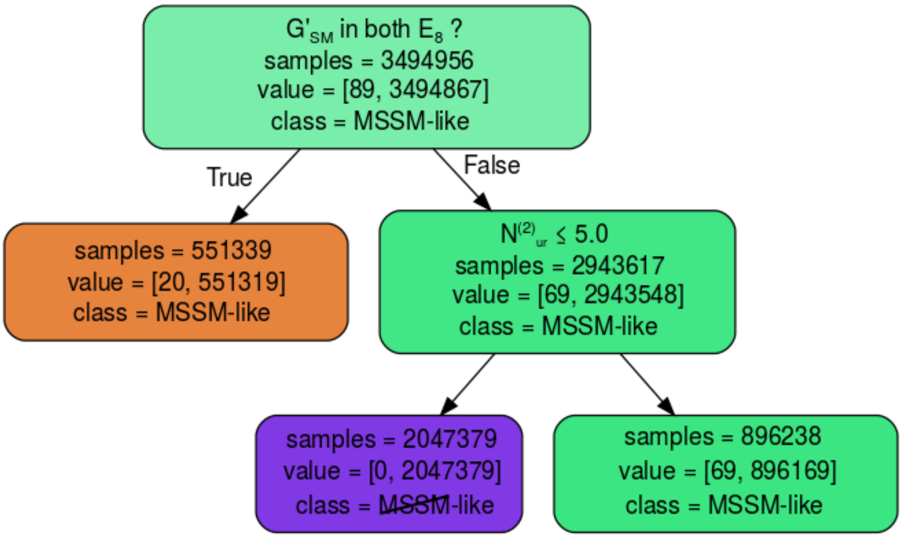


Figure 6.4: *Visualization of the decision tree trained on the* phenomenology *dataset from table 6.2. The numbers are evaluated on the training set D^{train}. We can extract the contrast pattern $N_{\text{ur}}^{(2)} \leq 5$ for ~~MSSM~~-like models from the first node in the right branch. For MSSM-like models the threshold condition can be inverted to $N_{\text{ur}}^{(2)} \geq 6$. The informative labels shown at each node state: (i) the threshold condition, e.g. G'_{SM}* in both E_8? *(ii) total number of models in this node, e.g.* samples=3 494 956 *(iii) the number of MSSM-like and ~~MSSM~~-like models in this node, e.g.* value = [89, 3 494 867] *(iv) the class assigned to all models in the node, e.g.* class = MSSM-like. *The final prediction for the models is given by the leaf nodes, i.e. the last nodes without further splitting. Figure adapted from [21].*

the increase for the probability to detect an MSSM-like model from

$$\hat{p}(\text{MSSM-like in } D_{\text{phenomenology}}) = \frac{512}{5\,216\,735} \approx 10^{-4} \quad \text{to} \tag{6.26}$$

$$\hat{p}(\text{MSSM-like in } D_{\text{hidden E}_8}) = \frac{875}{5\,153\,287} \approx 2 \cdot 10^{-4}\,, \tag{6.27}$$

which is consistent with the estimated growth rate in eq. (6.25). However, these are the expectations of obtaining any MSSM-like model. Thus, they are not necessarily inequivalent to the previously identified MSSM-like models. Unfortunately, the collected total number of inequivalent MSSM-like models was not adequate. The increase from 130 inequivalent MSSM-like models in the *phenomenology* dataset to 136 in the *hidden* E_8 dataset is rather small. Before we will investigate the reasons for this and present a solution that will lead to many new inequivalent MSSM-like models, the next section is dedicated to estimating the probability of missing MSSM-like models due to this constraint.

6.2.2 Confidence intervals on contrast mining patterns

To neglect areas of the heterotic orbifold landscape on statistical reasoning gives rise to the question of how reliable these constraints are. The possibility to miss MSSM-like models that are hidden in between the overwhelming amount of phenomenologically excluded models is still present, as no analytical solution exists. This analytic solution might be discovered by generating a lot of different MSSM-like models and recognizing the underlying pattern. However, until this can be done, it is possible to quantify the uncertainty by assigning confidence intervals to the resulting contrast mining patterns. Therefore, the probability of events that did not occur at all can be modeled by a binomial distribution. (Details on this and a derivation of the formula can be found in appendix B.6). For a confidence interval of 95% it turns out that the probability of an event A that did not occur in the sample is,

$$\Pr(A) \leq 1 - \sqrt[n]{(0.05)} \tag{6.28}$$

where n is the number of samples. Thus it is possible to assign a maximal probability to events related to finding MSSM-like models in the excluded area. From the dataset *phenomenology* with $W_5 = (0^{16})$ of table 6.2 one can infer that

$$\Pr\Big(\exists\; M : M \in D_{\text{MSSM-like}} \text{ but } N_{\text{ur}}^{(2)}(M) \ngeq 6\Big) \leq 1 - \sqrt[509]{0.05} \;=\; 0.0059\;. \tag{6.29}$$

This is the probability that a new point in $\mathbb{Z}^{128}$, which is MSSM-like, violates the contrast mining condition, eq. (6.24). Note that this probability only applies to MSSM-like models, which are very rare and difficult to find in the first place. Therefore, one can also set a probabilistic limit to the hypothetical case of actually finding an MSSM-like model in the excluded region. Concretely, to be able to filter such a model against the background of ~~MSSM~~-like models. This refers to compute the probability to find the $\mathbb{Z}^{128}$ configuration for the MSSM-like model in this region,

$$\Pr\;(\text{MSSM-like models in excluded area}) \;\leq\; 1 - \sqrt[N_{\text{exc}}]{0.05} \;\approx\; 3\cdot 10^{-6}\;, \tag{6.30}$$

where $N_{\text{exc}} = \left|\left\{M \in D^{W_5=(0^{16})}_{\textit{phenomenology}} \,\middle|\, N_{\text{ur}}^{(2)}(M) < 6\right\}\right| \approx 10^6$ is the number of models that reside in the excluded region. Let us explain why the estimates are based on the subset with $W_5 = (0^{16})$ of the whole dataset *phenomenology*. A detailed analysis shows that the constraint $c_{\text{hidden E}_8}$ eq. (6.24) is actually based on data points from the subset $W_5 = (0^{16})$. Hence, the subset without $W_5 = (0^{16})$ would lead to a higher value for the constraint $c_{\text{hidden E}_8}$. Thus, it is embedded into the more significant dataset, and the analysis for the second subset would have to take the higher value for $N_{\text{ur}}^{(2)}$ into account. As a remark, throughout the thesis and especially in chapter 8 we show that we are unable to find counterexamples to the derived contrast patterns even with an immense search effort.

6.2.3 The *dynamic hidden E_8* contrast pattern

In order to understand why the improved probability to find MSSM-like models did not equally increase the number of inequivalent MSSM-like models, the next step is to analyze the effect of the *hidden* E_8 contrast pattern in more detail. To do so, the equivalent MSSM-like $\mathbb{Z}_6$-II (1, 1) models from the *hidden* E_8 dataset are visualized with respect to the number of equivalent MSSM-like models that appear for various values of $N_{\text{ur}}^{(2)}(M)$ and inequivalent spectra, spectrum(M). Hence, in fig. 6.5 every inequivalent spectrum is represented by an individual bar. The bar height gives the number of models M that have the same spectrum in the dataset.

From fig. 6.5, it is clear that the models with small numbers of unbroken roots $N_{\text{ur}}^{(2)}(M) \in \{6, \ldots, 14\}$ are frequently sampled in the *hidden* E_8 dataset, while it seems to be very difficult to construct models with $N_{\text{ur}}^{(2)}(M) \geq 30$. This can be understood as an effect of the chosen search strategy. What appears in the contrast patterns is that the uniformly drawn coefficients for the gauge embedding matrix M, on average, cause a too strong breaking of the visible and hidden gauge group in order to allow for MSSM-like models. By redrawing again from the

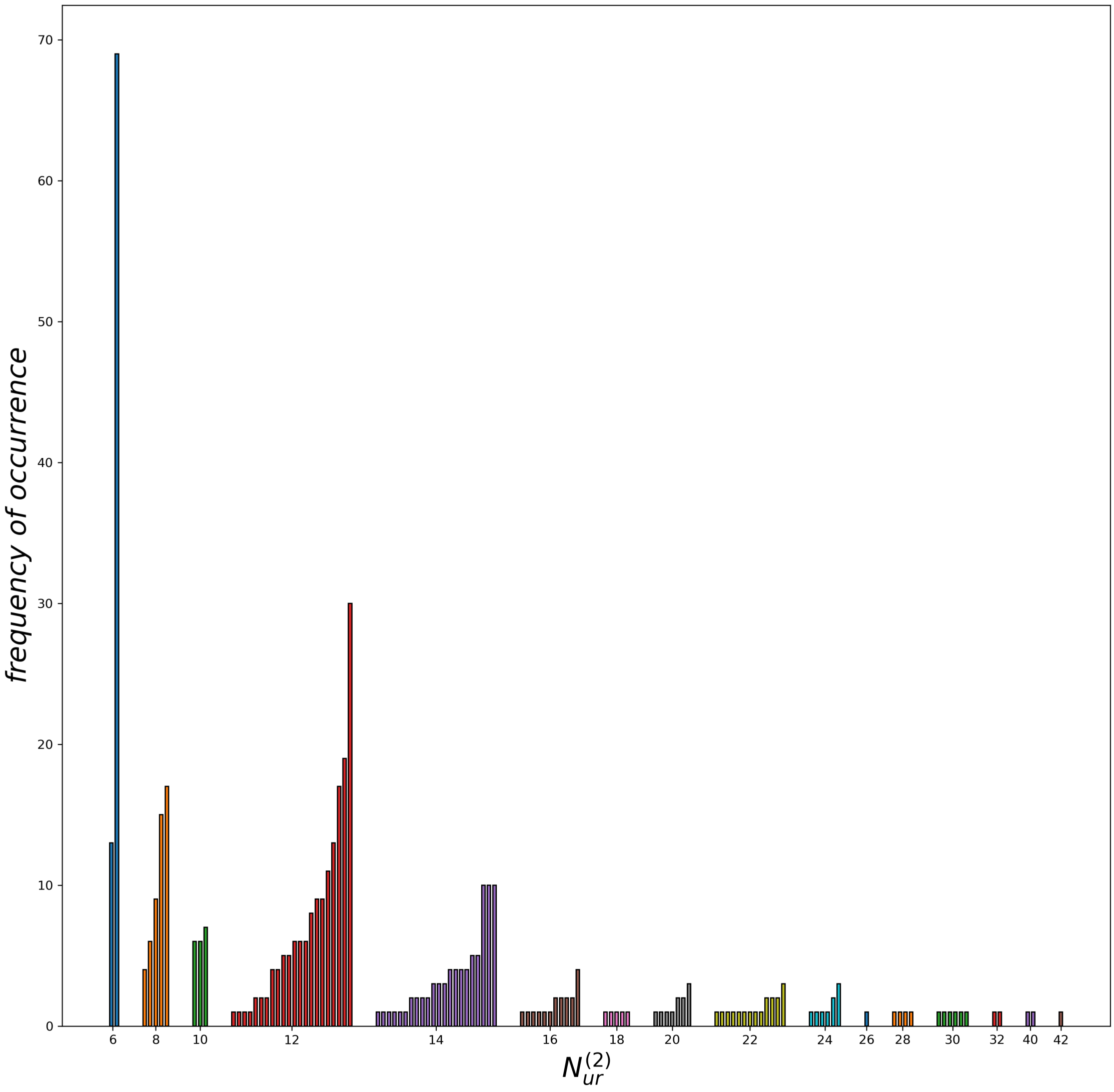


Figure 6.5: *Frequency of occurrence for inequivalent MSSM-like* $\mathbb{Z}_6$-II (1,1) *models for the number of unbroken roots* $N_{ur}^{(2)}(M)$ *from the hidden* E_8*. Based on the dataset* hidden E$_8$ *with the contrast pattern* $N_{\mathrm{ur}}^{(2)}(M) \geq 6$*. In detail, each bar represents an inequivalent MSSM-like model* M*. The number of duplicates of each model is shown by the height of the bar. As an example, for the green bars at* $N_{\mathrm{ur}}^{(2)}(M) = 10$ *there are three inequivalent MSSM-like models, each represented by one bar. These inequivalent models have 6, 6 and 7 duplicates in the whole dataset, respectively. Note that in this chart only those MSSM-like models appear that have* G'_{SM} *exclusively in the first* E_8*. Figure adapted from [21].*

uniform distribution, the probability mass function of the coefficients is cut-off for the constraint value. However, the newly drawn coefficients do then accumulate very close to the lower bound since these combinations have a higher likelihood to emerge. Nevertheless, it is unintended to modify the probability mass function since for further conservative investigations of the heterotic orbifold landscape an even exploration of the space $\mathbb{Z}^{128}$ is desired, and only the exception is made for the known pattern to set a Wilson line to zero. This pattern is easy to handle and compare. Moreover, especially for $N_{\mathrm{ur}}^{(2)}(M) = 22$, the bar chart shows a lot of distinct bars.

X	6	8	10	12	$\ldots$	20	22	$\ldots$	30	32	$\ldots$	40	42
$\mathrm{gr}(N_{\mathrm{ur}}^{(2)}(M) \geq X)$	1	5	6	6	$\ldots$	12	26	$\ldots$	48	75	$\ldots$	73	82

Table 6.4: *Growth rate* gr *for different threshold values X of the contrast pattern $N_{\mathrm{ur}}^{(2)}(M) \geq X$, estimated on the* hidden E_8 *dataset.*

This proposes that the diversity of inequivalent MSSM-like models lies in subspaces of the $\mathbb{Z}_6$-II (1,1) orbifold landscape where models have larger hidden sector gauge groups. Those areas are especially unlikely to reach due to the distribution for sampling. Furthermore, based on the dataset *hidden* E_8, it is possible to simulate the effect of higher threshold values X by computing the change of the growth rate for the contrast pattern $N_{\mathrm{ur}}^{(2)}(M) \geq X$. This leads to table 6.4. Therefore, it is not only the variety of models that seems to lie in the region with higher threshold values X, but also the growth rate is increasing and indicating that such models are easier to uncover as soon as the search reaches this subspace. Thus, it seems very promising to change the threshold value $X = 6$ of the contrast pattern $N_{\mathrm{ur}}^{(2)}(M) \geq X$ into a dynamic variable X, and in referring to this, the new constraint is named *dynamic hidden E_8*,

$$c_{\text{dynamic hidden } \mathrm{E}_8}(X) \;=\; \left\{N_{\mathrm{ur}}^{(2)}(M) \;\geq\; X\right\} . \tag{6.31}$$

A search based on the *dynamic hidden E_8* contrast pattern for various values of the threshold X and different constraints on the Wilson lines is performed. The new type of pattern, i.e. the dynamic variable X, introduces some freedom on the parameter handling in the search procedure. Details on this are postponed to the next section 6.2.4 in order to give a concise result discussion here. For evaluation of the contrast pattern the new dataset *dynamic hidden E_8* is obtained, see table 6.2. Compared to the *hidden E_8* dataset with 136 inequivalent MSSM-like $\mathbb{Z}_6$-II (1,1) models in total, it is now possible to find 415 inequivalent MSSM-like models with comparable search effort, i.e. a similar size of the dataset. This is already more than in any existing $\mathbb{Z}_6$-II (1,1) search [75, 79, 126, 127]. Hence, a significant improvement of the search for inequivalent MSSM-like models in the $\mathbb{Z}_6$-II (1,1) heterotic orbifold landscape is achieved. Moreover, this search is also able to solve the long-standing puzzle of the absence of MSSM-like models in the case $W_3 = W_4 = (0^{16})$. So far, it was not possible to find any MSSM-like model if M_5, of order $N_5 = 3$, is turned off, even though there is no theoretical obstruction for such a model to exist. The strongly increased probability to find MSSM-like models leads to the possibility of identifying two MSSM-like $\mathbb{Z}_6$-II (1,1) models with $W_3 = W_4 = (0^{16})$, $M_1^{W_3=(0^{16})}$ and $M_2^{W_3=(0^{16})}$, as can be seen in table 6.2. These models have hidden gauge groups,

$$G_{4D}^{(2)}\left(M_1^{W_3=(0^{16})}\right) \;=\; \mathrm{SU}(3) \times \mathrm{SU}(3) \times \mathrm{U}(1)^4 \;, \tag{6.32}$$

$$G_{4D}^{(2)}\left(M_2^{W_3=(0^{16})}\right) \;=\; \mathrm{SU}(5) \times \mathrm{U}(1)^4 \;. \tag{6.33}$$

In particular, $N_{\mathrm{ur}}^{(2)}(M_1^{W_3=(0^{16})}) = 12$ and $N_{\mathrm{ur}}^{(2)}(M_2^{W_3=(0^{16})}) = 20$ and consequently heavily benefit from the *dynamic hidden E_8* contrast pattern, as these models reside in regions of the landscape with high X value. Additionally, due to the vanishing Wilson lines in the $\mathbb{T}^2/\mathbb{Z}_3$ plane, these models are equipped with a phenomenologically appealing $\Delta(54)$ flavor symmetry [143]. Still note that the probability of finding a MSSM-like model with the condition $W_3 = W_4 = (0^{16})$ is far below the probabilities for the other Wilson line conditions.

In order to understand why the new search strategy is so successful and audit the first estimations made at the beginning of this section, a detailed analysis of the results is reasonable.

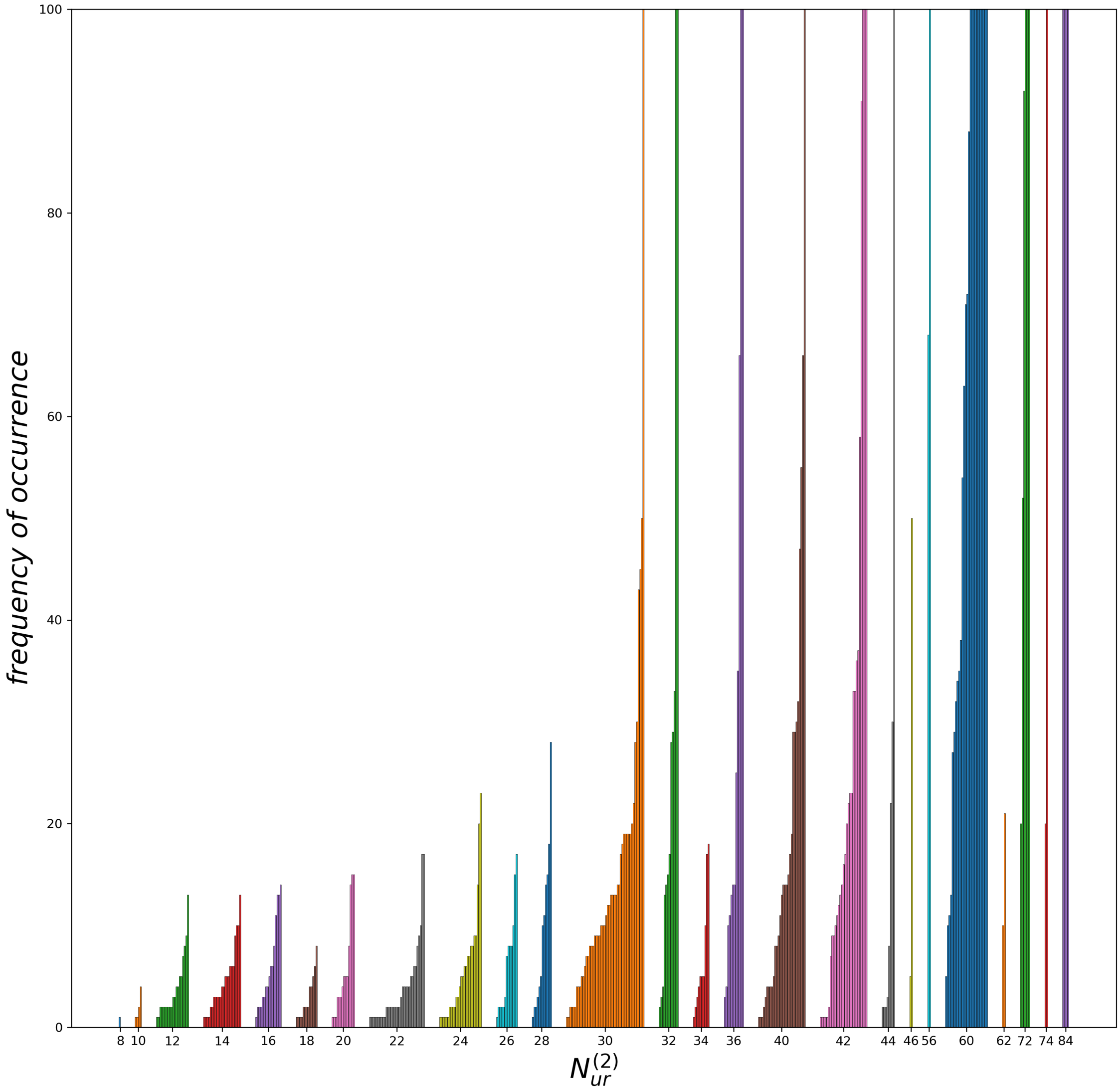


Figure 6.6: *Frequency of occurrence for inequivalent MSSM-like* $\mathbb{Z}_6$-II (1,1) *models for the number of unbroken roots* $N_{\mathrm{ur}}^{(2)}(M)$ *from the hidden* E_8*. Based on the dataset* dynamic hidden E$_8$ *with the contrast pattern* $N_{\mathrm{ur}}^{(2)}(M) \geq X$*. For details see fig. 6.5. Note that increasing the threshold value* X *leads to a more intense search in those areas of the* $\mathbb{Z}_6$-II (1,1) *orbifold landscape that were insufficiently sampled by the static search with* $X = 6$*. Figure adapted from [21].*

Astonishingly, it turns out that a huge fraction of the diversity of MSSM-like models lies in subspaces of the heterotic orbifold landscape where the hidden sector gauge group is large, see fig. 6.6. In more detail, using the *dynamic hidden* E_8 contrast pattern it is feasible to obtain novel MSSM-like models with $N_{\mathrm{ur}}^{(2)}(M) = X$ for $X \in \{34, 36, 44, 46, 56, 60, 62, 72, 74, 84\}$ and uncover plenty MSSM-like models for higher X values, e.g. with $X \in \{30, 40, 42\}$. The reason for this improvement, compared to the non-dynamic *hidden* E_8 (fig. 6.5), is that the new search strategy that led to fig. 6.6, adjusts to the regions with higher X values by constraining the probability mass function of the gauge embedding matrix M, i.e. parameters d_{kI} accordingly.

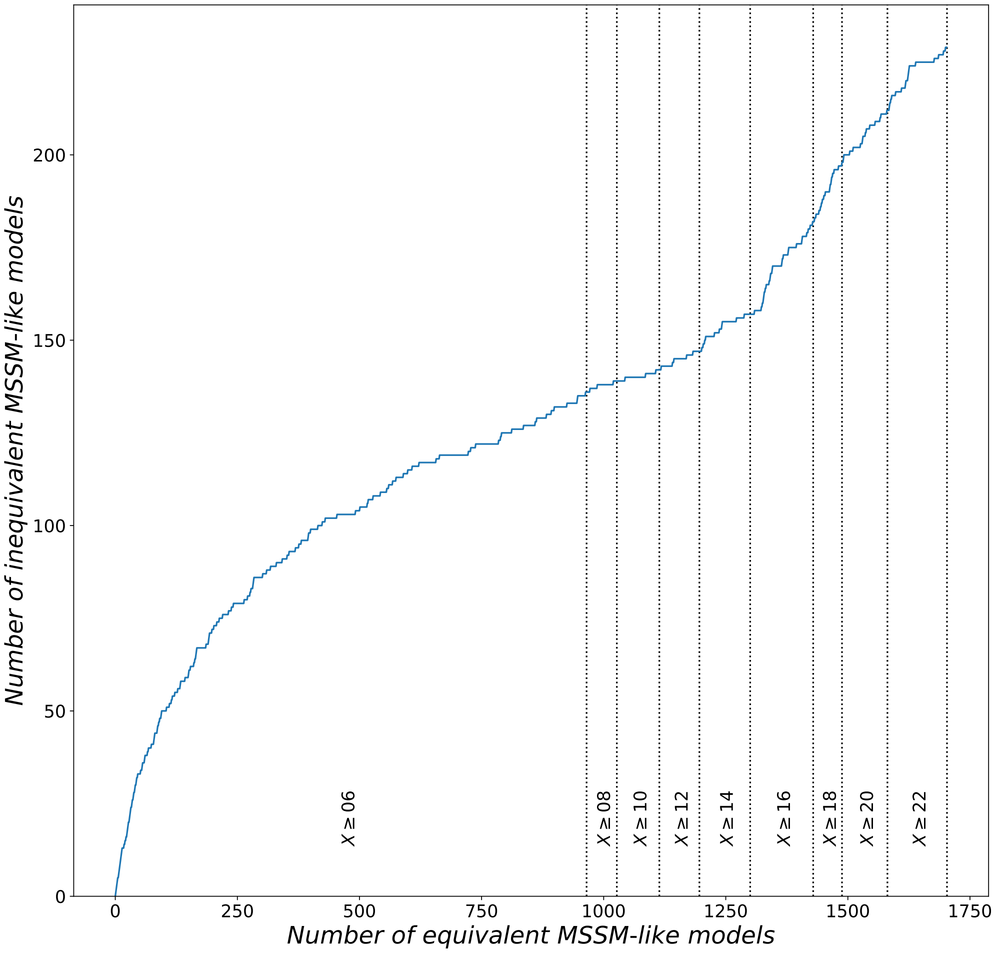


Figure 6.7: *Effect of the dynamic hidden E_8 search. On the vertical axis the number of inequivalent MSSM-like models and on the horizontal axis the number of equivalent MSSM-like models are given. They are chronologically ordered with respect to the search time. The dotted lines mark the X value used for the $N_{\text{ur}}^{(2)}(M) \geq X$ constraint.*

6.2.4 Inverse search strategy

Random searches are, in general, not exhaustive. This also accounts for the search developed in this thesis. During any random search process, the number of inequivalent MSSM-like models will follow a saturation curve [144]. This behavior is indicated by the first part $X \geq 6$ of fig. 6.7 where the inequivalent versus equivalent MSSM-like models are plotted. It shows a decreasing gradient of the function. Consequently, the effort for creating a new inequivalent MSSM-like model grows exponentially during sampling time. The same applies to the *dynamic hidden* E_8 search. However, it is possible to generate a small steep increase of novel MSSM-like spectra by changing the value for X, e.g. $X = 16$. After some time, the new constrained search follows the same behavior. Thus, any attempt to reach these results using a basic random search would take an unfeasible amount of computing time and is only a theoretical possibility rather than an alternative approach. Although it is possible to generate the steep increases, it remains an open question of how much computational effort should be invested for a certain value of X. Furthermore, the growth rates in table 6.4 and the actual search show, that it is much more efficient to search for high values of X, in order to get novel MSSM-like spectra quickly. Hence,

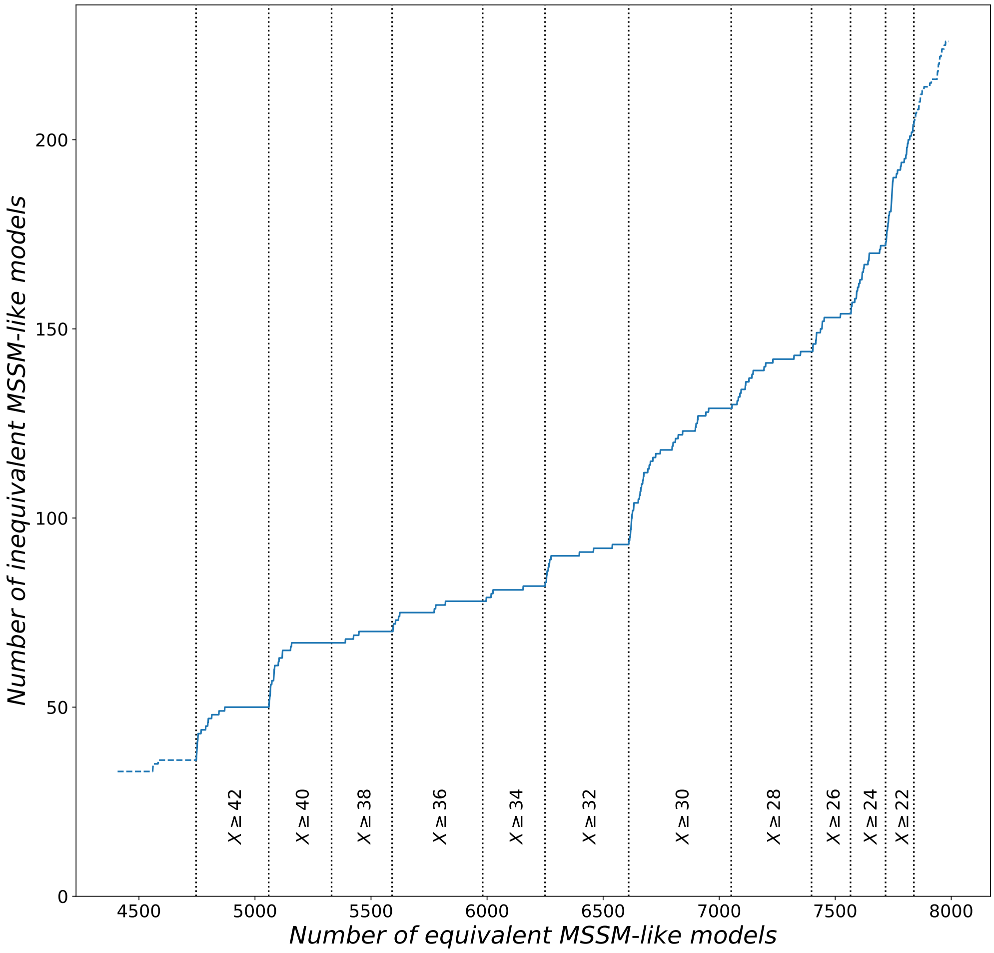


Figure 6.8: *Inverse search strategy for the dynamic hidden E_8 search. Plotted is a part of the cumulative function of new MSSM-like models. On the vertical axis the inequivalent number of MSSM-like models and on the horizontal axis the number of equivalent MSSM-like models are given. They are chronologically ordered with respect to the search time. The dotted lines mark the X value used for the $N_{\mathrm{ur}}^{(2)}(M) \geq X$ constraint.*

this leads to the suggestion of inverting the search process starting from high values of X and then easing the constraint until we hit the lower bound $X \geq 6$. As the lower regions always include the deeper ones this search gets broader and broader.

Additionally, this is supported by the fact that precisely the models in the higher X regions are the ones with very low probability to appear in an unconstrained random search. These models are the ones that pop up occasionally in the long tail of the saturation curve. This indicates the search to proceed since these models have compactification parameter configurations that are very unlikely to appear and hence need a long exploration time. The graph of inverse search in fig. 6.8 indicates that the inverse search rather strictly indicates if a particular subspace still has some novel MSSM-like models left or generates just already known ones, see e.g., $X \geq 42$. In other words, it shows the probability of finding a new inequivalent model. This quantity is related to the gradient of the curve. Hence, the sharp profile of the inverse search allows for a saturation check in order to quantify if a certain subregion should be explored further or not. One realization of this check is to investigate the inverse cumulative plot fig. 6.8. The estimate on the probability for new inequivalent MSSM-like models can be computed by[2],

$$\mathrm{Pr}\left(\text{inequiv. MSSM-like for } c_{\text{dyn.}}(X) \mid D_Y\right) \approx \frac{F_{D_Y}(D_{c_{\text{dyn.}}(X)}) - F_{D_Y}(\xi D_{c_{\text{dyn.}}(X)})}{(1-\xi)|D_{c_{\text{dyn.}}(X)}|} \tag{6.34}$$

where $D_Y = \{D_{c_{\text{dyn.}}(Y)} \text{ with } Y > X\}$ is the combined dataset of all previously constructed datasets with stronger constraints $c_{\text{dyn.}}(Y)$. The shorthand notation $c_{\text{dyn.}}(Y)$ refers to eq. (6.31). The cumulative function F_{D_Y} computes the number of inequivalent novel MSSM-like models of $D_{c_{\text{dyn.}}(X)}$ with respect to D_Y, i.e. only MSSM-like models that are not contained in D_Y can be counted as novel MSSM-like models. The approximation parameter ξ defines the proportion of the cumulative plot that is taken into account to approximate the gradient at the tail of the saturation sub-curve. A common value used in this thesis is to consider the last quarter, i.e. $\xi = \frac{3}{4}$. If the probability for novel MSSM-like models in eq. (6.34) vanishes, the search is assumed to be exhaustive for constraint $c_{\text{dyn.}}(X)$ and the value X can be lowered to explore the next region. This condition implies, that the last quarter of the cumulative graph is flat. Note that it is important for the analysis to consider all searches performed with higher values $Y > X$, i.e. the cumulative result. This indicates if the X-constrained search still produces novel MSSM-like spectra, or occasionally reaches the models from the already performed stronger constraints.

Note that the parameter ξ determines the strength of the estimator for the probability eq. (6.34). It can be taken to smaller values for a more exhaustive search or to larger values in order to get a more efficient, less exhaustive result. Nonetheless, since the construction of any orbifold model is independent, the search can be performed in parallel. We recommend to perform the search in parallel batches over multiple values X and check the saturation before new submissions to the batch-queuing system. As soon as a particular value of X can not fulfill the condition, this constraint is taken out of the submitted batch series, and a lower one is appended.

In fig. 6.8, not all subspaces are entirely flat in the last quarter. One should keep in mind that this plot does not indicate the overall probability to generate a MSSM-like model. Thus, even if there is a small remaining probability that novel MSSM-like models are in the subspaces, the search effort to find them, can be massive. The estimate of the computational effort is given by multiplying the probability eq. (6.34) with the overall probability of finding MSSM-like models in this subspace. The probability to find a MSSM-like model in the sub-spaces of $\mathbb{Z}_N$ orbifolds can be inferred from the sample proportions. A qualitative overview is given in fig. D.2. Indeed, this value provides an alternative, more efficiency-oriented indicator for exploring the landscape.

6.2.5 The *U-sector* contrast pattern

After an intense search with the static and dynamic hidden E_8 contrast pattern eqs. (6.24) and (6.31), we are equipped with a new and more informative dataset, that has 415 inequivalent MSSM-like models in $D^{N_{\text{ur}}} = D_{\text{hidden } E_8} \cup D_{\text{dynamic hidden } E_8}$, where $D^{N_{\text{ur}}}$ is obtained by combining the datasets *hidden* E_8 and *dynamic hidden* E_8 from table 6.2. Based on this new dataset, a search for further contrast patterns is reasonable. The same logic as in section 6.2.1 is applied, i.e. a decision tree is used to mine contrast patterns on the basis of the new combined dataset $D^{N_{\text{ur}}}$. It turns out that the remaining features $N^{(\alpha)}_{\mathrm{U}_a}(M)$ contain the next strongest contrast patterns for the data.

---

[2] Note, in practice, a large enough background set is necessary to estimate the probability. Likewise, for a very small number of MSSM-like values, we recommend to skip the probability computation and search for more models.

For computational reasons, it is necessary to downsample the background of ~~MSSM~~-like models. This implies choosing only a random fraction of ≈ 50% of the total dataset. Due to the enormous amount of data, this can be considered to be a valid approach. The decision tree has approximately the same statistics for the analysis in comparison to the whole dataset. Furthermore, for the limited and important MSSM-like models, all instances of these models are taken into account. As described in section 6.2.1, they are reduced to their inequivalent models to get a generalization measure based on this quantity. The $D^{\mathtt{train}}$ and $D^{\mathtt{val}}$ split is then performed with 33% of models in the validation set.

The decision tree hyperparameters are adjusted with respect to the same metric as before, i.e. `Recall`(MSSM-like) = 1. The goal is to classify all MSSM-like models correctly. However, it turns out that this is rather challenging due to two MSSM-like models. One of these models is misclassified during training, the other one during validation. These two MSSM-like models are the special $\Delta(54)$ models, where $W_3 = W_4 = (0^{16})$ from section 6.2.3. Because these models lie in a very specific subspace within the $\mathbb{Z}_6$-II (1,1) orbifold landscape and hence are tractable, it seems to be appropriate to sacrifice these types of models in the new contrast pattern, in order to get a sufficient reduction of the landscape. Furthermore, there are no new models expected so far in the subspace $W_3 = W_4 = (0^{16})$, at least for a reasonable amount of search effort. Note, that it is possible to tune the hyperparameters, i.e. the l_{12} from eq. (6.23), such that these models get also correctly classified as MSSM-like. However, then nearly all models are classified as MSSM-like and the expected growth rate eq. (6.18) of this pattern tends towards one, i.e. no gain at all.

Accepting the trade-off that two MSSM-like models are misclassified, a new contrast pattern can be identified from fig. 6.9, i.e.

$$c_{\text{U-sector}} = \left\{ N^{(1)}_{\text{U}_2}(M) \geq 2 \,,\, N^{(2)}_{\text{U}_3}(M) \leq 5 \right\} , \tag{6.35}$$

for a model M to have an enhanced likelihood of being MSSM-like. This contrast pattern is called *U-sector* constraint as it provides bounds on the bulk matter fields (section 2.4.1), depending on $\alpha = 1, 2$, respectively. Employing this new constraint on top of the former ones, the estimated growth rate reads,

$$\text{gr}\left(c_{\text{U-sector}}, D^{N_{\text{ur}}}_{\text{MSSM-like}}, D^{N_{\text{ur}}}_{\cancel{\text{MSSM}}\text{-like}}\right) \approx \frac{0.999}{\frac{1\,209\,687}{2\,215\,901}} \approx 1.8 \,. \tag{6.36}$$

Let us discuss some distinctions of the *U-sector* constraint. To begin with, the computation of gr(c) given in eq. (6.36): Contrary to the *hidden* E_8 contrast pattern, it is not ensured that the *U-sector* contrast pattern covers models which fulfill the constraint of G'_{SM}, section 6.1.2, in both E_8 factors. For $N^{(2)}_{\text{ur}}(M) \geq 6$ models with G'_{SM} in both E_8 factors fulfill the more restrictive condition $N^{(\alpha)}_{\text{ur}}(M) \geq 8$ for $\alpha = 1, 2$. This is not the case for the *U-sector* constraint. See fig. 6.9 and follow the right branch. This branch, according to the first node, only contains models with $G'_{\text{SM}} \subseteq G^{(1)}_{4D}$ and hence the contrast patterns might only hold for this preceding condition. Therefore, the growth rate in eq. (6.36) is estimated using only those models where G'_{SM} is solely in the first E_8 factor. However, remember that models with G'_{SM} in both E_8 factors are always contained in the *dynamic hidden* E_8 search.[3] Secondly, the arrangement of the *U-sector* constraints hides that $\Delta(54)$ MSSM-like models excluded by the constraint $N^{(2)}_{\text{U}_3}(M) \leq 5$ do obey the subsequent constraint $N^{(1)}_{\text{U}_2}(M) \geq 2$. The decision tree selects the order by taking the statistics into account. It is necessary to misclassify these two MSSM-like

---

[3] In chapter 8, we will argue that even the models with G'_{SM} in both E_8 factors are contained in the *U-sector* search.

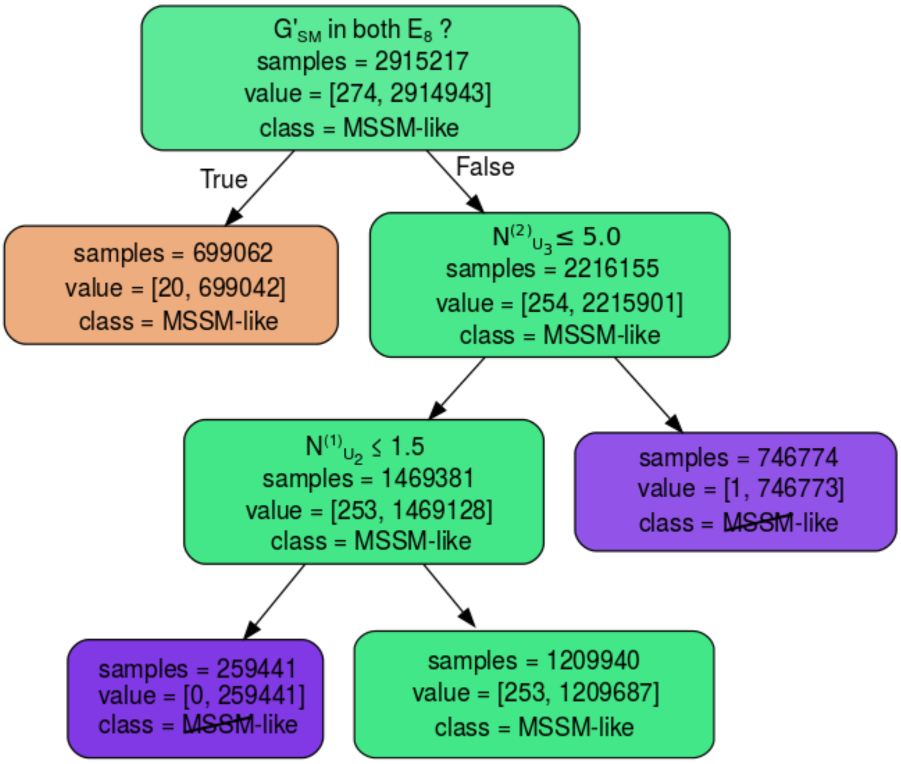


Figure 6.9: *Visualization of the decision tree trained on the* hidden E_8 *and* dynamic hidden E_8 *datasets from table 6.2. The numbers are evaluated on the training set* D^{train}*. The hyper-parameter* l_{12} *from eq.* (6.23) *is changed to* $l_{12} = 10^5$*. We obtain the contrast pattern* ($N^{(1)}_{U_2} \geq 2$ *and* $N^{(2)}_{U_3} \leq 5$) *for MSSM-like models from the right branch of the tree. Notice that this tree misclassifies one MSSM-like model after the* $N^{(2)}_{U_3}$*-split in order to get better performance. See also figure 6.4 for further details. Figure adapted from [21].*

models for optimization reasons. However, it is still appealing that all MSSM-like models (with $G'_{\text{SM}} \in E_8^{(1)}$) obey this constraint. This observation deserves further investigations towards analytic conditions for MSSM-like models.

Realizing the *U-sector* contrast pattern in the search algorithm represented by figure 6.3 and executing a search, the final dataset, called *U-sector* is collected, see table 6.2. Once again, the contrast data mining procedure shows its strength, and the probability of finding MSSM-like models increased further, as shown in figure 6.10. The *U-sector* contrast pattern generalizes to the $\mathbb{Z}_6$-II (1,1) landscape and uncovers several novel MSSM-like models. To analyze the benefit of the new constraint, the appropriate dataset for comparison is the *dynamic hidden E_8* dataset. The dynamic search is the background, and the $c_{\text{U-sector}}$ constraint is used on top of it. This results in a growth of 395 inequivalent MSSM-like models to 459 models. Finally, merging all datasets yields in total 468 inequivalent MSSM-like $\mathbb{Z}_6$-II (1,1) models. In conclusion, all former searches for MSSM-like $\mathbb{Z}_6$-II (1,1) models [75, 79, 126, 127] were significantly exceeded by rejecting those regions in the $\mathbb{Z}_6$-II (1,1) orbifold landscape where most likely no MSSM-like model exists and additionally balancing the search which is uniform in the $\mathbb{Z}^{128}$ space but not in the physical space of MSSM-like models.

Furthermore, some of our contrast patterns might even be necessary conditions for all MSSM-like models. In the following section 6.3, it is demonstrated how efficiently the derived constraint,

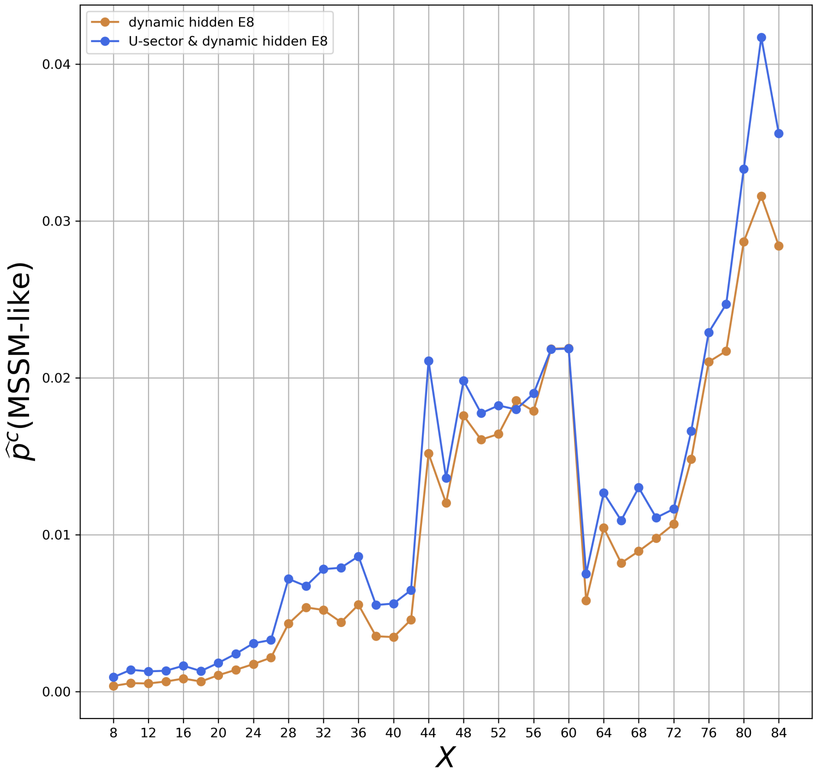


Figure 6.10: *Estimated probability $\hat{p}^c(MSSM\text{-}like)$ to find MSSM-like models for the threshold value X of two different dynamic hidden E_8 searches. The contrast pattern c refers to $N_{\mathrm{ur}}^{(2)} \geq X$ or $N_{\mathrm{ur}}^{(2)} \geq X$ in combination with the* U-sector *constraint. Figure adapted from [21].*

i.e. *hidden* E_8 contrast pattern, can be transferred to the other orbifold geometries and actually can find novel models. The lower bound of this constraint will be orbifold geometry specific.

6.3 Generalized contrast patterns

The contrast patterns were acquired solely on the $\mathbb{Z}_6$-II (1, 1) orbifold geometry. Nevertheless, the fundamental insights from this analysis can be applied to other geometries as well. The contrast patterns are based on the particle spectrum and not on properties of the geometry itself. The idea of the *hidden* E_8 contrast pattern can be used rather directly in the context of other orbifold geometries. Unfortunately, this is not the case for the *U-sector* contrast pattern. The number of bulk matter fields $N^{(\alpha)}_{\mathrm{U}_a}$ for $a = 1, 2, 3$ depends on the individual entries in the twist vectors of a given orbifold geometry, see eq. (6.21). These entries are not sorted according to some physical quantity, rather the sorting of the entries in the twist vectors is given in ascending order of rotation angles. Hence, it is not clear why a particular U-sector might be special among the different sectors, i.e. if a special status of a U-sector is related to the rotation angle or some nontrivial relation between all sectors. Permutation invariant constructions, i.e. the elementary symmetric polynomials in $N^{(\alpha)}_{\mathrm{U}_a}$,

$$N^{(\alpha)}_{\mathrm{U}_1} + N^{(\alpha)}_{\mathrm{U}_2} + N^{(\alpha)}_{\mathrm{U}_3} \tag{6.37}$$

$$N^{(\alpha)}_{\mathrm{U}_1} N^{(\alpha)}_{\mathrm{U}_2} + N^{(\alpha)}_{\mathrm{U}_1} N^{(\alpha)}_{\mathrm{U}_3} + N^{(\alpha)}_{\mathrm{U}_2} N^{(\alpha)}_{\mathrm{U}_3} \tag{6.38}$$

$$N^{(\alpha)}_{\mathrm{U}_1} N^{(\alpha)}_{\mathrm{U}_2} N^{(\alpha)}_{\mathrm{U}_3} \tag{6.39}$$

have been tried, but could not be used as a replacement of the contrast patterns which act directly on the individual U-sectors, neither did they contribute as distinct patterns.

Therefore, the *dynamic hidden* E_8 search is directly utilized and the U-sector thresholds are analyzed for the individual orbifold geometries $\mathbb{O}$. The result is summarized in table 6.7. In order to apply the *dynamic hidden* E_8 contrast pattern to all $\mathbb{Z}_N$ orbifold geometries the lower bound $X_{\mathrm{min}}(\mathbb{O})$ of $N^{(2)}_{\mathrm{ur}}(M) \geq X_{\mathrm{min}}(\mathbb{O})$ for each $\mathbb{Z}_N$ orbifold geometry has to be identified by,

$$X_{\mathrm{min}}(\mathbb{O}) \;=\; \min\left(\left\{ N^{(2)}_{\mathrm{ur}}(M) \;\middle|\; M \in D_{(\mathbb{O}\text{ from [126, 127]})} \right\}\right) \,. \tag{6.40}$$

Remember that $N^{(2)}_{\mathrm{ur}}(M) \geq X_{\mathrm{min}}(\mathbb{Z}_6\text{-II (1,1)}) = 6$. To get an estimate for these bounds, the datasets collected using the basic search technique fig. 4.1 of [126, 127] can be used as background searches. Therefore, we extract the dataset of the relevant orbifold geometry $\mathbb{O}$ from the combined dataset $D_{(\mathbb{O}\text{ from [126, 127]})}$ that correspond to the different $\mathbb{Z}_N$ orbifold geometries. Then, the conventional search in [126, 127] compares to the *phenomenology* search from our investigations in the previous chapters. Concretely, for this scan there is no lower bound on the hidden sector gauge group, i.e. $N^{(2)}_{\mathrm{ur}} \geq 0$. Thus, it is possible to analyze the results of the traditional search [126, 127] to derive the lower bounds $X_{\mathrm{min}}(\mathbb{O})$ for all $\mathbb{Z}_N$ orbifold geometries. As it is already known which constraints are searched for, the construction of a decision tree is not necessary. Note that at this point, it is assumed that similar patterns can be found in the other orbifold geometries. The actual training of the tree can interchange the patterns on the features. However, this straightforward approach will turn out to be useful. The results are stated in table 6.5. Moreover, the background datasets $D_{(\mathbb{O}\text{ from [126, 127]})}$ allow us to focus the *dynamic hidden* E_8 contrast pattern on thresholds greater than $X_{\mathrm{min}}(\mathbb{O})$, i.e. $N^{(2)}_{\mathrm{ur}} > X_{\mathrm{min}}(\mathbb{O})$, to save computational resources.

The final results of the search are given in table 6.6. Here for each $\mathbb{Z}_N$ orbifold geometry, the numbers of inequivalent MSSM-like orbifold models that were found using the *dynamic hidden* E_8 contrast pattern is compared to those numbers in the literature [126, 127]. The *dynamic hidden* E_8 search is capable of finding several new inequivalent MSSM-like orbifold models in nearly

orbifold	$\mathbb{Z}_4$	$\mathbb{Z}_6$-I	$\mathbb{Z}_6$-II	$\mathbb{Z}_7$	$\mathbb{Z}_8$-I			$\mathbb{Z}_8$-II		$\mathbb{Z}_{12}$-I	$\mathbb{Z}_{12}$-II
geometry $\mathbb{O}$	$\forall$	$\forall$	$\forall$	(1,1)	(1,1)	(2,1)	(3,1)	(1,1)	(2,1)	$\forall$	(1,1)
$X_{\min}(\mathbb{O})$	4	12	6	56	6	6	4	0	4	4	6

Table 6.5: *Lower bounds* $X_{\min}(\mathbb{O})$ *on the number of unbroken roots* $N_{\text{ur}}^{(2)}$ *from the hidden* E_8 *factor for MSSM-like orbifold models. Obtained from [126, 127] for the* $\mathbb{Z}_N$ *orbifold geometries* $\mathbb{O}$ *listed in the first column of table 6.6. In case of "* $\forall$*" all lattices for a given* $\mathbb{Z}_N$ *orbifold geometry have the same lower bound.*

all orbifold geometries. In particular, the numbers for the different $\mathbb{Z}_6$-II orbifold geometries, as well as the $\mathbb{Z}_{12}$-I case, have increased using the contrast patterns. Additionally, focusing on the effect of the models from the novel search on the combined dataset from the literature, it is noticeable that nearly all orbifold geometries get a boost of new models even in cases where the total number was lower than the one from the literature, i.e. $\mathbb{Z}_8$-II $(1,1)$, $\mathbb{Z}_8$-II $(2,1)$, $\mathbb{Z}_{12}$-II $(1,1)$. Moreover, we observe 13 additional MSSM-like models from [126, 127] for the orbifold geometry $\mathbb{Z}_6$-II $(1,1)$, which we used to develop all the constraints of this chapter. This geometry was of particular interest, and the intense searches performed could not find all MSSM-like models. However, these models satisfy all the constraints derived in section 6.2. Hence, an extended scan with contrast patterns can find these models. Furthermore, the appearance of the $\mathbb{Z}_7$ MSSM-like model is also a great achievement of our contrast patterns. So far, this model was only observed in orbifold-specific scans [31, 127]. Besides, the $\mathbb{Z}_6$-I orbifold geometry is unusual in terms of the $X_{\min}$ value. This is the only $\mathbb{Z}_N$ geometry (except of $\mathbb{Z}_7$) with a lower bound higher than needed for the SM gauge group. The lower bound $X_{\min}(\mathbb{Z}_6\text{-I}) = 12$ (see table 6.5), is computed for those models where G'_{SM} appears in one E_8 factor only. Thus, our search algorithm could miss MSSM-like $\mathbb{Z}_6$-I models where each E_8 factor contains G'_{SM}. Analyzing these models individually, a lower bound $X_{\min}(\mathbb{Z}_6\text{-I}) = 10$ for these cases is found. Furthermore, it seems that for some orbifold geometries like $\mathbb{Z}_8$-II $(1,1)$ the standard approach is advantageous. However, a comparison is difficult since it is not known how much computational effort was needed to produce the results from the literature. Moreover, for $\mathbb{Z}_8$-II $(1,1)$ our contrast patterns seem to be not as useful, since the lower bound for $N_{\text{ur}}^{(2)} = 0$ (see table 6.5) is practically not present and thus does not reduce this sub-landscape. Hence, it seems that our search algorithm is too involved for such geometries and the extra work in computing constraints is not rewarded. However, the merged datasets in table 6.6 shows that the contrast patterns could still significantly enhance the numbers of inequivalent MSSM-like orbifold models in these geometries. Thus, the search algorithm is able to find new MSSM-like orbifold models in corners of the landscape that are missed by the conventional approaches.

On the basis of the 'merged' datasets the decision tree approach to derive the *U-sector* constraints, as in the case of the $\mathbb{Z}_6$-II (1,1) orbifold geometry section 6.2.5, can be used. The results are given in table 6.7. In nearly all orbifold geometries it is possible to get a $\texttt{Recall}\big(D^{\texttt{val}}\big) = 1.00$ and no MSSM-like model is missed in the training set.

Only $\mathbb{Z}_6$-II orbifold geometries $(1,1)$, $(3,1)$ had a $\texttt{Recall}\big(D^{\texttt{val}}\big) = 0.99$, i.e. produced a *false negative.* For the $\mathbb{Z}_6$-II orbifold geometries $(1,1)$ and $(3,1)$ the decision tree sacrificed one MSSM-like model during training, while $\mathbb{Z}_6$-II $(2,1)$ misclassified two models during training. Overall, this similarity in the behavior of the different $\mathbb{Z}_6$-II geometries shows how strongly the MSSM-like patterns are connected to the point group of the orbifold geometries. Only, the $\mathbb{Z}_6$-II $(4,1)$ orbifold geometry can avoid these misclassifications.

number of inequivalent MSSM-like orbifold models

orbifold geometry		from [126]	from [127]	[126, 127]	dynamic hidden E_8	'merged'
$\mathbb{Z}_4$	(2,1)	128	138	161	125	179
	(3,1)	25	26	29	33	33
$\mathbb{Z}_6$-I	(1,1)	31	30	31	31	31
	(2,1)	31	30	31	31	31
$\mathbb{Z}_6$-II	(1,1)	348	363	390	468	481
	(2,1)	338	349	372	395	443
	(3,1)	350	351	390	415	482
	(4,1)	334	354	404	407	464
$\mathbb{Z}_7$	(1,1)	0	1	1	1	1
$\mathbb{Z}_8$-I	(1,1)	263	256	271	248	271
	(2,1)	164	155	164	144	164
	(3,1)	387	377	418	408	430
$\mathbb{Z}_8$-II	(1,1)	638	1 833	1 903	1 259	2 289
	(2,1)	260	489	508	349	555
$\mathbb{Z}_{12}$-I	(1,1)	365	556	568	610	625
	(2,1)	385	554	572	607	625
$\mathbb{Z}_{12}$-II	(1,1)	211	352	371	365	435

Table 6.6: *Inequivalent MSSM-like orbifold models for all $\mathbb{Z}_N$ orbifold geometries, see also [21]. The second and third column show the number of inequivalent models known in the literature. The fourth column states the number of inequivalent models combining the datasets of the literature. The number of inequivalent models constructed by the search fig. 6.3 using the* dynamic hidden E_8 *contrast pattern eq.* (6.31) *are listed in the fifth column. The last column, gives the final results, i.e. the numbers of inequivalent MSSM-like orbifold models obtained by merging the three datasets.*

Another special case is the $\mathbb{Z}_8$-II $(2,1)$ orbifold geometry. In order to get a growth rate larger than one the decision tree has to split the set of MSSM-like models into two sets at the first node with a constraint on $N^{(1)}_{\mathrm{U}_3}$. Then, for both sets a second split takes $N^{(2)}_{\mathrm{U}_3}$ into account, resulting in a growth rate $\mathrm{gr}(c) = 1.78$ for the first set containing only 3% of the inequivalent MSSM-like $\mathbb{Z}_8$-II $(2,1)$ models from $D^{\texttt{train}}$ and a growth rate of $\mathrm{gr}(c, D) = 1.01$ for the second set containing 97% of the models, respectively.

In this context, it is interesting that for the $\mathbb{Z}_8$-I $(1,1)$ geometry, one can increase the growth rate by tuning the hyperparameters and train another decision tree. In detail, this results in a constraint $c'_{\mathrm{U-sector}} = \big\{N^{(1)}_{\mathrm{U}_3} \geq 8, N^{(2)}_{\mathrm{U}_3} \leq 25\big\}$ that has $\texttt{Recall}\big(D^{\texttt{val}}\big) = 1.00$ and $\mathrm{gr}(c'_{\mathrm{U-sector}}) = 2.18$. The only disadvantage is that for this tree one MSSM-like model from the training data is misclassified.

Moreover, let us mention that the expected growth rate for the $\mathbb{Z}_4$ orbifold geometry is significantly higher than it was for the $\mathbb{Z}_6$-II. For this geometry the *U-sector* constraint could be even more important than the *hidden E_8* constraint. A detailed analysis like in the case of

orbifold geometry		$N^{(1)}_{\mathrm{U}_1}$	$N^{(1)}_{\mathrm{U}_2}$	$N^{(1)}_{\mathrm{U}_3}$	$N^{(2)}_{\mathrm{U}_1}$	$N^{(2)}_{\mathrm{U}_2}$	$N^{(2)}_{\mathrm{U}_3}$	gr(c)
$\mathbb{Z}_4$	(2,1)			≥ 4			≤ 1	5.32
	(3,1)			≥ 4			≤ 1	6.92
$\mathbb{Z}_6$-I	(1,1)	≥ 13		≥ 14				2.79
	(2,1)	≥ 13		≥ 14				2.78
$\mathbb{Z}_6$-II	(1,1)		≥ 2				≤ 5	1.81
	(2,1)		≥ 2				≤ 5	1.60
	(3,1)		≥ 2				≤ 5	1.70
	(4,1)		≥ 2				≤ 5	1.86
$\mathbb{Z}_8$-I	(1,1)			≥ 4			≤ 25	1.22
	(2,1)			≥ 4			≤ 25	1.23
	(3,1)			≥ 8				2.21
$\mathbb{Z}_8$-II	(1,1)			≥ 4			≤ 41	1.61
	(2,1)			≤ 3			≤ 1	1.78
				≥ 4			≤ 41	1.01
$\mathbb{Z}_{12}$-I	(1,1)	≤ 10		≥ 2				1.24
	(2,1)	≤ 10		≥ 2				1.24
$\mathbb{Z}_{12}$-II	(1,1)	≥ 2					≤ 5	1.71

Table 6.7: *Decision tree derived U-sector constraints for $\mathbb{Z}_N$ orbifold geometries, obtained from the merged datasets of table 6.6. The $\mathbb{Z}_7$ orbifold geometry is neglected, because only one MSSM-like model is available.*

$\mathbb{Z}_6$-II seems necessary.

Contrary to the expectation, that the configuration of the *U-sector* constraints is highly sensitive to the specific orbifold, table 6.7 shows global patterns. These patterns generalize over the orbifold geometries. Foremost, it is remarkable that for the hidden sector, nearly all constraints are '$\leq$' and in the U_3*-sectors*. For the visible sector, a clear preference to '$\geq$' is given. On a refined level, the twist vector does not have any significant effect, and the specific threshold value is correlated to the order of the point group. One should use this global pattern and take the weakest constraint among a family of orbifolds, e.g. $N^{(3)}_{\mathrm{U}_1} \geq 4$ for $\mathbb{Z}_8$ orbifolds. Overall, this indicates more structure in the heterotic orbifold landscape and that necessary conditions for MSSM-like models can follow from further considerations in this direction.

6.4 Chapter summary

In this chapter, we reduce the heterotic orbifold landscape massively. Analytical constraints based on the SM gauge group and the Weyl symmetry of the landscape exclude vast areas of the compactification parameter space. By a statistical analysis with decision trees, we uncover contrast patterns that distinguish between areas of the landscape with a tiny probability to hold a MSSM-like model and others. Based on these constraints, we develop a search algorithm fig. 6.3 that focuses on the phenomenologically viable part of the heterotic orbifold landscape. The apparent result is that it was possible to substantially improve the search strategy for MSSM-like models in the heterotic orbifold landscape table 6.2. Thus, it was possible to identify models with very special and far from generic properties in heterotic orbifolds, i.e. models with $W_3 = (0^{16})$ and hence $\Delta(54)$ flavor symmetry.

For the technique itself, it is to mention that it is challenging to uncover suitable features table 6.3 that have a monotonic behavior concerning the successive search. Moreover, for very restrictive contrast patterns, the random sampling of the d_{kI}-coefficients takes rather long. This is because the constraints have to modify the probability mass function of the coefficients heavily. Learning a better probability mass function for drawing the coefficients might be useful, see chapter 9 for potential methods. Finally, the contrast pattern technique seems to be far more valuable than expected on the task to find the characteristic features of MSSM-like models with a stringy origin. Further statistical evidence will be analyzed in chapter 8. They might help to find the necessary conditions for beyond the SM physics to connect to a UV-complete framework.

Chapter 7

Bottom-up guide for the heterotic orbifold landscape

In the previous machine learning approaches, we studied the patterns of the heterotic orbifold landscape in order to find explicit constructions of MSSM-like models, as were defined in section 2.6. However, an obvious question is if the heterotic orbifold landscape includes the exact MSSM of particle physics. The concrete orbifold parameters of a string compactification that lead to a model that is in agreement with all experimental facts from particle physics would be a definite answer. Due to the enormous number of 4D string models, the question arises, if it is possible to predict the orbifold geometry with the highest likelihood to generate a particular MSSM-like model, i.e. the exact MSSM. Indeed, the contrast mining approach from section 6.3, showed orbifold specific threshold values for the generic contrast patterns. Inspired by this observation, we will show in this chapter that the particular phenomenology of the MSSM-like models is strongly correlated to the point group P. Hence, in terms of phenomenological properties, the landscape is divided into subregions.

For this task, the dataset consists of the known MSSM-like orbifold models constructed in the literature, as well as novel MSSM-like models from the advanced search strategy fig. 6.3. The feature space is formed of phenomenological properties, like the number of various types of vector-like exotics, and the value of the gaugino beta function coefficients. The intention is to develop a classifier that predicts the orbifold origin of MSSM-like bottom-up models. Thus, we train the machine learning model to predict the orbifold point group P that has the highest probability of reproducing a given MSSM-like model, see fig. 7.1.

7.1 A feature space of vector-like exotics

The dataset D of MSSM-like orbifold models consist of 126 783 inequivalent MSSM-like string models[1] from the $\mathrm{E}_8 \times \mathrm{E}_8$ heterotic string, compactified on 52 Abelian orbifolds $\mathbb{O}$.[2] In order to obtain such a huge dataset, constructed models from several landscape investigations were merged. In detail, the largest dataset available for the $\mathbb{Z}_{N_1}$ orbifolds was constructed in section 6.3 were the datasets [126, 127] were extended by novel MSSM-like models created by the dynamic hidden E_8 contrast pattern search. This procedure is extended to the $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$

---

[1] The MSSM-like orbifold models are available as arXiv ancillary files of [20, 21].

[2] From the 138 Abelian orbifolds with $\mathcal{N} = 1$, 23 $\mathbb{Z}_2 \times \mathbb{Z}_2$ orbifolds forbid chiral matter spectra [118, 147]. Furthermore, the orbifolds are reduced to those that can generate non-Abelian flavour symmetries ($\mathbb{Z}_6 \times \mathbb{Z}_6$ is the only exception) as investigated in [127].

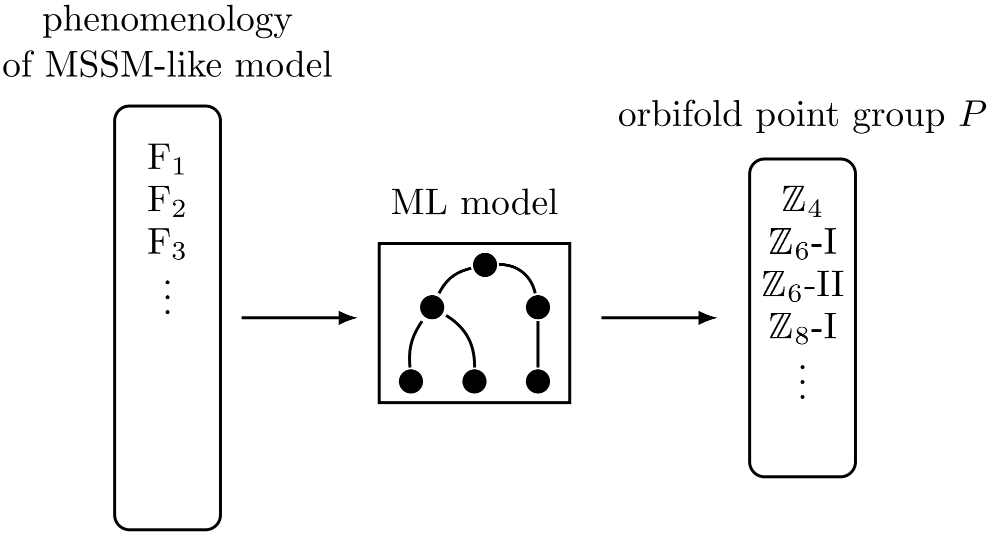


Figure 7.1: *Overview of the machine learning (ML) method. A predictive model (boosted decision tree) is trained to infer the probability* $p(P_i)$ *of the orbifold point groups* P_i *to reproduce a given MSSM-like bottom-up model. Note that* $\sum_i p(P_i) = 1$*. As input, we specify certain phenomenological features* F_k*. Figure adapted from [20].*

MSSM-like models using the same contrast pattern for the enhanced search algorithm and merge them with the models from [126, 127] to obtain the dataset D.

A review of the exotics in the inequivalent MSSM-like dataset can be found in table D.1 and table D.2. In the first row of the tables the different orbifold point groups P_i that produce the MSSM-like string model are given. There are 14 different point groups and hence $i = 1, \ldots, 14$. The second row states the number of inequivalent models per P_i. An important observation for the procedure is given by the fact that the dataset is highly imbalanced. In detail, the number of inequivalent MSSM-like models for the different point groups are spread over a wide range. As an example, $\mathbb{Z}_4$ contains only a few hundred MSSM-like models, while $\mathbb{Z}_2 \times \mathbb{Z}_4$ holds several ten-thousand MSSM-like models. Besides, since only one MSSM-like orbifold model, based on the $\mathbb{Z}_7$ point group, is part of the dataset, this point group is excluded from further considerations.

To predict the orbifold origin based on phenomenological features F_k as depicted in fig. 7.1 we compute the following phenomenological features of MSSM-like models:[3]

- **$\mathbf{F_1}$: # SM singlets**: The average MSSM-like model contains $\mathcal{O}(100)$ SM singlets. These can be understood as right-handed neutrinos, that are charged under hidden gauge group factors, i.e U(1)'s and the non-Abelian G_{hidden}. Recall from section 2.6 that SM singlets are relevant for the decoupling of vector-like exotics.

- **$\mathbf{F_2 - F_{36}}$: # vector-like exotics for each type**: Recall from section 2.6 that MSSM-like models are defined to allow for vector-like exotics. The total number of different vector-like exotics that appear in the landscape is of small cardinality, i.e. 43 distinct exotic particles (table D.1 ,table D.2). Thus, it is feasible to use the number of vector-like exotics of each type as features. Note, 8 types of vector-like exotics solely appear in the unique MSSM-like $\mathbb{Z}_7$ model and are neglected. Phenomenologically, vector-like exotics can be appealing for model building as new elementary particles. For instance,

---

[3] For convenience we use # as abbreviation for 'number of'.

so-called leptoquarks were proposed, to address flavor anomalies [148–151]. However, they can compromise gauge coupling unification and can lead to cosmological problems. Thus, we choose to search for "(almost) perfect MSSM models" that have (almost) no vector-like exotics.

- $\mathbf{F_{37}}$: **Heavy top quark from bulk**: This scenario describes a stringy mechanism that can explain the large mass hierarchy between the top quark and the lighter quarks. Therefore, we require an $\mathcal{O}(1)$ renormalizable Yukawa coupling for a left-chiral top quark doublet q_3, its right-chiral top quark partner $\bar{u}_3$ and the up-type Higgs h_u. If these fields originate from the twisted sectors their coupling can be suppressed if they are not located at the same fixed point. This is not the case for bulk matter. Actually, if the left-chiral top quark doublet q_3, its right-chiral top quark partner $\bar{u}_3$ and the up-type Higgs h_u distribute among the three untwisted sectors U_a, for $a = 1, 2, 3$ (see eq. (2.51)), respectively, the string selection rules [156, 157] and in especially the conservation of R-charges [31, 32, 158] allow for this unsuppressed coupling, which is related to the ten-dimensional gauge coupling constant [152–155]. Thus we incorporate this feature as binary-value, i.e. true or false, for a given model.

- $\mathbf{F_{38}}$: **# vanishing Wilson lines**: Non-Abelian flavor symmetries of the 4D effective theory can emerge in heterotic orbifolds from the twisted strings, that are localized on the fixed points of the extra-dimensions (section 2.4.2). If Wilson lines vanish, an additional permutation symmetry of some of the twisted strings can emerge such that the full flavor symmetry becomes non-Abelian. Because non-Abelian flavor symmetries are phenomenologically interesting [159], this feature is included indirectly in the feature space by the number of vanishing Wilson lines, see also [64, 67, 126, 127, 143]. A non-zero value indicates the presence of non-Abelian flavor symmetry.

- $\mathbf{F_{39}}$: **Anomalous U(1)**: In general 4D models from the heterotic orbifold landscape are equipped with an anomalous gauge factor $\mathrm{U}(1)_{\mathrm{anom}}$. However, it turns out that the anomaly of this pseudo-anomalous U(1) can be canceled by a universal Green–Schwarz anomaly cancellation mechanism [160]. Furthermore, this procedure is phenomenologically appealing, as it sets the scale for a Froggatt–Nielsen-like mechanism [161], where some SM singlets develop non-vanishing VEVs. Thus, these VEVs spontaneously break the additional U(1) factors and some of these VEVs decouple parts of the vector-like exotics from the spectrum. Moreover, they are responsible to give the lighter quarks and leptons masses. Even though 99% of the MSSM-like orbifold models possess an additional $\mathrm{U}(1)_{\mathrm{anom}}$, this interesting feature is included as a binary value.

- $\mathbf{F_{40}}$: **hidden sector beta-function coefficient:** The last feature is connected to spontanous supersymmetry breaking through hidden sector gaugino condensation [47, 50, 76, 126, 162–165]. Therefore, it is assumed that the vector-like exotics decouple without breaking the non-Abelian hidden sector gauge groups G_{hidden}. For a strongly coupled hidden sector the gauginos build a condensate and spontaneously break supersymmetry. To determine the energy scale Λ_{hidden} where this happens, we have to analyze the running of the coupling and the corresponding Landau pole where the coupling becomes infinite, for each G_{hidden}. It turns out that,

$$\Lambda_{\text{hidden}} \;=\; M_{\text{GUT}}\,\exp\left(-\frac{16\pi^2}{2\,b\,g^2(M_{\text{GUT}})}\right)\;, \tag{7.1}$$

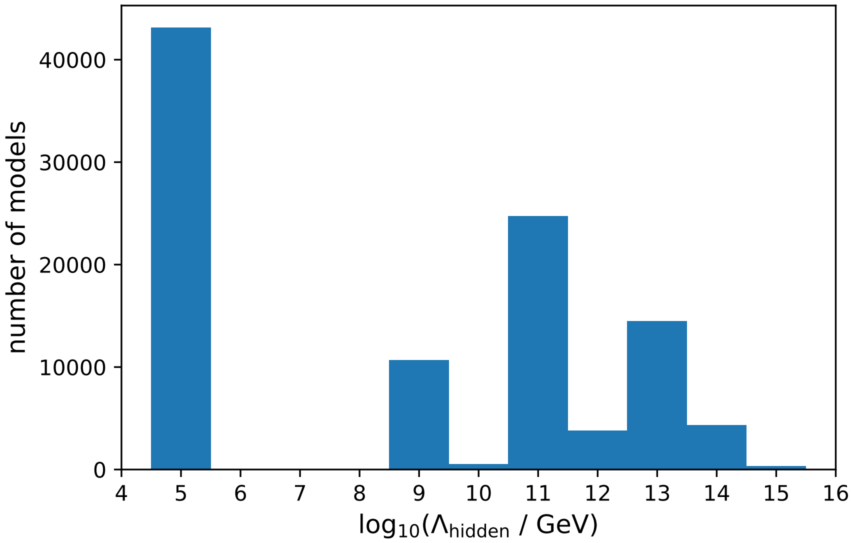


Figure 7.2: *Number of MSSM-like orbifold models from dataset D for the hidden sector gaugino condensation scale* Λ_{hidden}*. Excluded are a few models with smaller* Λ_{hidden} *as well as* $\approx 24\,000$ *models with* $b = 0$*, i.e.* $\Lambda_{\text{hidden}} = 0$*. Note, for the peak at* $\Lambda_{\text{hidden}} = 10^5\,\text{GeV}$ *the scale of supersymmetry breaking is far too low. It consists mainly of MSSM-like orbifold models from the point groups* $\mathbb{Z}_2 \times \mathbb{Z}_4$ *and* $\mathbb{Z}_4 \times \mathbb{Z}_4$ *(each contributing* $\approx 19\,000$ *models). Figure adapted from [20]*

with $M_{\text{GUT}} \approx 3 \cdot 10^{16}\text{GeV}$ and $g^2(M_{\text{GUT}}) \approx 1/2$. We assume in eq. (7.1) that the gauge coupling constants of the MSSM coincide at the GUT scale approximately with the one of the hidden sector gauge group. Thus, to determine the gaugino condensation scale, we can compute the known beta-function coefficient $b = 3C_2 - \sum_{\mathbf{r}} \ell(\mathbf{r})$ of G_{hidden} with $C_2 = N$ and $\ell(\mathbf{N}) = 1/2$ for SU(N), and $C_2 = 2(N-1)$ and $\ell(\mathbf{2N}) = 1$ for SO($2N$). The summation over $\mathbf{r}$ runs over the chiral matter in the spectrum, where $\mathbf{r}$ is given by the representation of the matter under G_{hidden}. The phenomenological preferred energy scale is $\Lambda_{\text{hidden}} \approx 10^{13}\text{GeV}$ for supersymmetry breaking at least around TeV scale, as the gravitino mass can be approximated by

$$m_{3/2} \approx \frac{\Lambda^3_{\text{hidden}}}{M^2_{\text{Planck}}} , \tag{7.2}$$

and supersymmetry breaking is communicated to the observable sector by gravity. Note, that for this estimate string threshold corrections [166–168] are neglected. Those can increase the value for Λ_{hidden}, such that a reasonable range to consider are three orders of magnitued [74]. This corresponds to the lower bound $\Lambda_{\text{hidden}} \gtrsim 10^{10}\text{GeV}$, i.e. $b \gtrsim 10$ [76]. In our MSSM-like models, the distribution of Λ_{hidden} is given in fig. 7.2 for all point groups and in fig. 7.3 for $\mathbb{Z}_6$-II (1,1) only.

These features define the basis that spans the whole feature space. The original dataset, consisting of the physical spectra, had 126 783 inequivalent MSSM-like string models. The attribute inequivalent is linked to the full orbifold geometry, i.e. inequivalence is defined here within each different orbifold geometry separately. The dataset transformed into the feature space corresponds to a set of 112 670 inequivalent data points $(F, \mathbb{O})$. However, neglecting the orbifold origin and comparing the feature vectors F results in 106 009 inequivalent vectors.

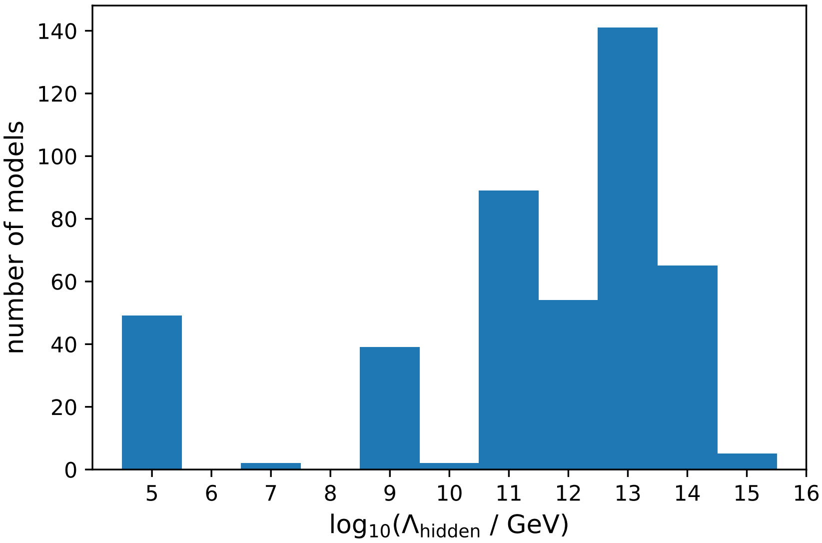


Figure 7.3: *Number of* $\mathbb{Z}_6$-II $(1,1)$ *MSSM-like orbifold models for hidden sector gaugino condensation scale* Λ_{hidden} *(models with* $b = 0$*, i.e.* $\Lambda_{\text{hidden}} = 0$*, are excluded). Figure adapted from [20].*

Hence, this introduces a substantial internal, unavoidable, misclassification error. In detail, this would not train a well-defined function but rather have identical input values (feature vectors F) that lead to different target values $\mathbb{O}$. As we will train a deterministic algorithm, in terms of its prediction, the final model will decide to map all instances of a certain F to the dominant $\mathbb{O}$. The reason for this ambiguity is the hidden sector that is not taken into account in detail, in our feature space, as the feature vector characterizes the properties concerning the MSSM only. Thus, models with different hidden sectors can yield the same feature vector [80].

However, combining the feature vectors with the point group, i.e. (F, P_i) results in 106 027 inequivalent data points. Contrary to the above situation, there is nearly no overlap if the point group P_i is taken as label. Hence, this label is a natural choice for the prediction task. This result connects strongly to the contrast patterns behavior concerning the point group (see section 6.3). The contrast patterns showed a strong correlation with the point groups and not with the specific lattice. Thus, the tuples (F, P_i) build the machine learning dataset D.

Closing this discussion, we point out that the reduction to the inequivalent subset of the data is necessary to avoid data leakage (appendix C.1). Concretely, the duplicates can get separated in test and training set, and hence the performance on the test set is no longer the performance to unseen data but to the duplicates of the training set (compare to section 6.2.1). As we want to predict the origin of the MSSM that is certainly not contained in the dataset, this generalization property is of utmost importance.

7.2 Point group classifier from 4D physics

To tackle this classification task, several methods from machine learning may be valuable. The two most widely applied and successful methods are neural networks and boosted decision trees. As our dataset is highly imbalanced, the neural network approach becomes disfavored. To handle imbalanced data in neural networks is still tricky [195], even though research in this direction leads to improvements, e.g. specialised loss functions [196].

For the non-neural network tools of machine learning, several hyperparameters exist to consider unbalanced data. In case of decision trees we already used in section 6.2.1 the `class_weight` argument to adjust the importance of the minority class. This is one reason why a boosted decision tree will be used in this imbalanced classification task. A boosted decision tree consists of an ensemble of simple classification trees (appendix C.6). A single predictive model from the ensemble is called a weak learner since they do not have the capacity to reach sophisticated results on the classification task. In the classification task here, a single decision tree was only able to reach a performance of `f1-macro` $= 0.66$, after hyperparameter optimization. Combining weak learners, it is possible to create a single strong learner. The resulting prediction is then a weighted majority vote of the collection of weak learners. However, in order to contribute different information to the voting, the weak learners have to be trained with different goals. In the boosting approach, this is achieved by a successive training of the weak learners, where the misclassified training data of the previous trees are weighted with a higher value in order to enforce the next weak learner to classify these models correctly. The weights of the previously correctly classified data are decreased accordingly, see also [197].[4]

Training of the classifier

The first task is to split our dataset D of inequivalent tuples (F, P_i) into 80% training $D^{\texttt{train}}$ and 20% test data $D^{\texttt{test}}$. The test set $D^{\texttt{test}}$ is kept aside for the evaluation of the trained classifier. $D^{\texttt{train}}$ is used to perform a grid search for the optimal hyperparameters, i.e. each possible combination of hyperparameter values, from a predefined parameter set, is used to train the classifier in a 5-fold cross validation (`CV`) (for an explanation of `CV` see appendix C.4). The metric to evaluate the trained classifiers is `f1-macro`, which is suitable for imbalanced data. We use for hyperparameter tuning the scikit-learn [189] module `GridSearchCV` and for the boosted decision tree the LightGBM implementation [200]. The large-scale hyperparameter search results in:

- `class_weight='balanced' (default=None)`: The argument 'balanced' automatically adjust the class weights $\texttt{w}_i$ inversely proportional to the class frequency of occurrence in the input data, i.e.

  $$\texttt{w}_i = \frac{\left|D^{\texttt{train}}\right|}{|P| \cdot |D^{\texttt{train}}(P_i)|} \,. \tag{7.3}$$

  where for this dataset, we have the number of classes $|P| = 14$ and $\left|D^{\texttt{train}}(P_i)\right| = \{F \in D^{\texttt{train}} \mid Y_{\text{correct}}(F) = P_i\}$ for each class. $\left|D^{\texttt{train}}\right|$ is the total number of data points.

- `learning_rate=0.2 (default=0.1)`: The learning rate for boosting. Hence, this value determines how strong the weights of the misclassified models get adjusted for the next weak learner.

- `min_child_samples=8 (default=20)`: In order to get good predictions on the minority classes it was beneficial to allow the individual trees to split to lower number of minimum samples per child node.

- `min_child_weight=0.01 (default=`10^{-3}`)`: Regularization method. Minimum sum of instance weight (hessian, i.e. second derivative of objective function) needed in a child node. Roughly speaking, a measure for the impurity of the data in the child node. Stops if the impurity is very low and hence regularizes the trees.

---

[4] In gradient boosted decision trees, the correction of the weights is generalized to the optimization of an arbitrary differentiable loss function [198, 199]. In our multiclass prediction this is the softmax objective function.

- `n_estimators=1500 (default=100)`: Number of weak learners to train.
- `num_leaves=50 (default=31)`: Maximum number of leaves a single decision tree learner can have.

Note, multiple classification algorithms were examined as alternatives: k-nearest-neighbors (kNN), linear discriminant analysis (LDA), random forest, support vector machines (SVM) and fully connected neural networks with softmax classification.[5] These other techniques are not able to outperform the `f1-macro` score of LightGBM. Only another boosted decision tree algorithm, XGBoost [202], yields comparable results. Besides, we constructed an ensemble of these different estimators, where the prediction of the ensemble is a weighted linear combination of the predictions of each individual estimator [203]. None of these approaches led to an improvement of more than $\sim 1\%$ for the `f1-macro` score. Hence, the positive effects of these ensembles do not justify the additional obscurification of the classification algorithm. In particular, the usage of LightGBM as a single estimator allows us to read out and interpret the inner structure of the boosted decision tree by visualizing the feature importance, see fig. 7.4. One obtains that, the SM singlets and the hidden sector beta function are frequently used to separate the different point groups.

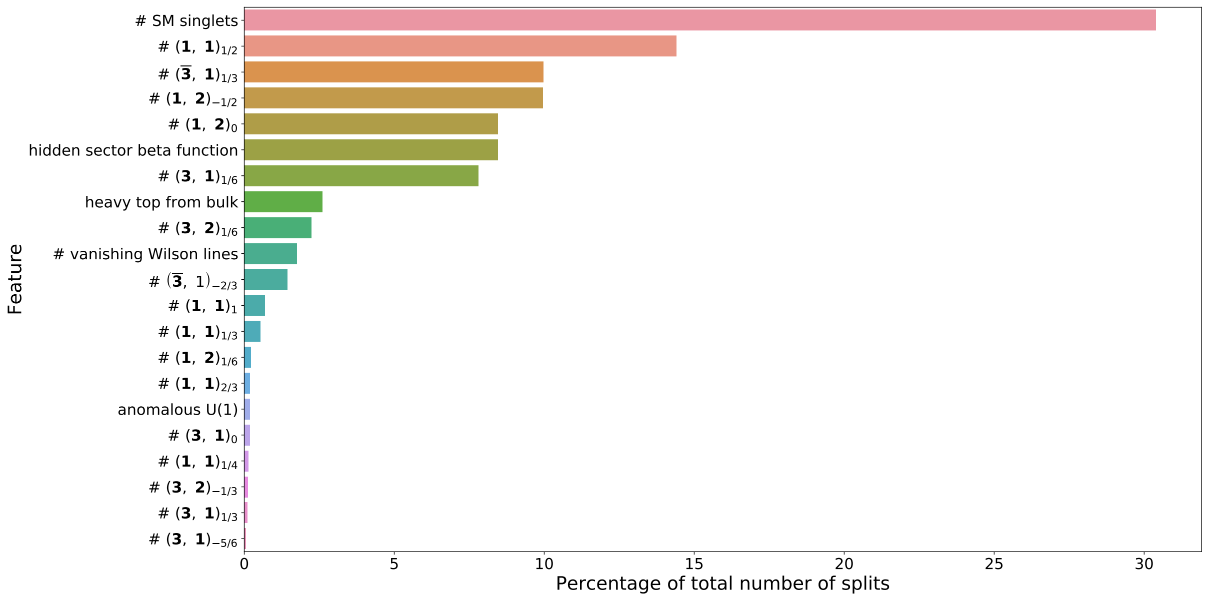


Figure 7.4: *The importance of the features for the boosted decision tree, based on the numbers of times the feature is used for a threshold condition in the model. Obtained from the* `feature_importances_` *method of LightGBM. We give the numbers as percentage and rank the features accordingly. As a result, the feature "# SM singlets" turns out to be used in 30% of the nodes and is the most important feature for our boosted decision tree. Figure adapted from [20].*

Evaluation of the predictions

The performance of the boosted decision tree on the test set $D^{\texttt{test}}$ is estimated in a classification report, see table 7.1. The overall high values for the `f1-score`, even for the minority classes,

---

[5] The implementations from scikit-learn [189] are utilized for the non-neural network algorithms. The Keras API [201] is used for the neural networks.

indicates that our classifier is well balanced and did not get heavily biased by the majority classes. For the upcoming predictions the `Precision` value is also of special interest. This value indicates how reliable a certain prediction is, i.e. for the lowest value of 0.75 in case of $\mathbb{Z}_8$-I, we have a 0.25 percent chance that this area belongs to a different point group P_i that is misclassified as *false positive* to the class $\mathbb{Z}_8$-I. However, the capability of the classifier to correctly identify the point group origin of the individual feature vectors is astonishing. It was unknown that the spectrum of vector-like exotics, in combination with some additional phenomenologically interesting features, allows for such a clear-cut inference of the point group.

point group P_i	`Precision`	`Recall`	`f`$_1$`-score`	support
$\mathbb{Z}_4$	0.85	0.85	0.85	33
$\mathbb{Z}_6$-I	0.88	0.88	0.88	8
$\mathbb{Z}_6$-II	0.82	0.75	0.79	305
$\mathbb{Z}_8$-I	0.75	0.78	0.76	125
$\mathbb{Z}_8$-II	0.86	0.84	0.85	444
$\mathbb{Z}_{12}$-I	0.95	0.90	0.92	232
$\mathbb{Z}_{12}$-II	0.69	0.68	0.69	82
$\mathbb{Z}_2 \times \mathbb{Z}_2$	0.88	0.83	0.85	255
$\mathbb{Z}_2 \times \mathbb{Z}_4$	0.97	0.98	0.98	8 155
$\mathbb{Z}_2 \times \mathbb{Z}_6$-I	0.84	0.59	0.69	211
$\mathbb{Z}_3 \times \mathbb{Z}_3$	0.99	1.00	0.99	483
$\mathbb{Z}_3 \times \mathbb{Z}_6$	0.87	0.88	0.87	1 019
$\mathbb{Z}_4 \times \mathbb{Z}_4$	0.98	0.98	0.98	9 113
$\mathbb{Z}_6 \times \mathbb{Z}_6$	0.93	0.96	0.94	741
macro avg	0.88	0.85	0.86	21 206

Table 7.1: *Classification report of the boosted decision tree classifier. Evaluated on D^{test} of 21 206 feature vectors. The support gives the number of feature vectors with point group P_i in D^{test}. In gray we highlight the three highest values for the `f1-score`. For the full classification matrix see [20].*

7.3 Predicting the stringy origin of the MSSM

After we have reviewed the predictions of our classifier, the next step is to slightly predict beyond the dataset space, towards the exact MSSM. Concretely, a feature vector without any vector-like exotics is given to the classifier to infer its orbifold origin. However, extrapolations with machine learning models are, in general, rather difficult as no confidence intervals are given for the predictions. In this particular case, the feature vector lies outside the border of the dataset D. However, we use additional features besides the numbers of vector-like exotics, e.g. the number of SM singlets and the hidden sector beta-function coefficient. These features are not a priori fixed for the exact MSSM spectrum. Hence, we can scan over a parameter space to predict the origin of exact MSSM spectra with a varying number of SM singlets and values for the hidden sector gaugino beta function. Thus, the inference for the MSSM also involves non-trivial features that are embedded in the orbifold dataset, i.e. orbifold models from the training set have values for these features above and below the values used in the feature vectors of this prediction task.

Moreover, the experience gained with regularized decision trees in the context of contrast mining (see chapter 6) indicates that these models can be utilized for predictions slightly beyond the dataset. The general procedure of a decision tree is to divide up the feature space into smaller subspaces. Each of those subspaces gets assigned to a class of the prediction task. Concretely, the class is identified based on the dominant class of the models from $D^{\mathtt{train}}$ that lie in this subspace. However, a decision tree necessarily leaves open subspaces at the boundary of the training set. Since the feature vector of the MSSM is outside the region of feature vectors of MSSM-like orbifold models, the MSSM will lie in one of these open subspaces. Still it will get a meaningful prediction assigned, due to the models from the training set, that reside in the same open subspace. Then, an ensemble of decision trees gives an additional regularization that improves the generalization to a broader range in feature space. In this way, regularized boosted decision trees can be used to get meaningful predictions.

Parameter scan for the exact MSSM

For the parameter scan we generate a new dataset $D^{\mathtt{MSSM}}$ of feature vectors $F^{\mathtt{MSSM}}$ that have a zero entry for all types of vector-like exotics, i.e. $F_k^{\mathtt{MSSM}} = 0$ for $k = 2, \dots, 36$. Furthermore, for each of the 4 combinations of the binary valued features "# vanishing Wilson lines" and "heavy top from bulk"[6], we generate the following sub-set of feature vectors:

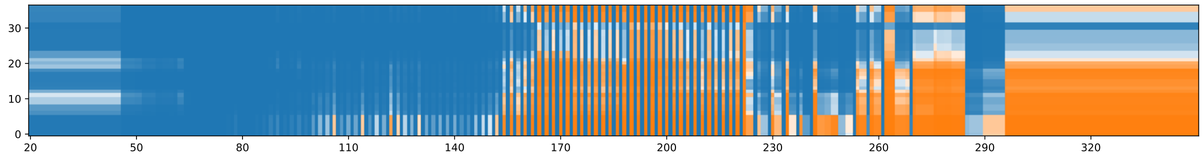
- For each $N \in \{20, \dots, 350\}$ and $b \in \{0, \dots, 36\}$ we create one feature vector with $F_1^{\mathtt{MSSM}} = N$ and $F_{40}^{\mathtt{MSSM}} = b$,

where the sets for N and b represent the values for the # SM singlets and the hidden sector beta function that appeared in D, respectively. Hence, in total we generate $37 \cdot 330$ feature vectors for each combination of the binary valued features. For each $F^{\mathtt{MSSM}}$ of the dataset $D^{\mathtt{MSSM}}$ we predict the corresponding point group P_i with the boosted decision tree, according to fig. 7.1. The outcomes of this prediction are illustrated in fig. 7.5. In order to emphasize how confident the boosted decision tree is, we introduced a transparency parameter α that indicates if one point group dominated the prediction, or other point groups reached comparable likelihoods. It turns out that overall, the parameter space is dominated by the two classes, $\mathbb{Z}_2 \times \mathbb{Z}_4$ and $\mathbb{Z}_4 \times \mathbb{Z}_4$. These classes build up the majority of MSSM-like orbifold models in the dataset D with 55 429, and 48 812 MSSM-like models, respectively. Moreover, they achieve the second and third highest `f1-scores` and thus are reliable, see table 7.1. Beyond that, for MSSMs without a heavy top from bulk, the $\mathbb{Z}_2 \times \mathbb{Z}_2$ orbifold point group (with 1 711 MSSM-like models) occupies a large fraction of the prediction area, especially in cases with SM singlets < 100 , see fig. 7.5c and fig. 7.5d. Furthermore, for MSSMs without a heavy top from bulk and without a vanishing Wilson line, several orbifold point groups become relevant in distinct areas of fig. 7.5d.

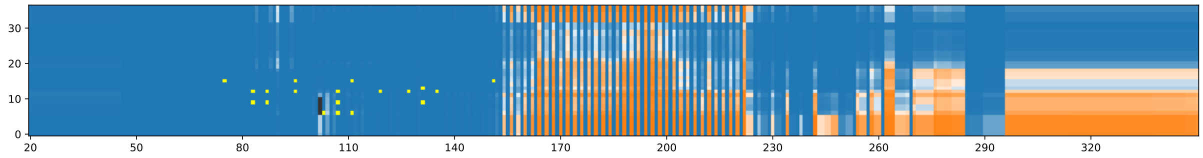
Finally, there is a further exceptional ingredient that strengthens the prediction for the $\mathbb{Z}_2 \times \mathbb{Z}_4$ orbifold point group. Namely, we observe the first "almost perfect" MSSM-like orbifold models (marked as yellow points in fig. 7.5b) in the orbifold geometry $\mathbb{Z}_2 \times \mathbb{Z}_4$ (2, 4) and (1,6). Their particle spectrum is very close to the exact MSSM and only contains additional Higgs-pairs. Thus, the boosted decision tree has clear evidence to estimate this point group as the origin of an exact MSSM spectrum. In the upcoming section, we will present one representative of the "almost perfect" MSSM-like orbifold models, as those models were unknown in the literature to exist.

---

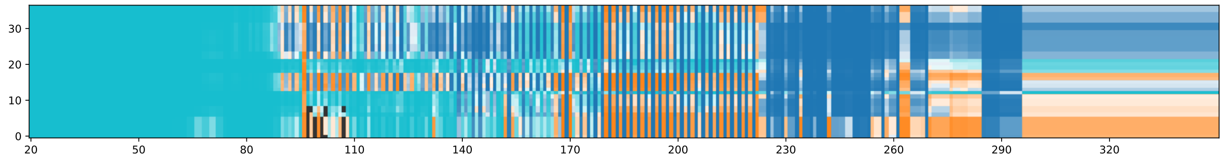
[6] For the "anomalous U(1)" feature the predictions turned out to be identical. Hence, this feature is set to true for the prediction task.



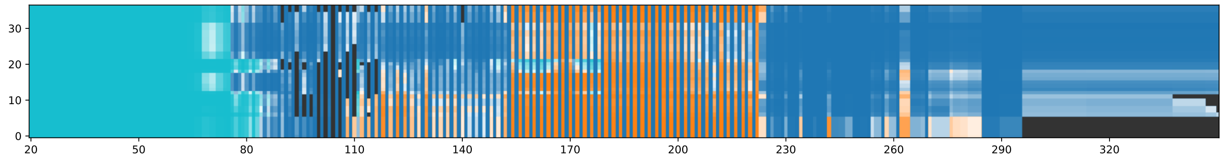
(a) *With heavy top from bulk and vanishing Wilson lines.*



(b) *With heavy top from bulk but without vanishing Wilson lines. The small black region corresponds to* $\mathbb{Z}_8$-I*. In yellow, we highlight 20 almost perfect MSSM-like orbifold models, see section 7.4.*



(c) *Without heavy top from bulk but with vanishing Wilson lines. The small black regions correspond to either* $\mathbb{Z}_6$-II *or* $\mathbb{Z}_8$-I*.*



(d) *Without heavy top from bulk and without vanishing Wilson lines. The black regions: (i)* 90 - 115 *SM singlets* $\mathbb{Z}_8$-I *(ii)* $b > 20$ *and* ≈ 90 *SM singlets* $\mathbb{Z}_{12}$-I *(iii)* $b > 30$ *and* 140 *SM singlets* $\mathbb{Z}_6$-I *(iv) SM singlets* > 290 *and* $b \lesssim 5$ $\mathbb{Z}_3 \times \mathbb{Z}_6$ *and for* $5 \lesssim b \lesssim 10$ $\mathbb{Z}_6 \times \mathbb{Z}_6$.

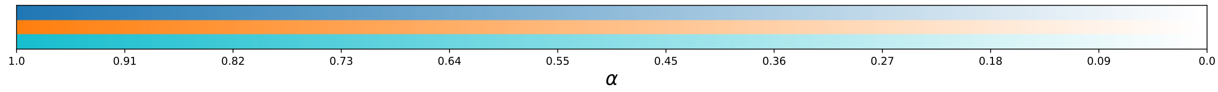


Figure 7.5: *Predicted orbifold point groups for the MSSM for four different cases (a) - (d) depending on the features "heavy top from bulk" and "vanishing Wilson lines": in each case, the horizontal axis corresponds to the number of SM singlets (i.e. right-handed neutrinos), while the vertical axis gives the beta-function coefficient* b *that sets the scale of supersymmetry breaking via hidden sector gaugino condensation. The colors are associated to point groups as follows: (i) blue:* $\mathbb{Z}_2 \times \mathbb{Z}_4$*, (ii) orange:* $\mathbb{Z}_4 \times \mathbb{Z}_4$*, (iii) turquoise:* $\mathbb{Z}_2 \times \mathbb{Z}_2$*. Black indicates a different point group than the three dominant ones. The transparency* $\alpha := p(1^{\text{st}}) - \frac{3}{4}p(2^{\text{nd}})$ *of each pixel indicates the difference between the highest and the second highest probabilities of the predictions,* $p(1^{\text{st}})$ *and* $p(2^{\text{nd}})$*, respectively. Note that* $\sum_i p(P_i) = 1$*. Hence, the color gets very transparent if* $p(1^{\text{st}}) \approx p(2^{\text{nd}})$*. Figures adapted from [20].*

7.4 An almost perfect MSSM-like orbifold model

For the particular case of "almost perfect" MSSM-like orbifold models, we present here an explicit construction from the orbifold geometry $\mathbb{Z}_2 \times \mathbb{Z}_4$ (2,4).[7] The entries of the gauge embedding matrix turn out to be

$$M = \begin{pmatrix} \boxed{\left(\ 0,\ 0,\ 0,\ 0,\ \tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{3}{2},\ \tfrac{3}{2}\right)} & \boxed{\left(\ 0,\ \tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{1}{2},\ 2\right)} \\ \boxed{\left(-\tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4},\ 1,-1,\ 0,-1,\ \tfrac{1}{4}\right)} & \boxed{\left(-\tfrac{3}{4},\ 0,\ 0,\ 0,\ 0,\ 0,\ \tfrac{1}{2},-\tfrac{3}{4}\right)} \\ & \\ \boxed{W_1^{(1)} = (0^8)} & \boxed{W_1^{(2)} = (0^8)} \\ \boxed{W_2^{(1)} = (0^8)} & \boxed{W_2^{(2)} = (0^8)} \\ \boxed{\left(-\tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{5}{4},-\tfrac{5}{4},\ \tfrac{5}{4},\ \tfrac{5}{4},-\tfrac{5}{4},-\tfrac{5}{4}\right)} & \boxed{\left(-\tfrac{3}{4},-\tfrac{5}{4},-\tfrac{1}{4},-\tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{5}{4},\ \tfrac{9}{4},\ \tfrac{3}{4}\right)} \\ \boxed{W_4^{(1)} = W_3^{(1)}} & \boxed{W_4^{(2)} = W_3^{(2)}} \\ \boxed{\left(\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0\right)} & \boxed{\left(\ \tfrac{3}{2},-\tfrac{1}{2},-1,\ 2,\ 0,-\tfrac{3}{2},\ 1,\ \tfrac{1}{2}\right)} \\ \boxed{W_6^{(1)} = W_3^{(1)}} & \boxed{W_6^{(2)} = W_3^{(2)}} \end{pmatrix} \tag{7.4}$$

with the orbifold specific geometric conditions (eq. (2.25)) $W_3 = W_4 = W_6$ and $W_1 = W_2 = (0^{16})$. Furthermore, the 4D gauge group reads

$$G_{4D} = \mathrm{SU}(3)_{\mathrm{flavor}} \times \mathrm{SU}(3)_{\mathrm{C}} \times \mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{SU}(5)_{\mathrm{hidden}} \times \mathrm{U}(1)_{\mathrm{Y}} \times \mathrm{U}(1)^6 \ , \tag{7.5}$$

with a gauged SU(3) flavor symmetry $\mathrm{SU}(3)_{\mathrm{flavor}}$. Concretely, the reason for this gauged flavor symmetry is that from the massless matter spectrum given in table 7.2 we observe that the quark doublets (q), up-type quarks ($\bar{u}$), electrons ($\bar{e}$) and up-type Higgses (h_u) transform as triplets under the $\mathrm{SU}(3)_{\mathrm{flavor}}$. These fields live all in the bulk of the orbifold. Moreover, one down-type quark ($\bar{d}$) is a singlet of $\mathrm{SU}(3)_{\mathrm{flavor}}$ and lives in the U_3-sector. Finally, the $T_{(1,3)}$ twisted sector provides two down-type quarks ($\bar{d}$) and lepton doublets (ℓ).

Moreover, with respect to the MSSM this model contains exactly three additional Higgs-pairs (h_u, h_d) and ten flavons f^0 that transform as triplets of $\mathrm{SU}(3)_{\mathrm{flavor}}$. These flavons contribute $10 \times 3 = 30$ SM singlets to the in total 75 SM singlets (s^0). Their VEVs are important in this model to explain quark and lepton masses using the $\mathrm{SU}(3)_{\mathrm{flavor}}$ as family symmetry [169, 170].

To estimate the hidden sector gaugino condensation scale, we have to analyze the chiral spectrum with respect to the hidden sector gauge group factor $\mathrm{SU}(5)_{\mathrm{hidden}}$. It turns out that the massless spectrum contains two **5**-plets and two $\bar{\mathbf{5}}$-plets, that are SM singlets s^0. Thus, these SM singlets can decouple without breaking the SM gauge group or $\mathrm{SU}(5)_{\mathrm{hidden}}$. Hence, there is no remaining matter charged under $\mathrm{SU}(5)_{\mathrm{hidden}}$ that can contribute to the beta-function coefficient. Thus, with $C_2 = 5$ we obtain $b = 3 \cdot 5 = 15$, which relates to the phenomenologically interesting condensation scale $\Lambda_{\mathrm{hidden}} \approx 10^{12}$ GeV eq. (7.1).

The entire set of 20 almost perfect MSSM-like orbifold models is available as arXiv ancillary files [20]. They originate all from the $\mathbb{Z}_2 \times \mathbb{Z}_4$ point group. In detail, one model was found in the (1,6) orbifold geometry and 19 in the (2,4) orbifold geometry. These 20 almost perfect MSSM-like orbifold models are very similar in their massless matter spectrum. To get an overview concerning their hidden sector gaugino condensation scale, we marked these 20 almost perfect

---

[7] Note, for this geometry, the two rotational generators are roto-translations.

sector	#	irrep	labels
U_1	1	$(\mathbf{3};\mathbf{3},\mathbf{2};\mathbf{1})_{\frac{1}{6}}$	q_i
	1	$(\mathbf{1};\mathbf{1},\mathbf{2};\mathbf{1})_{-\frac{1}{2}}$	ℓ_i or h_d
	2	$(\mathbf{1};\mathbf{1},\mathbf{1};\mathbf{1})_0$	s_i^0
U_2	1	$(\bar{\mathbf{3}};\mathbf{1},\mathbf{2};\mathbf{1})_{-\frac{1}{2}}$	ℓ_i or h_d
	1	$(\mathbf{3};\mathbf{1},\mathbf{2};\mathbf{1})_{\frac{1}{2}}$	h_u
U_3	1	$(\mathbf{3};\bar{\mathbf{3}},\mathbf{1};\mathbf{1})_{-\frac{2}{3}}$	$\bar{u}_i$
	1	$(\mathbf{1};\bar{\mathbf{3}},\mathbf{1};\mathbf{1})_{\frac{1}{3}}$	$\bar{d}_i$
	1	$(\mathbf{3};\mathbf{1},\mathbf{1};\mathbf{1})_1$	$\bar{e}_i$
	2	$(\mathbf{1};\mathbf{1},\mathbf{1};\bar{\mathbf{5}})_0$	s_i^0
	1	$(\mathbf{1};\mathbf{1},\mathbf{1};\mathbf{1})_0$	s_i^0
$T_{(1,3)}$	22	$(\mathbf{1};\mathbf{1},\mathbf{1};\mathbf{1})_0$	s_i^0
	2	$(\mathbf{1};\bar{\mathbf{3}},\mathbf{1};\mathbf{1})_{\frac{1}{3}}$	$\bar{d}_i$
	2	$(\mathbf{1};\mathbf{1},\mathbf{2};\mathbf{1})_{-\frac{1}{2}}$	ℓ_i or h_d
	10	$(\bar{\mathbf{3}};\mathbf{1},\mathbf{1};\mathbf{1})_0$	f_i^0
	2	$(\mathbf{1};\mathbf{1},\mathbf{1};\mathbf{5})_0$	s_i^0

Table 7.2: *Massless matter spectrum of an "almost perfect" MSSM-like orbifold model originating from the* $\mathbb{Z}_2 \times \mathbb{Z}_4$ $(2,4)$ *orbifold geometry. The third column gives the irreducible representations with respect to the gauge group factors in the order* $G_{4D} = \mathrm{SU}(3)_{\text{flavor}} \times \mathrm{SU}(3)_{\mathrm{C}} \times \mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{SU}(5)_{\text{hidden}} \times \mathrm{U}(1)_{\mathrm{Y}} \times \mathrm{U}(1)^6$.

MSSM-like orbifold models as yellow points in fig. 7.5b. For further details on these models see [20].

However, a detailed phenomenological investigation of these models is not possible. Unfortunately, those orbifold geometries correspond to the two exceptions for the $\mathbb{Z}_2 \times \mathbb{Z}_4$ point group for which R-symmetries are not understood [158, 171–173]. Hence, to analyze the almost perfect MSSM-like orbifold models, the R-symmetries have to be re-analyzed for these orbifolds.

Finally note that, even though the $\mathbb{Z}_2 \times \mathbb{Z}_4$ point group is one of the majority classes, i.e. 48 812 MSSM-like orbifold models, the specific geometries $\mathbb{Z}_2 \times \mathbb{Z}_4(1,6)$ and $\mathbb{Z}_2 \times \mathbb{Z}_4(2,4)$ of this point group lead to only 82 and 320 MSSM-like orbifold models, respectively. Thus, further investigation on the impact of the specific lattice for the same point group might reveal why these minority classes achieve almost perfect MSSM-like models.

7.5 Chapter summary

In conclusion, the trained predictive model shows that the heterotic orbifold landscape gets partitioned into several sub-landscapes that strongly correlate to phenomenological properties. Concretely, the point group P that underlies an orbifold geometry leaves a particular imprint on the particle spectrum of vector-like exotics and some additional phenomenological features. This imprint can be exploited to assign likelihoods to the point group's capability of creating a certain MSSM-like particle spectrum. Furthermore, we present the first "almost perfect" MSSM-like orbifold models, see table 7.2. They originate from the $\mathbb{Z}_2 \times \mathbb{Z}_4$ $(1,6)$ and $(2,4)$ orbifold geometries. These findings may encourage further research to complete the understanding of R-symmetries for these two $\mathbb{Z}_2 \times \mathbb{Z}_4$ orbifold geometries [158, 171–173].

Moreover, fig. 7.3 implies, that it is possible to utilize the hidden sector gaugino condensation scale to formulate a refined definition of MSSM-like models. In detail, the dataset D of inequivalent MSSM-like spectra can be reduced by $\approx 50\%$ if models with $\Lambda_{\text{hidden}} \lesssim 10^5$ GeV are excluded. In the upcoming chapter 8, we show how this refined definition of MSSM-like models can be used for contrast mining patterns, from section 6.2, to further reduce the landscape.

Chapter 8

Interplay of machine learning results

In this section, we are revealing the interplay among the machine learning approaches and illustrate how synergy effects can be used for further improvements. Concretely, the effects of the different machine learning approaches are studied among each other, e.g.: Where lie the new contrast mining MSSM-like models from table 6.6 in the 2-dimensional autoencoder clustering of chapter 5 ? How does the phenomenological property of hidden sector gaugino condensation from chapter 7 correlate with the contrast mining pattern on the hidden E_8 eq. (6.24)? Each of the machine learning approaches individually is able to produce insights into the heterotic orbifold landscape. However, it is important to examine how these concepts interact with each other. Note, for clarity, we will refer to the large-scale random scan of type fig. 4.1, observed in chapter 5, as $\texttt{C}^2\texttt{PAP}$ random search, since it was generated by runs on the `Computational Center for Particle and Astrophysics` compute cluster.

Therefore, let us start by examining the behavior of the special MSSM-like models generated in the contrast mining procedure. First, regarding the prediction of the decision tree $\texttt{DT}_{\texttt{Cluster}}$ trained to decide on which island the MSSM-like models are located, see eq. (5.14). The predictions are summarized in table 8.1. Although the models of the contrast mining techniques originate from areas in the landscape that are very different from the $\texttt{C}^2\texttt{PAP}$ random search (see fig. 6.6), nevertheless, $2/3$ of the contrast mining MSSM-like models are assigned to the fertile islands. This is the same ratio as obtained for the $\texttt{C}^2\texttt{PAP}$ and Mini-Landscape datasets from chapter 5. Hence, this confirms the reliable predictive power of the autoencoder clustering that extends beyond the $\texttt{C}^2\texttt{PAP}$ random search.

Consequently, the question arises: Are those patterns different from the contrast mining patterns (particularly the phenomenological patterns eq. (6.13)), or do the patterns of the two methods correlate? Hence, it is crucial how the complete contrast mining dataset of $\mathbb{Z}_6$-II (1,1) interacts with the clustering, i.e., how the ~~MSSM~~-like models distribute in the latent space. The position of the contrast mining models from table 6.2, projected by $\texttt{AE}_{\text{encoder}}$ eq. (5.11) into the 2-dimensional chart[1] of the heterotic orbifold landscape, is shown in fig. 8.1. Although these models already obey the SM gauge group in the first E_8 and have a hidden gauge group bigger than $N_{\text{ur}}^{(2)} \geq 6$, they still populate all the clusters from the autoencoder approach. This implies that the autoencoder was not able to recognize these underlying conditions in its clustering. The result indicates that the two methods are orthogonal to each other, and therefore one method can improve on the other since they found different patterns that do not correlate. Moreover, the MSSM-like models from contrast mining populate the fertile islands of the landscape in similar proportion to other datasets, e.g. compare with table 8.1. Note the autoencoder latent space

---

[1] The ability to project new instances into the same 2-dimensional space is not given for another powerful 2-dimensional projection technique named t-SNE [190].

	region	Contrast Mining	$\texttt{C}^2\texttt{PAP}$
	R_0	156	65
fertile islands	R_1	63	48
	R_2	63	21
	R_3	47	11
	R_4	35	18
	R_5	19	6
	R_6	20	3
	R_7	2	2
	R_8	2	2
	R_9	4	1
	R_{10}	36	12
	R_{11}	34	6
	total	481	195

Table 8.1: *Classification of 481 contrast mining MSSM-like* $\mathbb{Z}_6$-II (1,1) *models within the various regions* R_i *of the landscape, as predicted by the decision tree* $\texttt{DT}_{\texttt{Cluster}}$ *eq.* (5.14). *For comparison the 195 MSSM-like models from the* $\texttt{C}^2\texttt{PAP}$ *random search of chapter 5 are given.*

can be used as a filter in the contrast mining approach. This filtering is more effective than in the $\texttt{C}^2\texttt{PAP}$ random search, as the contrast mining approach forces the algorithm to produce models with involved particle content. The computational time on the physically more intricate models is significantly increased for the `orbifolder`, compared to the models excluded from the contrast mining patterns. Hence, the fast decision if the model lies in the fertile island can save computational time if applied after the sophisticated search algorithm fig. 6.3.

Furthermore, merging the 195 MSSM-like models from the $\texttt{C}^2\texttt{PAP}$ random search in chapter 5 with the 481 MSSM-like models of the contrast mining search, chapter 6, results in 517 inequivalent models. This is an unexpected increase in inequivalent MSSM-like models given table 6.6, from which it appeared that new models are complicated to find. Moreover, other MSSM-like models from the Mini-Landscape do not contribute further models. In this context, it is crucial that the 195 MSSM-like models likewise extend the old scans in the literature [126, 127] by 89 models. Hence, this demonstrates that the guided searches with $W_5 = (0^{16})$ and the Mini-Landscape are different from the $\texttt{C}^2\texttt{PAP}$ random search and make up the majority of MSSM-like models known in the literature. Even though the $\texttt{C}^2\texttt{PAP}$ random search was intense, it only led to 195 inequivalent MSSM-like models. This scan corresponds to the phenomenologically unconstrained search, i.e. 1 out of 10^7 modular invariant gauge embedding matrices M results in an equivalent MSSM-like model (see the last paragraph in section 6.1.2). This implies, that in order to find 195 inequivalent MSSM-like models it is reasonable to assume that $\mathcal{O}(10^3)$ equivalent MSSM-like models were produced (compare to table 6.2), thus we can estimate that $\mathcal{O}(10^{10})$ modular invariant $\mathbb{Z}^{128}$ points where tested in the $\texttt{C}^2\texttt{PAP}$ random search.

Astonishingly, all new models from the $\texttt{C}^2\texttt{PAP}$ random search of the $\mathbb{Z}_6$-II (1,1) landscape obey the contrast mining conditions, in particular, they have the properties $N_{\mathrm{ur}}^{(2)} \geq 18$ and fulfill the U-sector condition. This observation is a great success for the generalization of the contrast mining patterns. No MSSM-like model of $\mathbb{Z}_6$-II (1,1) scans in the literature, and even the large-scale $\texttt{C}^2\texttt{PAP}$ random search was able to violate the contrast patterns.[2] Note that the

---

[2] Only the special $\Delta(54)$ models with $W_3 = (0^{16})$ do not obey all U-sector conditions, however, they were found by the contrast mining search itself, see section 6.2.5.

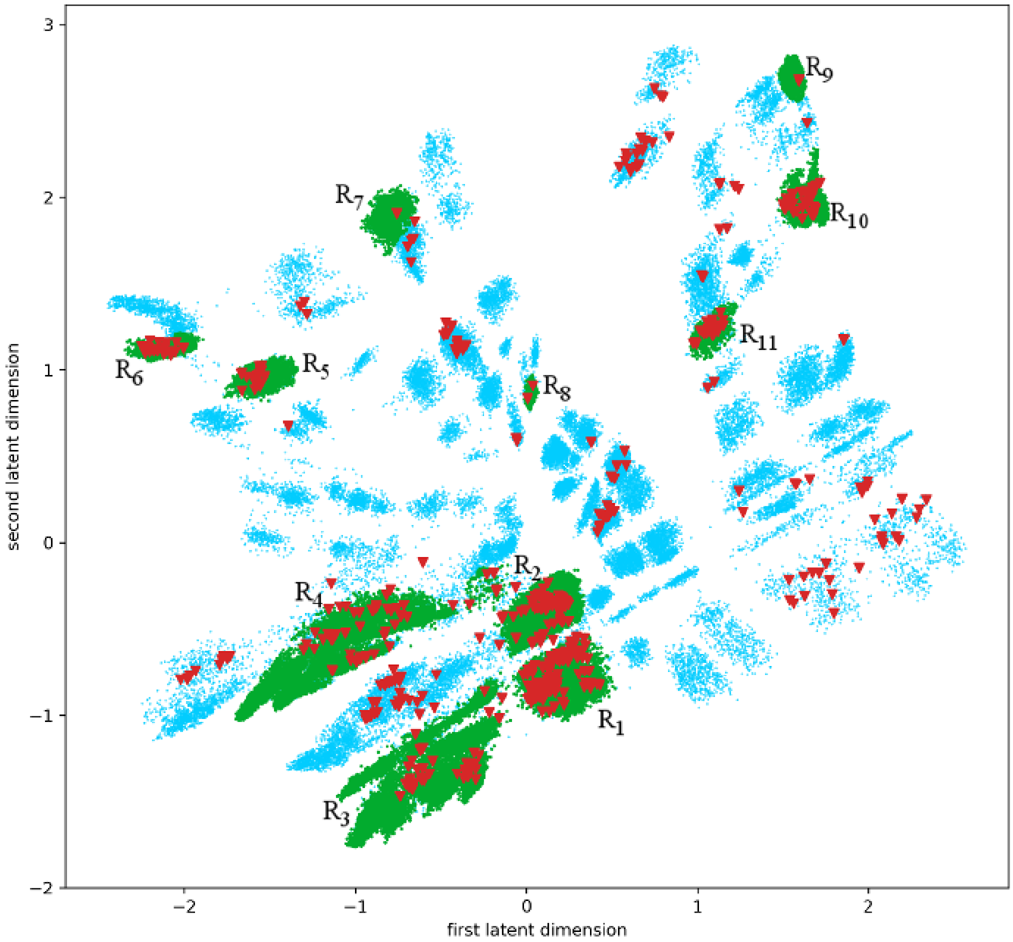


Figure 8.1: *Projection* $\mathtt{AE}_{\text{encoder}}$ *eq.* (5.11) *of the models constructed by the contrast mining successive search (fig. 6.3) with the* hidden E_8 *eq.* (6.24) *and SM gauge group eq.* (6.13) *constraint. It is fascinating that still all areas, in particular all blue areas, are populated by this very specific dataset. The red triangles visualize the combined set of the 481 MSSM-like models from contrast mining and the 195 MSSM-like models from the* `C`^{2}`PAP` *random search. In total, there are 517 inequivalent MSSM-like models.*

`C`^{2}`PAP` random search, as discussed in chapter 6, scans especially in the area excluded by the hidden E_8 constraint, i.e. $N_{\text{ur}}^{(2)} < 6$. Nonetheless, it was impossible to find a model that violates this condition. Identical arguments hold for the U-sector constraints.

Let us stress the importance of the contrast mining constraints with additional observations on the models that can host G'_{SM} in both E_8 factors. So far, those models were excluded from the analysis as it can not be decided during the construction process, which E_8 becomes the hidden E_8 sector. However, they do not violate the constraints either. It is always possible to choose one of both E_8 factors (there exists the $\mathbb{Z}_2$ of both E_8 factors section 4.2) as the hidden sector such that all the contrast mining patterns are satisfied. In other words, not a single model of this type violates the conditions for the U-sector.[3] For some models, both variants satisfy the contrast conditions, i.e. it is irrelevant which E_8 factor is assigned to be the hidden sector. These are significant indications that the contrast mining patterns are fundamentally connected

---

[3] Recall from section 6.2.5 that the hidden E_8 lower bound is trivially fulfilled.

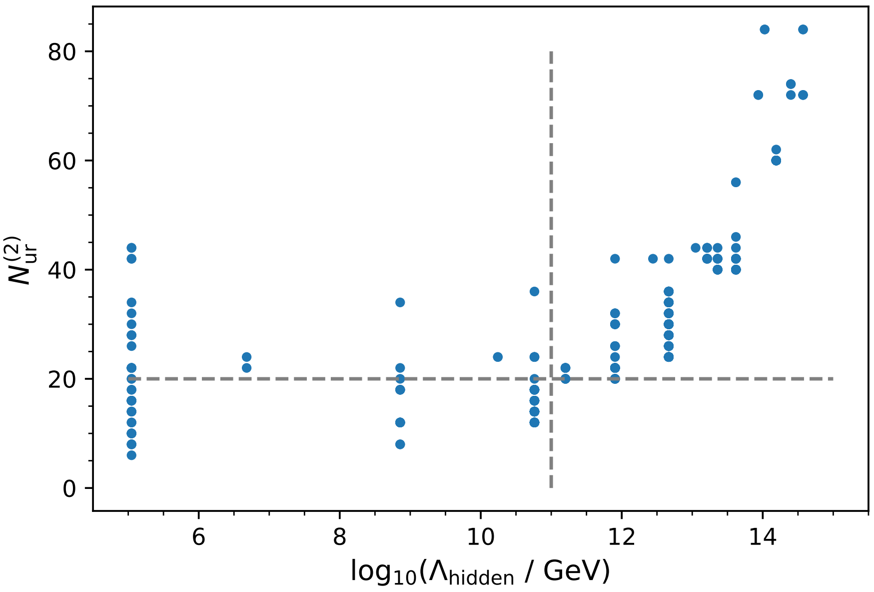


Figure 8.2: *Contrast mining pattern* dynamic hidden E_8 *eq.* (6.31) *correlates strongly with the gaugino condensation scale. The vertical dashed line is at* $\ln(\Lambda_{\mathrm{hidden}}/\mathrm{GeV}) = 11$ *and the horizontal line at* $N_{\mathrm{ur}}^{(2)} = 20$*. We see with this new phenomenological ingredient the observed contrast patterns gain importance, and the lower bound on the* hidden E_8 *could be raised.*

to the property of a model to be MSSM-like.

Finally, it is interesting that the efforts from predicting the orbifold origin of the exact MSSM (chapter 7) link to the contrast mining constraints. In particular, we want to analyze the connection of the hidden sector gaugino condensation scale to the contrast patterns. A direct investigation of the hidden sector gaugino condensation is not possible during the successive creation of the gauge embedding. The resulting hidden gauge group and the matter content depend non-linearly on the subsequent Wilson lines. Furthermore, the condensation scale is connected to the decoupling assumption of the vector-like exotics, see section 7.1. Interestingly, the observed pattern for the hidden E_8 unbroken roots correlates phenomenologically beneficial with the hidden sector gaugino condensation scale, see fig. 8.2. Concretely, higher values X for the hidden E_8 constraints generically increase the gaugino condensation scale and hence make the models phenomenologically more viable for supersymmetry breaking at the TeV scale. Moreover, it appears from fig. 8.2 that the contrast pattern, here in the case of $\mathbb{Z}_6$-II (1,1), can improve to reduce the landscape further by taking this supplementary phenomenological property into account. Recall the estimated growth rate $\mathrm{gr}\left(N_{\mathrm{ur}}^{(2)} \geq 20\right) = 12$ from table 6.4, which indicates that for the threshold in fig. 8.2 the reduction of the landscape would be significant. Furthermore, remember from fig. 7.2 that this effect is even stronger for the whole MSSM-like orbifold landscape as other geometries produce even more models in the region of small supersymmetry breaking scales. This connection can lead to an immediate impact of the gaugino hidden sector condensate on the scans, even though it can not be tested directly.

To summarize, the above results suggest that we uncovered some building blocks that can be iteratively applied to reinforce each other and lead to a virtuous circle of further insights into the heterotic orbifold landscape.

Chapter 9

Conclusion and outlook

In this thesis, we have studied several phenomenological aspects of orbifold compactifications in heterotic string theory. After a concise introduction to the string theory framework in chapter 2, the starting point in chapter 3 is an analysis of a perturbatively vanishing cosmological constant Λ at the one-loop level, in the context of non-supersymmetric heterotic orbifolds. The one-loop vacuum amplitude that computes Λ can be factorized according to the (twisted) sectors of the orbifold geometry. We show that for a model-independent vanishing, it is sufficient that each of these sectors preserves a local Killing spinor. Investigating this option, this results in a non-existence proof that for non-supersymmetric heterotic orbifolds, there exists always at least one twisted sector without a local Killing spinor. Moreover, we connect this to a more general group theoretical no-go conjecture that we verify for $\mathcal{O}(100\,000)$ finite groups. This observation emphasizes how important further considerations of a general formalism for non-Abelian orbifold point groups are. Furthermore, the study of the model details, in non-supersymmetric theories, seems to be unavoidable in tackling the issue of the cosmological constant in this setting. For both concerns, our analysis provides the basis for such investigations.

In light of the obstacles associated with the cosmological constant for non-supersymmetric theories, we study the structure of the supersymmetric heterotic orbifold landscape with Abelian point groups. The ultimate goal is to confirm the importance of string theory by explicitly constructing a string model that reproduces the spectrum of the MSSM in the low energy limit. Hence, the searches focus on MSSM-like models that allow for vector-like exotics beside the exact MSSM spectrum. In chapter 4, we introduce a statistical procedure to solve the intricate modular invariance conditions for the gauge embedding compactification parameters. Here we also explore the parameter space neighborhood of an MSSM-like model in the $\mathbb{Z}_6$-II (1,1) landscape. This makes clear how disordered the landscape appears as the local neighborhood consists of entirely different physical models as well as numerous string-inconsistent models. We develop a statistical method to explicitly compute the extraordinary low probability to find a set of modular invariant compactification parameters and thus verify the necessity for a guiding principle when exploring the landscape.

Therefore, we pioneer in this thesis in applying the new and powerful techniques of machine learning to the string landscape. In chapter 5, we construct a representation of the $\mathbb{Z}_6$-II (1, 1) orbifold compactification parameters, that is invariant under the symmetries of the landscape. This enables us to train an autoencoder neural network to project the $\mathcal{O}(100)$ compactification parameters highly non-linearly to a 2-dimensional representation. The resulting chart reveals the clustering of MSSM-like models, even though the neural network had no access to the related physics of the parameters. In this representation, the physically realistic models are traceable for the first time. Concretely, the MSSM-like models accumulate and populate specific regions, so-called fertile islands. In those islands, the proportion of MSSM-like models is approximately 20 times higher than in the rest of the projected landscape. The significant dimensional reduction of the parameter space in combination with a physically meaningful clustering illustrates the high informational content of the 2-dimensional chart. Hence, the search technique based on the uncovered fertile islands provides a prime example for a systematic search in the landscape.

Inspired by the vast area in the autoencoder clustering that does not contain realistic 4D models, we work out in chapter 6 an advanced search strategy that directs the construction process towards areas that can produce phenomenologically viable models. Therefore, we analytically reduce the search space by developing a method to divide out the Weyl group (with $\mathcal{O}(10^{17})$ elements) of the landscape. Furthermore, the search is informed about the necessity to reproduce the SM gauge group, which excludes 90% of the $\mathbb{Z}_6$-II (1,1) orbifold parameter space. Proceeding to these substantial reductions of the landscape, a data-driven approach, so-called contrast mining, is utilized. This statistical analysis yields constraints with clear physical interpretation. Namely, lower bounds on the number of unbroken roots in the hidden E_8 and the number of bulk matter fields emerge. Imposing these bounds raises the probability of obtaining MSSM-like models tremendously. Moreover, we develop a systematic search that reaches corners of the parameter space that hold many inequivalent MSSM-like models but are hardly accessible by the normal random search. Finally, since the improved search for MSSM-like models is not only superior by orders of magnitude but also easily generalizable to all Abelian orbifold geometries, a considerably more comprehensive dataset of MSSM-like models is obtained. This leads to the first practical results of machine learning in string theory. Concretely, the search solves long-standing issues and can find special models, such as those with $\Delta(54)$ flavor symmetry ($W_3 = W_4 = (0^{16})$) in the $\mathbb{Z}_6$-II (1,1) orbifold geometry or MSSM-like models in $\mathbb{Z}_7$ (1, 1) geometry, that previous random searches could not discover. This substantially improved datasets of MSSM-like models are published as ancillary files of [20, 21].

The generalization of the contrast patterns showed that the specific numerical threshold values on the number of unbroken roots in the hidden E_8 and the bulk matter fields correlate with the point group P that underlies an orbifold geometry. Hence, in chapter 7, we investigate the phenomenology of MSSM-like models and discover that a boosted decision tree is suitable to infer the point group P the MSSM-like model originates from. In detail, phenomenological properties like the various types of vector-like exotics (see table D.1 and table D.2), the number of SM singlets, the existence of non-Abelian flavor symmetries, and the hidden sector beta function relevant for supersymmetry breaking via gaugino condensation are considered. This observation is an important step that indicates that the specific orbifold compactification leaves an imprint on the particle spectrum and divides the orbifold landscape into phenomenological sub-regions. Consequently, it is also possible to predict the most likely origin of perfect MSSM models, i.e. models with additional SM singlets but no vector-like particles in the spectrum. This prediction is performed over various values for the number of SM singlets and the hidden sector beta function. Almost the whole parameter space is occupied by $\mathbb{Z}_2 \times \mathbb{Z}_4$, $\mathbb{Z}_4 \times \mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ and some rare exceptions of other point groups. In this regard, we find the first almost perfect models from heterotic orbifolds, that solely have Higgs-like fields as additional vector-like particles from orbifolds with $\mathbb{Z}_2 \times \mathbb{Z}_4$ point group. Hence, these astonishing results provide a natural indicator to focus future searches on the orbifold geometries related to the predicted point groups.

The individual machine learning results mean enormous progress in understanding the landscape. However, there is even more in the interesting interplay between the different approaches. In chapter 8, we point out that the new type of MSSM-like models from the advanced contrast mining search still accumulate in the fertile islands of the autoencoder clustering. This strengthens the generalization property of the islands that operate on entirely new models. Furthermore, the contrast mining patterns of chapter 6 are orthogonal to the patterns that lead to the clustering of the MSSM-like models in chapter 5. Concretely, the autoencoder projects a huge amount of not MSSM-like models, from the advanced search to the region outside of the fertile islands. Thus, these methods can improve on each other as their provided information is not correlated, and both identify complementary regions of the landscape that do not lead to MSSM-like models. Moreover, we increase in chapter 8 the statistical evidence for contrast mining patterns to have an analytical connection to the MSSM-like models. The large-scale search of the $\mathbb{Z}_6$-II (1,1) landscape from chapter 5 scans mainly in the region of possible counterexamples. Even though this large-scale scan finds several novel MSSM-like models, they all confirm the contrast mining conditions. Further developments of contrast patterns in combination with investigations to find their analytical source is a fascinating research direction.

In addition, we show in chapter 8 that the hidden sector gaugino condensation scale of MSSM-like models correlates with the number of unbroken roots of the hidden E_8, such that the constraint from chapter 6 can be improved. This may spark a renewed interest in the detailed computation of the hidden sector gaugino condensation scale, i.e. the threshold corrections, as this feature is crucial to predict the orbifold origin of the perfect MSSM in chapter 7. The precise supersymmetry breaking scale can now directly be translated to contrast mining constraints on the hidden sector. Thus, this phenomenological property gets incorporated in further reduction of the landscape and a refined search for MSSM-like models.

Besides, the methods presented in this thesis are not limited to the heterotic orbifold landscape. This work already inspired the first attempts in this direction. In particular, the autoencoder clustering of chapter 5 is applied to other compactification schemes based on Calabi-Yau threefolds with line bundles [174, 175]. Similar to our findings, there exist clusters of phenomenologically realistic models in this schemes. Thus, the developed techniques are far more general and can be used in different compactification attempts. To generalize the prediction of the perfect MSSM origin, from chapter 7, across compactification techniques, seems an important step to classify compactification methods and to focus the search effort in the future.

Constraints to reduce the landscape and the construction mechanism fig. 6.3 reshape the probability mass function of the compactification parameters to approximate the distribution for MSSM-like models. A straightforward implication would be to use the large set of MSSM-like models, gathered in chapter 6, to learn the probability distribution of the compactification parameters for MSSM-like models. In the context of neural networks, several approaches to estimate joint probabilities of utmost complexity exist. Those models correspond to the type of generative models, e.g. variational autoencoders (VAE) [204], generative adversarial neural networks (GANs) [205, 206] and autoregressive models, like PixelCNN [185, 207]. Primarily GANs have been used in string theory related investigations [176, 177]. Initial tests suggest that an autoregressive approach is conceivable for the heterotic orbifold landscape. These models would extend the procedure fig. 6.3 and predict a distribution for every single compactification parameter based on all previously drawn parameters. Thus, further exploration to find the distribution of compactification parameters that relate to MSSM-like models seems promising.

During the investigations of this thesis it was observed that neural networks have difficulties to approximate the function of the `orbifolder`. Indeed, the modulo operations of orbifold projections are challenging to model for neural networks and need very high capacity, i.e. many adjustable parameters. Hence, it seems beneficial in physical applications to implement the system knowledge in the neural network architecture. Learnable activation functions [208] based on periodic functions, e.g. Fourier expansions, could be useful to allow the network to learn the modulo operation. A first attempt to improve the extrapolation behavior on Calabi-Yau threefold hypersurfaces in toric varieties [178] used equation learners [209], which utilize a set of base functions, inspired from natural science, as activation functions. This could be a joint improvement for machine learning and physics to include more concrete functional dependencies into the neural network to make it more efficient and precise.

An essential parallel aspect of the results of this thesis is that the search for MSSM-like models is vastly improved for inequivalent and most importantly, equivalent MSSM-like particle spectra in chapter 6. As we reduce the amount of symmetry that can lead to exactly equivalent physical models, these models may still vary in additional model parameters. Hence, there is a necessary demand to compute more details of the models, e.g. the Yukawa couplings and the decoupling scenarios of the vector-like exotics towards the MSSM [77]. The intriguing results for almost perfect MSSM models and the new insight on the phenomenological imprint of an orbifold geometry, seem to be an exciting exploration task to close the gap and find the perfect MSSM in string theory.

Appendix A

Representation theory of SO(6), Spin(6) and SU(4)

A.1 SO(6)

For the so(6) Lie algebra the generators J_{ij} in the vector representation can be given by,

$$(J_{ij})_{kl} \;=\; \delta_{ik}\delta_{jl} - \delta_{jk}\delta_{il} \;, \tag{A.1}$$

for $i, j = 1, \ldots, 6$. $J_{ij} = -J_{ji}$ for $i < j$ and $i, j = 1, \ldots, 6$ and form a basis of anti-symmetric 6×6 matrices. J_{ij} generates a rotation in the (X^i, X^j)-plane. The exponential map from the Lie algebra so(6) to the Lie group SO(6) for a specific element θ is specified by the anti-symmetric parameters ω_{ij} as

$$\theta \;=\; \exp\left(\tfrac{1}{2}\,\omega_{ij}\, J_{ij}\right) \;. \tag{A.2}$$

A.2 Spin(6)

The Euclidean Clifford algebra in six-dimensions is generated by 8×8 Hermitian gamma matrices Γ_i, $i = 1, \ldots, 6$ [179],

$$\left\{\Gamma_i, \Gamma_j\right\} \;=\; 2\,\delta_{ij}\,\mathbb{1}_8 \;, \qquad \tilde{\Gamma}^2 \;=\; \mathbb{1}_8 \;, \qquad \left\{\Gamma_i, \tilde{\Gamma}\right\} \;=\; 0 \;. \tag{A.3}$$

where the chirality operator $\tilde{\Gamma} = \mathrm{i}\,\Gamma_1\,\Gamma_2 \ldots \Gamma_6$. With the chirality operator we can define the Weyl-spinors using the projectors,

$$P^{(\pm)} \;=\; \frac{\mathbb{1}_8 \pm \tilde{\Gamma}}{2} \quad \text{with} \quad \left(P^{(\pm)}\right)^2 \;=\; P^{(\pm)} \quad \text{and} \quad P^{(+)}\,P^{(-)} \;=\; 0 \;. \tag{A.4}$$

The charge conjugation matrix C has the properties

$$C\,\Gamma_i\,C^{-1} \;=\; \Gamma_i^T \;, \qquad C\,\tilde{\Gamma}\,C^{-1} \;=\; -\,\tilde{\Gamma}^T \;, \qquad C^\dagger \;=\; -\,C^T \;=\; C \;. \tag{A.5}$$

It follows that, $C\,P^{(\pm)}\,C^{-1} = (P^{(\mp)})^T$ and $(C\Gamma_i)^T = -C\Gamma_i$ are anti-symmetric matrices. With the anti-Hermitian spin generators Σ_{ij},

$$\Sigma_{ij} \;=\; -\,\Sigma_{ji} \;=\; \tfrac{1}{4}\left[\Gamma_i, \Gamma_j\right] \;, \qquad C\,\Sigma_{ij}\,C^{-1} \;=\; -\,\Sigma_{ij}^T \;, \quad \text{and} \quad \mathrm{Tr}(\Sigma_{ij}) = 0 \;, \tag{A.6}$$

we can define the spin representation as

$$D_{\mathbf{s}}(\theta) \;=\; \exp\left(\tfrac{1}{2}\,\omega_{ij}\,\Sigma_{ij}\right) \quad\text{with}\quad C\,D_{\mathbf{s}}(\theta)\,C^{-1} \;=\; D_{\mathbf{s}}(\theta^{-1})^T = D_{\mathbf{s}}(\theta)^* \;, \tag{A.7}$$

The generators Σ_{ij} fulfill the so(6) Lie algebra that can be defined from the generators J_{ij} of the vector representation of so(6) given in eq. (A.1). Furthermore, $D_{\mathbf{s}}(\theta)$ is a unitary matrix, since $D_{\mathbf{s}}(\theta)^\dagger = D_{\mathbf{s}}(\theta^{-1}) = D_{\mathbf{s}}(\theta)^{-1}$.

Thus, the vector and spinor representation are related by

$$D_{\mathbf{s}}(\theta)^T\,C\Gamma_i\,D_{\mathbf{s}}(\theta) \;=\; \theta_{ij}\,C\Gamma_j \;. \tag{A.8}$$

The quadratic form of the left hand side shows that Spin(6) is the double cover of SO(6), as $-D_{\mathbf{s}}(\theta)$ and $D_{\mathbf{s}}(\theta)$ result in the same SO(6) group element θ. This means that the generators θ_α of the geometrical point group $P \subset \mathrm{SO}(6)$, can be connected in two different ways to the corresponding Spin(6) generators: $D_{\mathbf{s}}(\theta_\alpha)$ or $-D_{\mathbf{s}}(\theta_\alpha)$. Contrary, $D_{\mathbf{s}}(\theta)$ determines θ uniquely, i.e.

$$\theta_{ij} \;=\; \tfrac{1}{8}\,\mathrm{Tr}\big[D_{\mathbf{s}}(\theta^{-1})\,\Gamma_i\,D_{\mathbf{s}}(\theta)\,\Gamma_j\big] \;, \tag{A.9}$$

using eq. (A.8) and eq. (A.7).

A.3 SU(4)

For a convenient chiral basis we give the six-dimensional Clifford algebra in terms of the five-dimensional Clifford algebra. Concretely, we observe for the gamma matrices Γ_i, chirality operator $\tilde{\Gamma}$ and charge conjugation C that,

$$\Gamma_i \;=\; \begin{pmatrix} 0 & \bar{\gamma}_i \\ \gamma_i & 0 \end{pmatrix} \;,\quad \Sigma_{ij} \;=\; \begin{pmatrix} \sigma_{ij} & 0 \\ 0 & \bar{\sigma}_{ij} \end{pmatrix} \;,\quad \tilde{\Gamma} \;=\; \begin{pmatrix} \mathbb{1}_4 & 0 \\ 0 & -\mathbb{1}_4 \end{pmatrix} \;,\quad C \;=\; \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix} \;, \tag{A.10}$$

where $\bar{\gamma}_\kappa \;=\; \gamma_\kappa$, $\kappa \;=\; 1,\ldots,5$, and charge conjugate matrix c, satisfying $c\,\gamma_\kappa\,c^{-1} \;=\; \gamma_\kappa^T$ and $c^\dagger \;=\; c^{-1} \;=\; -c^T$ in five dimensions and $-\bar{\gamma}_6 \;=\; \gamma_6 \;=\; \mathrm{i}\,\mathbb{1}_4$. Similarly, we express the spin generators Σ_{ij} eq. (A.6) in terms of anti-Hermitian su(4) generators,

$$\sigma_{ij} \;=\; \tfrac{1}{4}(\bar{\gamma}_i\gamma_j - \bar{\gamma}_j\gamma_i) \quad\text{and}\quad \bar{\sigma}_{ij} \;=\; \tfrac{1}{4}(\gamma_i\bar{\gamma}_j - \gamma_j\bar{\gamma}_i) \;, \tag{A.11}$$

that are all traceless. The two sets of 15 generators related to each other by $\bar{\sigma}_{ij} = -c^{-1}\,\sigma_{ij}^T\,c$.

This allows us to use the the projectors eq. (A.4) to define the two irreducible chiral representations $D_{\mathbf{s}}^{(\pm)}(\theta) = P^{(\pm)}D_{\mathbf{s}}(\theta)$, from the reducible spin representation eq. (A.7). Those can be identified with the SU(4)-matrices

$$D_{\mathbf{4}}(\theta) \;=\; \exp\left(\tfrac{1}{2}\,\omega_{ij}\,\sigma_{ij}\right) \qquad D_{\bar{\mathbf{4}}}(\theta) \;=\; \exp\left(\tfrac{1}{2}\,\omega_{ij}\,\bar{\sigma}_{ij}\right) \;, \tag{A.12}$$

generated by the matrices σ_{ij} and $\bar{\sigma}_{ij}$, as

$$D_{\mathbf{s}}^{(+)}(\theta) \;=\; \begin{pmatrix} D_{\mathbf{4}}(\theta) & 0 \\ 0 & 0 \end{pmatrix} \;,\qquad D_{\mathbf{s}}^{(-)}(\theta) \;=\; \begin{pmatrix} 0 & 0 \\ 0 & D_{\bar{\mathbf{4}}}(\theta) \end{pmatrix} \;, \tag{A.13}$$

respectively. Note, that the four-dimensional representations $\mathbf{4}$ and $\bar{\mathbf{4}}$, are connected by charge conjugation (A.7) as $D_{\bar{\mathbf{4}}}(\theta) = c^{-1}\,D_{\mathbf{4}}(\theta^{-1})^T\,c$.

A.4 An explicit Spin(6) basis

A convenient choice for the Γ-matrices in six dimensions is given by

$$\begin{aligned} \Gamma_1 &= \sigma_1 \otimes \mathbb{1}_2 \otimes \mathbb{1}_2 \ , & \Gamma_3 &= \sigma_3 \otimes \sigma_1 \otimes \mathbb{1}_2 \ , & \Gamma_5 &= \sigma_3 \otimes \sigma_3 \otimes \sigma_1 \ , \\ \Gamma_2 &= \sigma_2 \otimes \mathbb{1}_2 \otimes \mathbb{1}_2 \ , & \Gamma_4 &= \sigma_3 \otimes \sigma_2 \otimes \mathbb{1}_2 \ , & \Gamma_6 &= \sigma_3 \otimes \sigma_3 \otimes \sigma_2 \ , \end{aligned} \tag{A.14}$$

where the chirality operator reads

$$\tilde{\Gamma} = \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \ . \tag{A.15}$$

Furthermore, the corresponding Cartan algebra of Spin(6) can be represented by commuting spin generators $\Sigma_1 = \Sigma_{12}$, $\Sigma_2 = \Sigma_{34}$ and $\Sigma_3 = \Sigma_{56}$ given by

$$\Sigma_1 = \tfrac{\mathrm{i}}{2}\, \sigma_3 \otimes \mathbb{1}_2 \otimes \mathbb{1}_2 \ , \quad \Sigma_2 = \tfrac{\mathrm{i}}{2}\, \mathbb{1}_2 \otimes \sigma_3 \otimes \mathbb{1}_2 \ , \quad \Sigma_3 = \tfrac{\mathrm{i}}{2}\, \mathbb{1}_2 \otimes \mathbb{1}_2 \otimes \sigma_3 \ . \tag{A.16}$$

Appendix B

Statistical methods

B.1 Probabilities

In a frequentist [210] perspective the probability $\Pr(X)$ of an event X is given by the proportion of occurrence $\widehat{\Pr}(X)$ for observed events,

$$\widehat{\Pr}(X) = \frac{|X|}{N} \quad \xrightarrow{N \to \infty} \quad \Pr(X) \ , \tag{B.1}$$

which converges to the true probability for an infinite number of trials N.

The basic computation rules for probabilities $\Pr(X)$ of an event X, can be summarized by the sum and product rule [194],

$$\begin{aligned} \Pr(X) &= \sum_Y \Pr(X, Y) \ , & \text{(B.2)} \\ \Pr(X, Y) &= \Pr(X|Y) \Pr(X) \ , & \text{(B.3)} \end{aligned}$$

where $\Pr(X|Y)$ is the conditional probability for X, given that event Y appeared and $\Pr(X, Y)$ is the joint probability of X and Y. Even though these building blocks are simple, they are frequently used and can lead to superior solutions, see fig. 4.1. Moreover, they are a crucial concept in generative deep learning models, i.e. so-called autoregressive models.

B.2 Distributions

In the context of this thesis, some distributions are frequently used and are summarized here.

Let us start with the discrete uniform distribution,

$$d \sim \mathcal{U}\{a, b\} \ . \tag{B.4}$$

Using this distribution one draws d from the integer set $X = \{a,\ a+1, \dots,\ b\ \}$, i.e. $a, b \in \mathbb{Z}$, with equal probability for each element of the set $\Pr(x \in X) = \frac{1}{b-a+1}$ to occur as result. As the distribution is discrete it corresponds to a probability mass function for d.

The most famous probability distribution used in this thesis is the normal distribution (or Gaussian),

$$\mathcal{N}(\mu, \sigma^2) \ = \ \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \ . \tag{B.5}$$

Specified by the parameters mean μ and standard variance σ, it gives the probability density function for the random variable $x \sim \mathcal{N}(\mu, \sigma^2)$. The normal distribution is used for statistical approximations that are based on the central limit theorem. This theorem ensures that the distribution of the means μ, of any distribution, are converging towards a normal distribution. In this context, also the confidence intervals are easy to compute, and several well-defined approximations exist.

Finally, there is the binomial distribution. This distribution is used in scenarios like the results of a coin flip. In this thesis, something similar happens when drawing d_{kI} from the discrete Uniform distribution and observe if the resulting vector M_k fulfills the corresponding modular invariance conditions or not. Moreover, for some statistical experiments, the distribution of the samples can not be approximated by the normal distribution, e.g. because the event of interest does not occur at all in the sampling due to a very tiny probability of appearance. The binomial distribution reads,

$$\mathcal{B}(t,\ p) = \binom{t}{n} (1-p)^{(t-n)}\ p^n \ , \tag{B.6}$$

which is parameterized by $t \in \mathbb{N}$ the number of trials and $p \in [0,1]$ the probability of the desired event to occur. This probability mass function then assigns different probabilities to the resulting n, i.e. the likelihood of a certain number of occurrences, such that $n \sim \mathcal{B}(t,p)$. This interpretation will later be inverted as the number of occurrences n is set such that one obtains the probability for the number of trials to reach n.

Data points that are drawn independently from the same distribution are said to be independent and identically distributed, which is often abbreviated by i.i.d. .

B.3 Estimating probabilities

The purpose of this section is to provide the statistical details for the statistics of the heterotic orbifold landscape (chapter 4) and the confidence intervals on the contrast patterns from section 6.2.2.

B.4 Binomial distribution and the probability of succeeding

In section 4.1 one draws the coefficients d_{kI} from a discrete uniform probability mass function eq. (B.4) in order to statistically solve the modular invariance conditions (MI) eq. (2.46). The building block is the probability,

$$\Pr\left(M_k = \text{MI} \;\middle|\; \{M_i\}_{i=1}^{k-1} = \text{MI}\right) \quad \text{with } \{M_i\}_{i=1}^{k-1} = \{M_1, \ldots, M_{k-1}\} \ , \tag{B.7}$$

to draw a set of d_{kI} such that the corresponding 16-vector M_k satisfies the corresponding modular invariance conditions, i.e. all modular invariance conditions that only contain the already drawn M_i from $i = 1, \ldots, k-1$ and the new M_k. This can be modeled by a binomial probability distribution. Therefore, let us denote the probability of success by p. Hence, the probability distribution to find n modular invariant 16-vectors within t trials is given by eq. (B.6). During the process of constructing modular invariant gauge embedding matrices, we can count the number of failures until we find M_k such that modular invariance is fulfilled. This means we can keep track of the case $n = 1$ counting the number t of trials. Note that the binomial factor equals one as we only consider the single combination for the successful drawing to appear, i.e. at the

Event	$\mathbb{E}(t)$	$\Pr\left(M_k \,\middle|\, \{M\}_{i=1}^{k-1} = \mathrm{MI}\right)$	σ
$V_1 = \mathrm{MI}$	6.04	$1.66 \cdot 10^{-1}$	$2.31 \cdot 10^{-4}$
$V_2 = \mathrm{MI} \mid V_1 = \mathrm{MI}$	1.00	1.00	0.00
$W_1 = \mathrm{MI} \mid \{V_1, V_2\} = \mathrm{MI}$	1.00	1.00	0.00
$W_2 = \mathrm{MI} \mid \{V_1, V_2, W_1\} = \mathrm{MI}$	1.00	1.00	0.00
$W_3 = \mathrm{MI} \mid \{V_1, ..., W_2\} = \mathrm{MI}$	36.04	$2.775 \cdot 10^{-2}$	$1.019 \cdot 10^{-4}$
$W_4 = \mathrm{MI} \mid \{V_1, ..., W_3\} = \mathrm{MI}$	1.00	1.00	0.00
$W_5 = \mathrm{MI} \mid \{V_1, ..., W_4\} = \mathrm{MI}$	282.57	$3.5389 \cdot 10^{-3}$	$3.6841 \cdot 10^{-5}$
$W_6 = \mathrm{MI} \mid \{V_1, ..., W_5\} = \mathrm{MI}$	1131.99	$8.83400 \cdot 10^{-4}$	$1.84311 \cdot 10^{-5}$

Table B.1: *Results from a dataset D in $\mathbb{Z}_6$-II (1,1) with 10^6 constructed modular invariant gauge embedding matrices with $\beta = 2$. The first column describes the event to occur in the successive construction of gauge embedding matrices. The second column shows the mean number of trials in D, necessary to find a 16-vector that fulfills the corresponding modular invariance conditions. The third column is the resulting estimate for the probability of finding a 16-vector eq.* (B.11). *The last column gives the standard deviation eq.* (B.12) *for the probability of the third column.*

last step. This means we stop the trials t as soon as a valid vector is drawn. This constraint appears to form the specifics of our counting technique. Hence,

$$\Pr(t) = (1-p)^{(t-1)} p \ . \tag{B.8}$$

Then we can compute the expectation value of t, i.e. the mean of t:

$$\mathbb{E}[t] = \sum_t t \ \Pr(t) = \sum_{t=1}^{\infty} t \ (1-p)^{(t-1)} \ p = p \ \sum_{t=1}^{\infty} t \ (1-p)^{(t-1)} = p \ \frac{1}{p^2} = \frac{1}{p} \ , \tag{B.9}$$

where we used the geometric series,

$$f \ = \ \sum_{t=0}^{\infty} a_0 \ q^t \ = \ \frac{a_0}{1-q} \quad \Rightarrow \quad \frac{df}{dq} \ = \ \sum_{t=1}^{\infty} a_0 \ t \ q^{(t-1)} \ = \ \frac{a_0}{(1-q)^2} \ . \tag{B.10}$$

Therefore, the probability to find a valid 16-vector can be estimated by the mean of all the trials constructing this 16-vector,

$$\Pr\left(M_k = \mathrm{MI} \,\middle|\, \{M_i\}_{i=1}^{k-1} = \mathrm{MI}\right) \ = \ \frac{1}{\mathbb{E}(t)} \ , \tag{B.11}$$

with t the number of trials until a consistent 16-vector was drawn. To be precise the probability of success p is given by $\Pr\left(M_k = \mathrm{MI} \,\middle|\, \{M_i\}_{i=1}^{k-1} = \mathrm{MI}\right)$. The label k is inherently in t as only the trials for the specific 16-vector M_k are used in the computation. The results on a dataset generated according to the successive search in fig. 4.1 with $\beta = 2$ are given in table B.1

B.5 Confidence interval and margin of error

A standard procedure to estimate the margin of error of a variable for a certain confidence interval is given by approximating the sampling distribution by the normal distribution $\mathcal{N}\left(\mu|\sigma^2\right)$.[1] Especially in our case, this is a valid approach since the central limit theorem guarantees that the distribution of the mean of any distribution converges to the normal distribution. Remember that we will primarily be interested in the mean as in eq. (B.11). Our goal is to find the range within the unknown actual probability $\Pr(A)$ lies with a certain confidence level. As we have no direct access to the probability $\Pr(A)$, we will take finite samples from the distribution behind $\Pr(A)$ and compute the sample proportion $\widehat{\Pr}(A)$. This sample proportions distribute as the normal distribution with $\Pr(A)$ as the mean μ. Hence, with 95% the $\widehat{\Pr}(A)$ are in the range of $2\sigma_{\Pr(A)}$ of the true probability $\Pr(A)$. This statement can be reformulated in a more practical one, as it is equivalent to a 95% probability that $\Pr(A)$ is within $2\sigma_{\widehat{\Pr}(A)}$ of $\widehat{\Pr}(A)$. Hence, we can estimate the true probability $\Pr(A)$ on the statistics of our samples to which we have full access. The standard error $\sigma_{\widehat{\Pr}(A)}$ is an unbiased estimator for $\sigma_{\Pr(A)}$ such that we find,

$$\sigma_{\widehat{\Pr}(A)} = \sqrt{\frac{\widehat{\Pr}(A)\left(1-\widehat{\Pr}(A)\right)}{n}} \,. \tag{B.12}$$

Error propagation

With eq. (B.12) we can estimate the margin of error for the probability of finding an individual M_k vector in the successive construction of the gauge embedding matrix M. To estimate the error for the full gauge embedding matrix M we can use table B.1 and the error propagation formula [211],

$$\sigma_f = \sqrt{\sum_x \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2} \,, \tag{B.13}$$

where in our case f corresponds to $\Pr(M = \mathrm{MI})$ and the sum over x goes over the probabilities for finding the M_k vectors, i.e. $\Pr\left(M_k = \mathrm{MI} \,\middle|\, \{M_i\}_{i=1}^{k-1} = \mathrm{MI}\right)$ for $k = 1, \ldots, 8$. Assuming a vanishing covariance between the different probabilities.

B.6 Probability of absent events: The Rule of Three

In case a specific event X did not appear in the sampling procedure, the typical method of using the normal distribution to compute the margin of errors is not a valid approximation. Then the so-called 'Rule of Three' method can be applied [212, 213]. In detail, the method is based on the binomial distribution, i.e. on the binomial one-sided confidence interval for the binomial distribution eq. (B.6) with t number of trials and n number of times event X appears. For example X can be MSSM-like models that do not fulfill the contrast pattern constraint eq. (6.24). Then, p is the probability of such an absent event to occur in a broader scan, e.g. to find an MSSM-like model that does not fulfill the constraint. Hence, since the absent event X does not occur at all ($n = 0$), we find,

$$\Pr(X \text{ is absent}) = (1-p)^t \,. \tag{B.14}$$

---

[1] This approximation is only valid for certain constraints on the total number of samples n and the probability p of a binary event [180], e.g. a popular qualification is that the approximation holds if $np \geq 5$ and $n(1-p) \geq 5$.

The aim is to estimate the probability p, which can be done in a certain confidence interval (CI). The probability of observing only negative results can have any value provided that it is more significant than $(1 - \mathrm{CI})$. Otherwise, it is assumed to be unlikely for p to show no positive results. This leads to the inequality,

$$(1-p)^t \geq (1-\mathrm{CI}) \; . \tag{B.15}$$

This can be reformulated into an upper bound for the probability p of the event X to appear,

$$p \leq 1 - \sqrt[t]{(1-\mathrm{CI})} \; . \tag{B.16}$$

The 'Rule of Three' then uses for CI=0.95 and p close to zero, the approximations $\ln(1-p) = -p + \mathcal{O}(p^2)$ and $\ln(0.05) \approx -3$ to find,

$$p \geq \frac{3}{t} \tag{B.17}$$

hence the name 'Rule of Three'.

Appendix C

Machine learning

This section is dedicated to providing detailed explanations on the general concept of machine learning (illustrated in fig. C.1), as well as on some exemplary techniques, which were used in this thesis (see also [181] for an overview of these techniques applied to string theory). Machine learning is a set of routines to turn information in the form of data into knowledge. In detail, the goal is to learn a function

$$Y_{\text{predicted}} : \mathcal{X} \to \mathcal{Y} \ , \tag{C.1}$$

that maps the data of the input space $\mathcal{X}$ to the space of the output or target space $\mathcal{Y}$. The spaces $\mathcal{X}$ and $\mathcal{Y}$ are often identified with $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

The input space $\mathcal{X}$ is usually called *feature space*, and the dimension n corresponds to the number of different *features*, i.e. they label the coordinates of the space $\mathcal{X}$. These so-called features can be raw numerical values from the input data, e.g. the individual pixel values of an image. However, they can also be sophisticated properties computed from the raw data. In our case, physical quantities. For example, the number of surviving roots on fixed points, computed from the gauge embedding matrix M (section 2.2.2). In this thesis, M represents the raw data from heterotic orbifold compactifications. This kind of creating powerful features is called *feature engineering* (see appendix C.2 for details). The elements $x \in \mathcal{X}$ are then the corresponding feature vectors. The dataset D is in general a sequence of tuples (x_k, y_k), enumerated by k. The machine learning algorithm is trained to learn the mapping $Y_{\text{predicted}}(x_k) = y_k$ for any possible pair (x_k, y_k) even if they are not covered in the dataset D. This property is called the generalization ability of the machine learning model.

In the context of this thesis, two different kinds of learning scenarios appear. In chapter 5 a neural network, i.e. autoencoder, is trained in a so-called *unsupervised* process. In contrast, chapter 6 and chapter 7, make use of *supervised* training of decision tree algorithms. These two branches of machine learning refer to the conditions if the dataset D contains so-called labels $y \in \mathcal{Y}$ for the target space or not. In particular, in supervised learning, the model learns from the space $\mathcal{X} \times \mathcal{Y}$ and masters to approximately reproduce the same labels for the training data that were given. In the unsupervised learning scheme, the algorithm has no examples of $\mathcal{Y}$ but rather determines a meaningful mapping based on patterns of the data in $\mathcal{X}$. In most cases, the dimension of $\mathcal{Y}$ or other general concepts of the target space are given instead of specific examples.

C.1 Splitting the dataset and data leakage

In order to verify if the general idea of learning the function Y_{predict} was successful, the generalization ability of the model has to be estimated after the training process. Therefore, the procedure in machine learning is as follows: The complete dataset D is split in order to have a held-out dataset with data that was not used during training. The performance on this dataset estimates how well the machine learning model generalizes to new unseen data. In detail, the dataset D is often split into three datasets:

- $D^{\texttt{train}}$: Dataset used to train the inner parameters of the machine learning model explicitly.
- $D^{\texttt{val}}$: Dataset used to adjust the hyperparameters of the machine learning model and to estimate the generalization ability. Hyperparameters are arguments on the machine learning algorithm that define properties of the training in general. Most importantly, they are used for regularization in order to maximize the generalization to new data by avoiding overfitting to the training data.
- $D^{\texttt{test}}$: Dataset for the final evaluation. This data is not used to adjust any (hyper) parameters of the machine learning model and solely determines the final generalization ability. If there is no flaw in the splitting of the dataset, this estimate coincides with the real-world application of the machine learning model. More precisely, the application of the machine learning model to the feature space $\mathcal{X}$ such that any point of the system can appear without limitation to those data points contained in D.

A common way to perform this splitting is given by a random division of the dataset D. In this case an element $(x_k, y_k) \in D$ has the probability $p_{\texttt{test}}$ to be send to $D^{\texttt{test}}$, the probability $p_{\texttt{val}}$ to be send to $D^{\texttt{val}}$ and the probability $p_{\texttt{train}}$ to be send to $D^{\texttt{train}}$, where $p_{\texttt{test}} + p_{\texttt{val}} + p_{\texttt{train}} = 1$. This can for example be done in a successive approach by `train_test_split` from the sklearn package [189], which divides the data in two sets. A different arrangement for small datasets is called cross-validation. Detailed explanations are given in appendix C.4. This technique was utilized in chapter 7, even though it is a computationally more intensive method.

While the above set up gives a general procedure that should ensure the learning of a function $Y_{\text{predicted}}$ that generalizes well on new data, it is in practice important to know the specific goal of the prediction as well as the structure of the dataset. This system knowledge allows avoiding overfitting that was not prevented by the dataset split.[1] This difficulty refers to the term *data leakage*. The naming is based on the issue that the data separation was not done carefully enough. Concretely, information of the $D^{\texttt{test}}$ ($D^{\texttt{val}}$) set 'leaked' into the training process, such that the algorithm could overfit on these data. Then applying the algorithm in a real-world application, the generalization to completely new data is much worse than expected from the evaluation on $D^{\texttt{test}}$ ($D^{\texttt{val}}$). The algorithm underperforms in real-world applications and does not match the estimated performance of the development environment. A dataset with several duplicates gives a simple case of this situation. The random split then might move one instance of the duplicate pair towards $D^{\texttt{test}}$ while the other instance remains in $D^{\texttt{train}}$. Hence the machine learning algorithm sees the duplicate of the instance during training and can fit directly towards the specific data. If then, truly new real-world data does not have the property to be a duplicate of the data already acquired for the machine learning approach, the trained model will have poor performance at least compared to the estimate from the performance on $D^{\texttt{test}}$.

---

[1] For this situation in the context of physics see section 6.2.1 for the MSSM-like classification task or at the end of section 7.1 in context of the prediction of the orbifold origin of the MSSM.

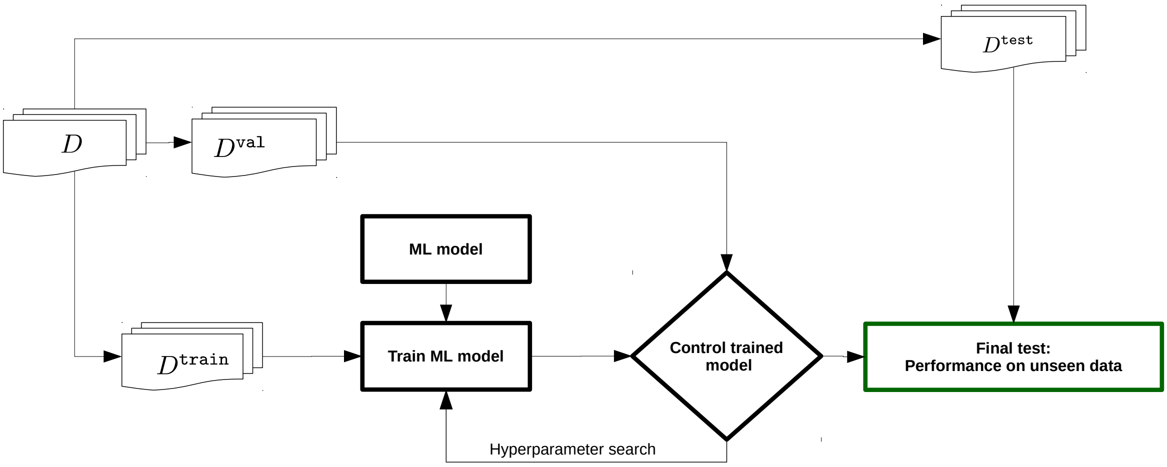


Figure C.1: *Illustration of machine learning (ML) workflow.*

C.2 Feature engineering

In general terms, *feature engineering* denotes the process of computing useful quantities from the raw data, in order to improve the learning task. This process is complicated and demands a lot of domain knowledge to develop these features.

The machine learning methods that are especially successful without these higher-level features are neural networks (see appendix C.7.2). Neural networks generate features on their own during training. Concretely, each hidden layer constructs a higher level of abstraction from the first layers towards the later. This is used in so-called *transfer learning* where the first layers are reused in another prediction task by only changing the last layer, assuming that the first layers learned basic principles. For example, in image analysis, the early layers learn basics like edge detection. However, this is one of the biggest open problems of neural networks. It is, in general, unfeasible to extract the meaning of the features that a neural network has learned on its own [214–216]. Hence, these features hardly yield any knowledge gain. Simultaneously, this is one of the reasons why the transfer of neural network concepts towards the application in string theory is so difficult. The architectures and the successful pre-trained models do not trivially generalize to the string theory data.

The alternative approach, used in this thesis, is based on physical intuition and knowledge of the system in order to create features. The machine learning techniques are used to quantify if our educated guess for a particular feature, or combinations of multiple features, leads to a correlation between these features and the property of the task at hand. The advantage of this approach is that by construction, we have a straightforward physical interpretation of our features.

C.3 Regularization

Regularization in the context of machine learning refers to the concept of additionally constraining the freedom of the model parameters to avoid overfitting. Overfitting is the phenomenon that the machine learning model parameters are fine-tuned to the training data $D^{\texttt{train}}$. Consequently, the trained model has a poor performance on unseen data, even if it comes from the same origin (distribution) as the training data.

In the context of this thesis, this was used in several approaches. In detail, the information bottleneck layer of the autoencoder in chapter 5 was a strong regularization of the neural network. Decision trees, especially in chapter 6, were regularized by limiting the `depth` of the tree, i.e. the number of splits that can be performed. Finally, in the case of the boosted decision tree, the ensemble technique itself is a form of regularizing the algorithm. Moreover, the individual trained trees were regularized by limiting the number of leaf nodes `num_leaves` and the minimum number of datapoints to allow for further splitting `min_child_samples`.

C.4 Cross-validation

Partitioning the available data D into three sets $D^{\texttt{train}}$, $D^{\texttt{val}}$, $D^{\texttt{test}}$, drastically reduces the number of samples which can be used for the statistical learning algorithm. One possible solution to this problem is a procedure called cross-validation (CV). The validation set $D^{\texttt{val}}$ is in this approach a dynamic subsequence of the training data $D^{\texttt{train}}$ and not a separate dataset. In detail, the classical approach is k-fold CV. Here $D^{\texttt{train}}$ is split into k equal subsets $D_k^{\texttt{train}}$. One of these k sets is then defined to be the validation set $D_y^{\texttt{train}} = D^{\texttt{val}}$. Then the following procedure is applied for each of the k folds: First, a model is trained using $k-1$ of the folds as training data $\bigcup_{k \neq y} D_k^{\texttt{train}}$. Secondly, the resulting model is validated on the remaining part of the data, i.e. $D^{\texttt{val}} = D_y^{\texttt{train}}$. Let us stress that this is done k times each time with the next chunk as the validation set, i.e. $y = 1, \ldots, k$. The performance measure reported by k-fold cross-validation is then the average of the values computed in the loop over y. This approach can be computationally expensive but does not waste data, which is a major advantage in problems where the number of samples is small. Often this technique is combined with a *grid search* for the optimal hyperparameters, as e.g. scikit-learn [189] provides with `GridSearchCV`. Then the CV is done for every point in the grid of hyperparameters, i.e. all combinations of the values for the hyperparameters are explicitly tested. The trained model with the best-averaged performance measure after CV is then selected as the best hyperparameter set. Note, there are also random approaches `RandomizedSearchCV` and Bayesian hyperparameter optimization [217] based on Gaussian processes fitted to the already used hyperparameters and their performance results.

C.5 Classification

In general machine learning tasks are divided into regression and classification tasks. Regression tasks refer to predict continuous target values. In a classification tasks the machine learning approach is asked to predict abstract classes as target value y, i.e. the labels do not have a inherent order like values on $\mathbb{R}$. As example one can define the two classes MSSM-like and $\cancel{\text{MSSM}}$-like which are the labels of interest when searching the landscape for physical models. This kind of binary classification task has different outcomes as shown in table C.1, i.e. there exist misclassification errors.

Before we discuss the measurement of the performance of such a classification, let us generalize to n-classes. While in binary classification tasks like above the classes can simply the represented by the binary values 0 and 1, e.g. by a sigmoid function, this is not possible for

		predicted class	
		Positive	Negative
true class	Positive	*true positives*	*false negative*
	Negative	*false positive*	*true negatives*

Table C.1: *Special case with two different classes. One is called positive, while the complementary is named negative. In our case positive can be set to MSSM-like and negative to* $\cancel{\mathit{MSSM}}$*-like.*

multiclass classification. A common method for this situation is one-hot encoding,

$$\begin{aligned} \mathtt{OHEnc} : \{\mathtt{c}_1, \mathtt{c}_2, \dots \mathtt{c}_n\} &\to \{0,1\}^n \\ \mathtt{c}_x &\mapsto (\delta_{1x}, \delta_{2x}, \cdots, \delta_{nx}) \ . \end{aligned} \tag{C.2}$$

This function maps categorical variables $\mathtt{c}_x$ to vectors of length n such that only one entry is 1 and all others are zero, see eq. (C.2). The generalization of the binary classification case shown in table C.1 is the so-called confusion matrix C, where C_{ij} contains the number of observation with true class $\mathtt{c}_i$ and predicted class $\mathtt{c}_j$. Hence, we can generalize the classification types for a class $\mathtt{c}_i$ in the multiclass setting as,

$$\#\text{ of true positives for } \mathtt{c}_i : C_{ii} \ , \tag{C.3}$$

$$\#\text{ of false positives for } \mathtt{c}_i : \sum_{j \neq i} C_{ij} \ , \tag{C.4}$$

$$\#\text{ of false negatives for } \mathtt{c}_i : \sum_{j \neq i} C_{ji} \ . \tag{C.5}$$

The performance of a predictive machine learning model can be measured by the accuracy that is defined by the number of correct predictions divided by the total number of all predictions. Or in terms of the confusion matrix C,

$$\mathtt{Accuracy} \ = \ \frac{\sum_i C_{ii}}{\sum_{ij} C_{ij}} \ . \tag{C.6}$$

However, for an imbalanced dataset this normal accuracy measurement can be misled in the following way: Assume a classification task with two classes A and B, where the class A builds the majority of the dataset with 99% of all instances. Any prediction method can now achieve an accuracy of 99% simply by predicting always class A, but never class B.

In order to avoid such a behavior, there exist several metrics for imbalanced classification tasks. The one we are using are based on the types of predictions from eqs. (C.3) to (C.5). This allows us to define three different metrics for an imbalanced classification task,

$$\mathtt{Precision}(\mathtt{c}_i) \ = \ \frac{C_{ii}}{\sum_j C_{ij}} \tag{C.7}$$

$$\mathtt{Recall}(\mathtt{c}_i) \ = \ \frac{C_{ii}}{\sum_j C_{ji}} \tag{C.8}$$

$$\mathtt{f1\text{-}score}(\mathtt{c}_i) \ = \ 2\,\frac{\mathtt{Precision}(\mathtt{c}_i) \cdot \mathtt{Recall}(\mathtt{c}_i)}{\mathtt{Precision}(\mathtt{c}_i) + \mathtt{Recall}(\mathtt{c}_i)} = 2\,\frac{C_{ii}}{\sum_j C_{ji} + \sum_j C_{ij}} \tag{C.9}$$

Finally, one can define the "`f1-macro`" as the average of the `f1-scores`

$$\texttt{f1-macro} = \sum_{i=1}^{n} \texttt{f1-score}(\texttt{c}_i) \tag{C.10}$$

for all classes. Then, to deal with imbalanced data, the `f1-macro` can be used to measure the performance of the model.

Let us briefly illustrate the benefit of using the `f1-macro` on the example from the beginning of this section. In this case, we have $\texttt{f1-score}(A) \approx 1$ and $\texttt{f1-score}(B) = 0$. Consequently, the $\texttt{f1-macro}={}^1\!/_2\big(\texttt{f1-score}(A) + \ \texttt{f1-score}(B)\big) \approx 0.5$, which rates the naive classification of a binary classification task as insufficient.

C.6 Decision tree

A decision tree[2] (fig. 6.4) is a supervised learning algorithm and therefore makes use of the labels y of $D^{\texttt{train}}$ in order to find correlations in the data, based on the features f (remember that features are called the dimensions of the space $\mathcal{X}$) to predict the label y,

$$\begin{aligned} \texttt{DT} : \mathcal{X} &\to \mathcal{Y} \\ x &\mapsto y \ . \end{aligned} \tag{C.11}$$

It recursively partitions the feature space such that the samples with the same labels are grouped together. Doing so he uses only so called thresholds on certain features f and hence splits the space orthogonal to the feature axes. Let the data at node m be represented by $D^{\texttt{train}}(m)$. For each possible split $s = (f_j, t_m)$ consisting of a feature f_j and threshold t_m, the tree partitions the data into complementary subsets $D_L^{\texttt{train}}(m, s)$ and $D_R^{\texttt{train}}(m, s)$,

$$D_L^{\texttt{train}}(m, s) = \left\{ \ (x, y) \in D^{\texttt{train}}(m) \ \middle| \ f_j \leq t_m \ \right\} \tag{C.12}$$

and consequently $D_R^{\texttt{train}}(m, s) = D^{\texttt{train}}(m)/D_L^{\texttt{train}}(m, s)$. The impurity G at node m is computed using an impurity function H,

$$G(D^{\texttt{train}}(m), s) = \frac{n_L}{N_m} H(D_L^{\texttt{train}}(m, s)) + \frac{n_R}{N_m} H(D_R^{\texttt{train}}(m, s)) \tag{C.13}$$

where $N_m = n_L + n_R$ with $n_{L/R} = |D_{L/R}^{\texttt{train}}(m, s)|$. The algorithm selects the parameters s^* that minimizes the impurity G,

$$s^* = \operatorname{argmin}_s G(D^{\texttt{train}}(m), s) \tag{C.14}$$

This is done recursively for subsets $D_L^{\texttt{train}}(m, s^*)$ and $D_R^{\texttt{train}}(m, s^*)$ until the regularization parameters prevent further training, e.g. maximum number of splits or minimal number of samples per node $N_m \leq \texttt{min\_samples}$ are violated. If no regularization constrain is exceeded the decision tree can train until the final nodes (leaf nodes) consists of only one element, i.e. $N_m = 1$.

If a target y is a classification outcome taking on values $\texttt{c}_k$ for $k = 1, \ldots, K$ for node m with N_m observations, let

$$p_m(k) = \frac{1}{N_m} \sum_{(x_i, y_i) \in D^{\texttt{train}}(m)} \delta(y_i, \texttt{c}_k) \tag{C.15}$$

---

[2] For an introduction to tree based methods see [197, 218] and the documentation of [189].

be the proportion of class k observations in node m. A common measures of impurity is the Gini index,

$$H(D^{\texttt{train}}(m)) = \sum_{k=1}^{K} p_m(k) \ (1 - p_m(k)) \ . \tag{C.16}$$

A huge benefit when we have to deal with imbalanced data is that the classification decision tree, `DecisionTreeClassifier`, of sklearn [189] allows for balancing the importance of the data with class weights (`class_weight = 'balanced'`), i.e. those weights raise the importance of a class with small frequency of occurrence and decreases the weights of classes that dominate the dataset.

C.7 Neural networks

Neural networks are a specific technique from machine learning. They are special since neural networks are universal function approximators [219, 220] for continuous functions. This means they can approximate the effective mapping of a wide variety of interesting functions. A neural network is a computational graph [221] where the building blocks of a neural network are so-called neurons (see fig. C.2). These neurons n_i are computational functions,

$$n_i(x) = f\left(\sum_j W_{ij} x_j + b_i\right) \ , \tag{C.17}$$

where the scalar valued W_{ij} are the so-called weights of neuron n_i. The parameter b_i is the bias value and f is the activation function. The weight vector and bias are adjustable parameters of the function to fit the data during training. The activation function adds non-linearity to this object. The universal function approximation theorem holds only for such induced non-linearities. Putting many of these neurons together results in a neural network (see fig. C.3). The basic approach is to align them in so-called layers. The number of neurons in one layer define the corresponding width. Stacking several of those layers on top of each other results in a deep neural network where the depth is defined by the number of layers stacked on each other.

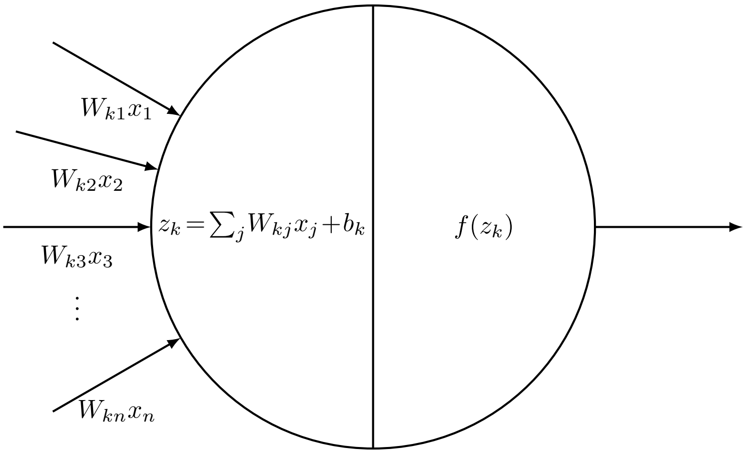


Figure C.2: *Illustration of a neuron.*

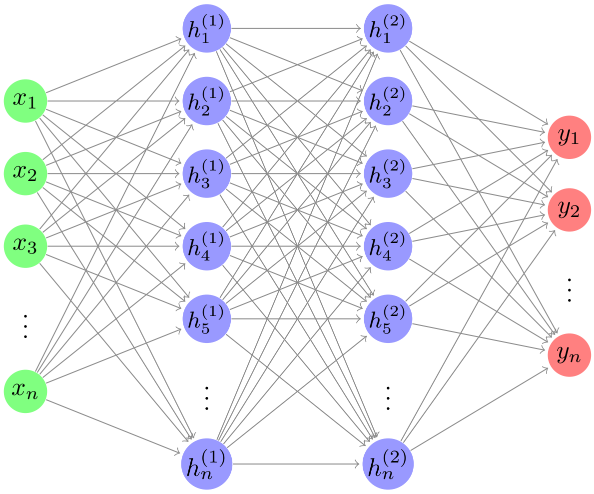


Figure C.3: *Illustration of a* feedforward neural network *(FNN). The green neurons x_i build the input layer for the data, The blue neurons $h_i^{(1)}$ and $h_i^{(2)}$ represent the two hidden layers. Finally, the red neurons y_i form the output layer.*

There are special layers like the input, output and hidden layer. Input layer simply feed the data into the network. Hidden layers are all the layers in between the input and output layer. They are called hidden since their output is not accessed but seen as the internal operations of the function. At the end we have the the output layer. This layer gives the prediction (or image of the function) of the neural network for the input.

We can then formalize the function of the neural network as,

$$n_o^{\text{out}} = f^{(n)} \left(\sum_m W_{om}^{(n)} \ ... \ f^{(2)} \left(\sum_i W_{ki}^{(2)} \ f^{(1)} \left(\sum_j W_{ij}^{(1)} x_j + b_i^{(1)} \right) + b_k^{(2)} \right) ... + b_o^{(n)} \right) \tag{C.18}$$

where the superscript (n) enumerates the layer and n_o^{out} is the final output for each dimension o of the target space $\mathcal{Y}$. One can obtain that without the non-linearity of the activation function f the network would be a simple linear mapping, i.e. matrix multiplication. Hence, adding additional layer could be collapsed to a single matrix multiplication. In this sense, the non-linearities are the reason why multiple layers stacked on each other can lead to a more sophisticated function approximator.

C.7.1 Training neural networks

The concept of adjusting the parameters in a neural network is very general among this class of techniques. First we define a quantity that we want to optimize, i.e. the objective function [221, p.78] (also called cost function or loss function). It is defined as,

$$L(n^{\text{out}}, y) \tag{C.19}$$

where y is the ground truth, i.e. the supervised target labels in $D^{\texttt{train}}$. The other part n^{out} is the prediction of the neural network and is the neural network realization of $Y_{\text{predicted}}$ for a generic machine learning function. In order to train a neural network one uses a stochastic gradient descent approach. First the parameters of the neural network are randomly initialized. Then the input x_k of $D^{\texttt{train}}$ are processed by the neural network and give a first prediction.

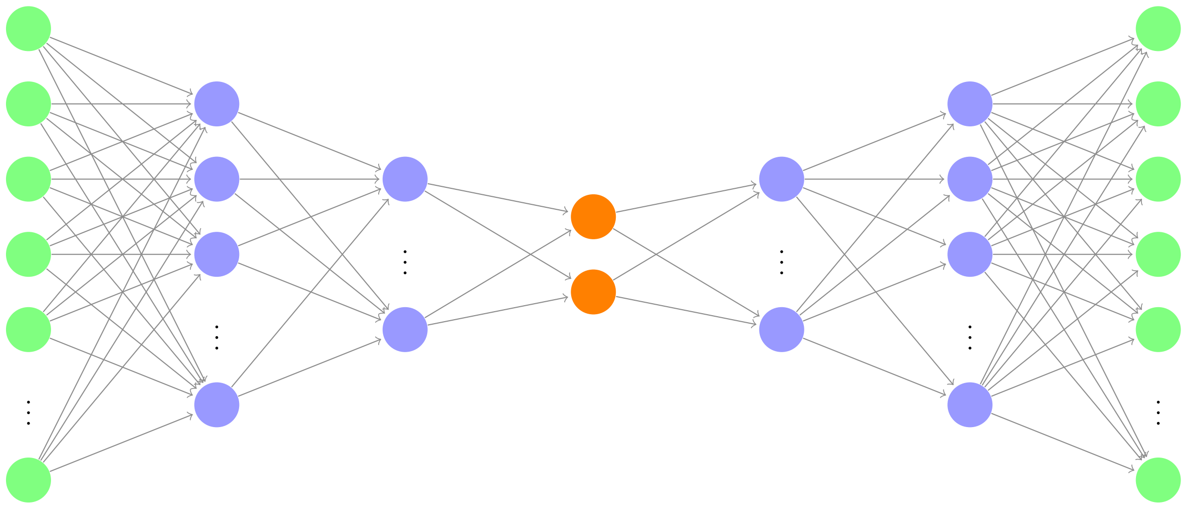


Figure C.4: *Conceptual architecture of an autoencoder neural network with latent layer of dimension two in orange. To emphasize the autoencoding concept, i.e. output equals input, the input and output layer are both in green.*

This first prediction is used to compute the value of the loss function L with the ground truth values y from $D^{\texttt{train}}$. As the goal of this approach is to minimize L the gradient of L with respect to all adjustable parameters w is computed,

$$\nabla_w L(w) \; , \tag{C.20}$$

in terms of a computational graph, i.e. recursively applying the chain rule of calculus. Then each parameter is updated according to the negative gradient correspondent to its contribution to L. The algorithm to compute the negative gradient with respect to the adjustable parameters is called back propagation. As this is done in a iterative way τ labels the number of iterations. Then the updates of the parameters read,

$$w(\tau + 1) = w(\tau) - \alpha \nabla_w L(w(\tau)) \; . \tag{C.21}$$

As the stochastic gradient approach is taken because an exact Newtonian optimization is computationally not feasible for large datasets, the step length is still needed. This is introduced as a hyperparameter α the so-called learning rate [194]. A more advanced version of stochastic gradient descent is given by the widely used Adam optimizer [222]. As the gradient also backpropagates through the activation functions, one should be aware that an activation function with zero gradients can bring the training to a standstill, e.g. dying ReLU and vanishing gradient with sigmoid activation function are two specific issues that interfere with the learning algorithm.

C.7.2 Autoencoder

Neural networks allow for various architectures, i.e. complicated arrangements of the individual neurons that lead to different computational graphs and allow to solve different problems. In some architectures the goal is to encode some known structure of the input into the neural network. One example are convolutional neural networks that detect translations in images of our 3D world.

Other architectures force the neural network to perform a task in a specific way. This is the case for autoencoder neural networks [186] (see fig. C.4 for the network architecture) that were used in chapter 5 and are explained in detail here. A remarkable property of autoencoders is that they are unsupervised. This is achieved by a simple trick, i.e. the neural network is trained to reproduce the input at the output layer and hence the labels are given by the input itself.

$$L(n^{\text{out}}, x_{\text{truth}}) \tag{C.22}$$

The specialty in the architecture of autoencoders is that the neural network has a so-called information bottleneck. There exists a so-called latent layer that has a small dimension compared to the input dimension; hence the neural network has to encode the information from the input through the hidden layers into the latent layer and then decode this information back towards the output layer, in such a way, that it can reproduce the input as good as possible. This forces the neural network to find redundancies in the feature space (such as irrelevant features) in order to find a non-linear map to a lower dimensional parameter space. Thus, an autoencoder yields a lower-dimensional, “compressed”, representation of the feature space in the latent space.

Appendix D

Aspects of orbifolds

D.1 Simple roots of $\mathrm{E}_8 \times \mathrm{E}_8$

The simple roots α_I of $\mathrm{E}_8 \times \mathrm{E}_8$ are given by the simple roots of E_8 as,

$$\alpha_I = \left(\alpha_I^{\mathrm{E}_8},\ \alpha_0^{\mathrm{E}_8}\right) \quad \text{for } I = 1, \ldots, 8 \tag{D.1}$$

$$\alpha_I = \left(\alpha_0^{\mathrm{E}_8},\ \alpha_{(I-8)}^{\mathrm{E}_8}\right) \quad \text{for } I = 9, \ldots, 16 \tag{D.2}$$

where we introduced $\alpha_0^{\mathrm{E}_8} = (0,0,0,0,0,0,0,0)$ and the simple roots basis for E_8 ,

$$\begin{array}{lcccccccccccl}
\alpha_1^{\mathrm{E}_8} & = & (& 0, & 1, & -1 & 0, & 0, & 0, & 0, & 0 &) \,, \\
\alpha_2^{\mathrm{E}_8} & = & (& 0, & 0, & 1 & -1, & 0, & 0, & 0, & 0 &) \,, \\
\alpha_3^{\mathrm{E}_8} & = & (& 0, & 0, & 0, & 1, & -1, & 0, & 0, & 0 &) \,, \\
\alpha_4^{\mathrm{E}_8} & = & (& 0, & 0, & 0, & 0, & 1, & -1, & 0, & 0 &) \,, \\
\alpha_5^{\mathrm{E}_8} & = & (& 0, & 0, & 0, & 0, & 0, & 1 & -1, & 0 &) \,, \\
\alpha_6^{\mathrm{E}_8} & = & (& 0, & 0, & 0, & 0, & 0, & 0 & 1, & -1 &) \,, \\
\alpha_7^{\mathrm{E}_8} & = & (& \frac{1}{2}, & -\frac{1}{2}, & -\frac{1}{2}, & -\frac{1}{2}, & -\frac{1}{2}, & -\frac{1}{2}, & -\frac{1}{2}, & \frac{1}{2} &) \,, \\
\alpha_8^{\mathrm{E}_8} & = & (& 0, & 0, & 0, & 0, & 0, & 0, & 1, & 1 &) \,.
\end{array} \tag{D.3}$$

Analogously, this is defined for the dual basis

$$\alpha_I^* = \left(\alpha_I^{*\mathrm{E}_8},\ \alpha_0^{*\mathrm{E}_8}\right) \quad \text{for } I = 1, \ldots, 8 \tag{D.4}$$

$$\alpha_I^* = \left(\alpha_0^{*\mathrm{E}_8},\ \alpha_{(I-8)}^{*\mathrm{E}_8}\right) \quad \text{for } I = 9, \ldots, 16 \tag{D.5}$$

such that $\alpha_I^* \cdot \alpha_J = \delta_{IJ}$, with $\alpha_0^{\mathrm{E}_8} = \alpha_0^{*\mathrm{E}_8}$. The dual root system basis is then given by,

$$\begin{array}{lcccccccccccl}
\alpha_1^{*\mathrm{E}_8} & = & (& 1, & 1, & 0, & 0, & 0, & 0, & 0, & 0 &) \,, \\
\alpha_2^{*\mathrm{E}_8} & = & (& 2, & 1, & 1, & 0, & 0, & 0, & 0, & 0 &) \,, \\
\alpha_3^{*\mathrm{E}_8} & = & (& 3, & 1, & 1, & 1, & 0, & 0, & 0, & 0 &) \,, \\
\alpha_4^{*\mathrm{E}_8} & = & (& 4, & 1, & 1, & 1, & 1, & 0, & 0, & 0 &) \,, \\
\alpha_5^{*\mathrm{E}_8} & = & (& 5, & 1, & 1, & 1, & 1, & 1, & 0, & 0 &) \,, \\
\alpha_6^{*\mathrm{E}_8} & = & (& \frac{7}{2}, & \frac{1}{2}, & \frac{1}{2}, & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}, & -\frac{1}{2} &) \,, \\
\alpha_7^{*\mathrm{E}_8} & = & (& 2, & 0, & 0, & 0, & 0, & 0, & 0, & 0 &) \,, \\
\alpha_8^{*\mathrm{E}_8} & = & (& \frac{5}{2}, & \frac{1}{2}, & \frac{1}{2}, & \frac{1}{2}, & \frac{1}{2}, & \frac{1}{2}, & \frac{1}{2}, & \frac{1}{2} &) \,.
\end{array} \tag{D.6}$$

D.2 Twisted sectors of $\mathbb{Z}_6$-II (1,1)

For completeness we illustrate here the higher order twisted sectors of $\mathbb{Z}_6$-II (1,1). Note that for $T_{(2)}$ and $T_{(3)}$ fixed tori appear.

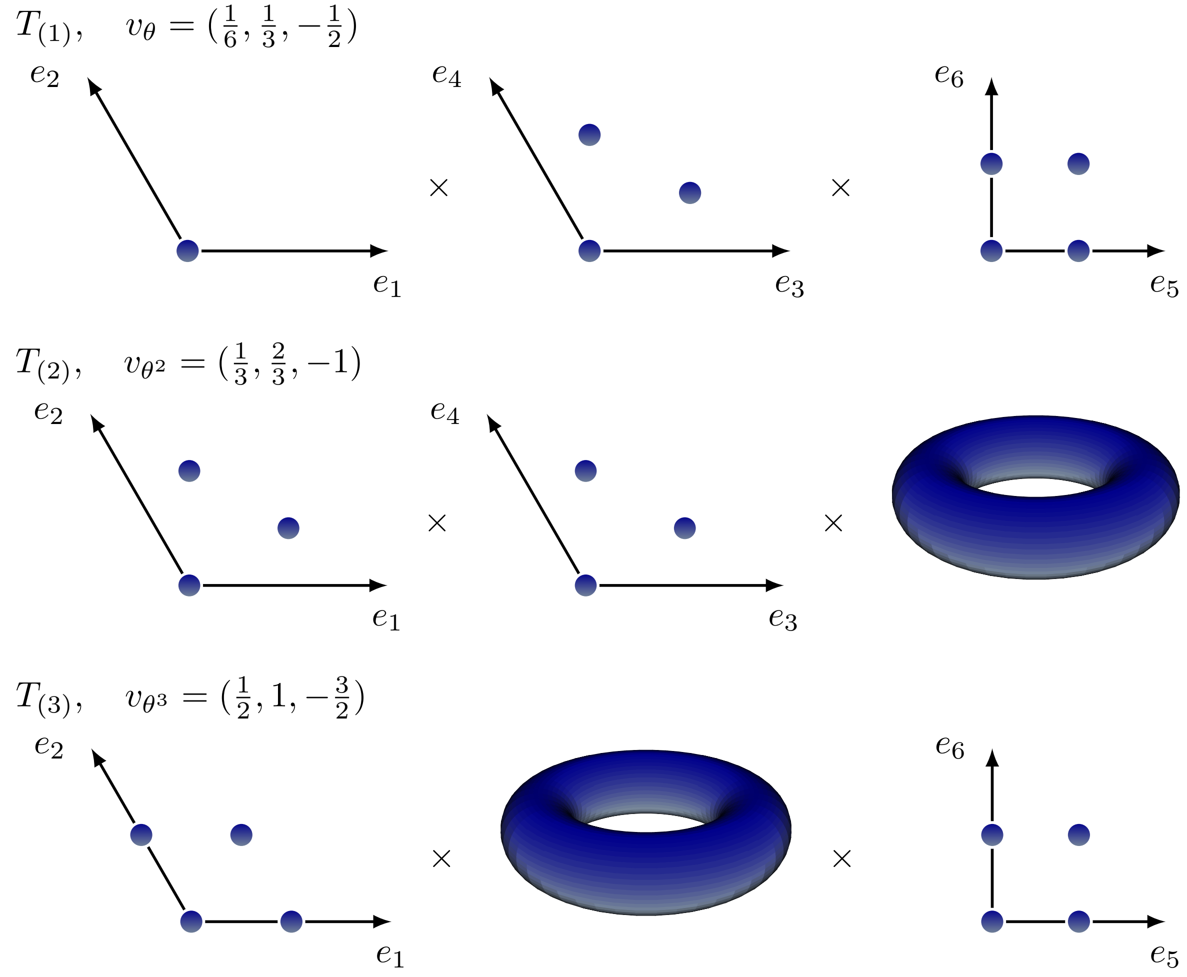
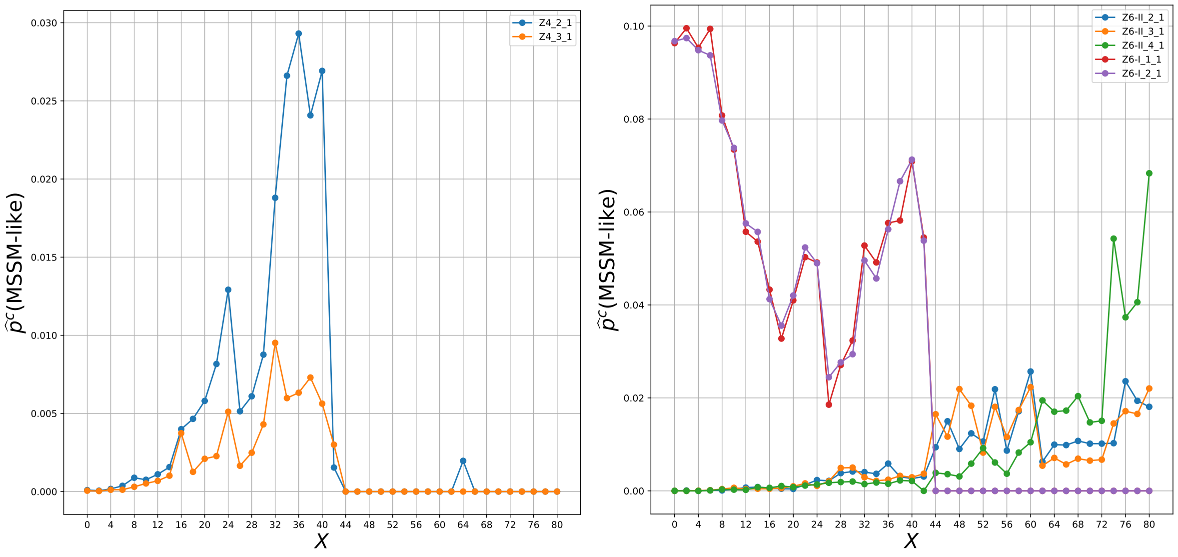


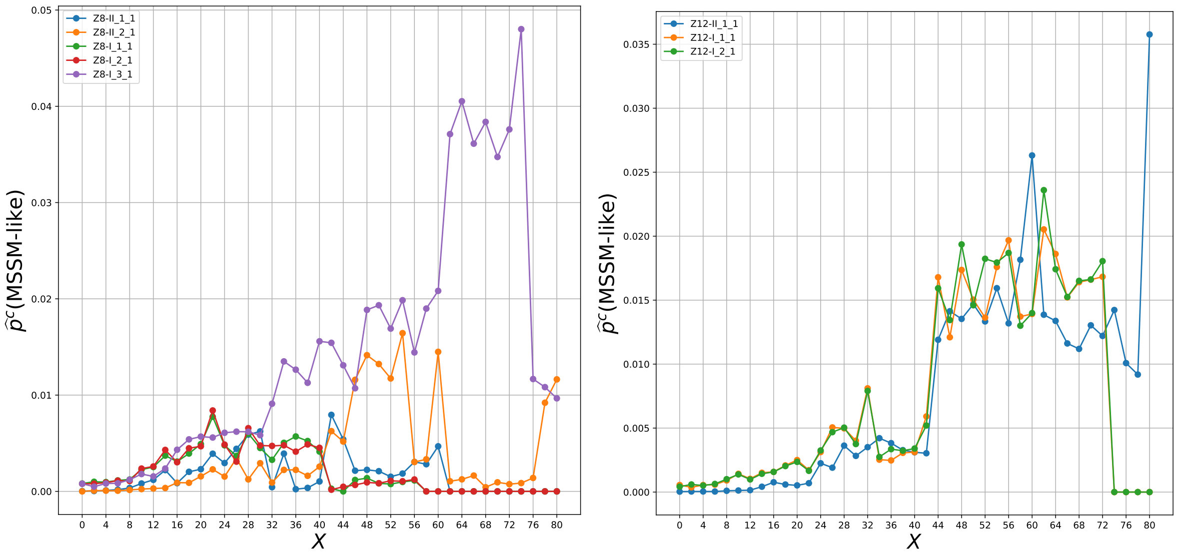
Figure D.1: *Illustration of the fixed points in* $\mathbb{Z}_6$ -II (1,1) *on* $\Gamma = \mathrm{SU(3)} \times \mathrm{SU(3)} \times (\mathrm{SU(2)})^2$. *The twisted sectors* $T_{(4)}$ *and* $T_{(5)}$ *have the same fixed point structure as* $T_{(2)}$ *and* $T_{(1)}$*, respectively.*

D.3 Probability for MSSM-like models in $\mathbb{Z}_N$ orbifolds

As in the main text for $\mathbb{Z}_6$-II (1,1) (fig. 6.10) we give here the estimated probabilities to find an MSSM-like model in the subspaces for higher $N_{\mathrm{ur}}^{(2)}(X)$ computed from the dataset observed by the scan that led to table 6.5.



(a) *Orbifold geometries based on* $\mathbb{Z}_4$ **(b)** *Orbifold geometries based on* $\mathbb{Z}_6$



(c) *Orbifold geometries based on* $\mathbb{Z}_8$ **(d)** *Orbifold geometries based on* $\mathbb{Z}_{12}$

Figure D.2: *For the* dynamic hidden E_8 *search, we display the probability* $\hat{p}^c$*(MSSM-like) (estimated on the samples) to find MSSM-like models under the constraint of the respective contrast pattern* c *as a function of the threshold value* X*, where* c *is* $N_{\mathrm{ur}}^{(2)} \geq X$.

D.4 Vector-like exotics in the heterotic orbifold landscape

	$\mathbb{Z}_4$	$\mathbb{Z}_6$-I	$\mathbb{Z}_6$-II	$\mathbb{Z}_7$	$\mathbb{Z}_8$-I	$\mathbb{Z}_8$-II	$\mathbb{Z}_{12}$-I	$\mathbb{Z}_{12}$-II	$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\mathbb{Z}_2 \times \mathbb{Z}_4$	$\mathbb{Z}_2 \times \mathbb{Z}_6$-I	$\mathbb{Z}_3 \times \mathbb{Z}_3$	$\mathbb{Z}_3 \times \mathbb{Z}_6$	$\mathbb{Z}_4 \times \mathbb{Z}_4$	$\mathbb{Z}_6 \times \mathbb{Z}_6$
# MSSM	212	62	1,870	1	865	2,844	1,250	435	1,711	55,429	1,095	3,337	5,153	48,812	3,707
$(\mathbf{3},\mathbf{2})_{1/6}$	1.89	0	31.60	0	4.05	25.00	4.00	24.14	8.53	14.04	31.60	14.59	19.04	15.31	39.60
$(\bar{\mathbf{3}},\mathbf{1})_{-2/3}$	50.47	0	28.66	0	5.55	44.13	3.28	35.63	8.36	14.35	22.92	18.91	18.84	14.93	29.92
$(\bar{\mathbf{3}},\mathbf{1})_{1/3}$	100	100	99.95	100	99.54	100	100	100	100	98.19	99.73	94.76	99.83	99.03	99.00
$(\mathbf{1},\mathbf{2})_{-1/2}$	96.23	35.48	92.19	100	93.99	94.94	78.96	91.72	92.17	92.20	96.26	90.62	99.20	95.07	95.50
$(\mathbf{1},\mathbf{1})_{1}$	1.89	0	28.66	0	5.55	44.09	3.28	35.63	8.36	14.36	22.92	11.30	18.92	14.98	29.92
$(\mathbf{1},\mathbf{1})_{0}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$(\mathbf{3},\mathbf{2})_{-1/2}$	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0	0
$(\mathbf{3},\mathbf{2})_{1/2}$	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0	0
$(\mathbf{3},\mathbf{2})_{-1/3}$	0	0	0	0	0.23	1.05	0.24	0	0	2.27	0.09	0	0.43	0.53	0.32
$(\mathbf{3},\mathbf{2})_{-1/6}$	0	0	0	0	0	0	0.64	0	0	0	0	7.19	0.02	0	0
$(\mathbf{3},\mathbf{2})_{-1/12}$	2.83	0	0	0	0	0	0	0	0	0.00	0	0	0	0.00	0
$(\mathbf{3},\mathbf{1})_{0}$	0	0	6.84	0	0	0	25.12	0	0	0	0	85.77	9.99	0	0.35
$(\mathbf{3},\mathbf{1})_{-1/2}$	0	0	0	0	0	0	1.12	0	0	0	0	0	0.06	0	0
$(\mathbf{3},\mathbf{1})_{1/2}$	0	0	0	0	0	0	0.80	0	0	0	0	0	0.04	0	0
$(\mathbf{3},\mathbf{1})_{-2/3}$	0	0	0	0	0	0	0.32	0	0	0	0	5.60	0.06	0	0
$(\mathbf{3},\mathbf{1})_{1/3}$	0	0	4.71	0	0	0	18.72	0	0	0	0	73.24	6.85	0	0.27
$(\mathbf{3},\mathbf{1})_{-5/6}$	0	0	2.89	0	0	0.74	0	0	0	0.12	0	0	0.62	0.44	0.76
$(\mathbf{3},\mathbf{1})_{-1/6}$	0	0	0.21	0	0	0	4.32	0	0	0	0	0	0.45	0	0
$(\mathbf{3},\mathbf{1})_{1/6}$	67.45	93.55	66.47	0	87.63	74.30	69.20	66.90	85.21	90.85	86.58	0	84.84	95.75	98.17
$(\mathbf{3},\mathbf{1})_{-7/12}$	2.83	0	0	0	0	0	0.32	0	0	0	0	0	0	0.01	0
$(\mathbf{3},\mathbf{1})_{-1/12}$	41.04	0	0	0	0.23	0.56	1.76	0	0	0.08	0	0	0	0.13	0
$(\mathbf{3},\mathbf{1})_{5/12}$	4.72	0	0	0	0	0.14	0.80	0	0	0.03	0	0	0	0.06	0
$(\mathbf{3},\mathbf{1})_{2/21}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{3},\mathbf{1})_{5/21}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{2})_{0}$	82.55	100	85.94	0	99.31	93.95	76.32	90.34	84.57	96.49	98.17	0	89.93	99.26	99.57
$(\mathbf{1},\mathbf{2})_{1/3}$	0	0	0.43	0	0	0	4.72	0	0	0	0	0	0.66	0	0
$(\mathbf{1},\mathbf{2})_{2/3}$	0	0	0	0	0	0	0.80	0	0	0	0	0	0.02	0	0
$(\mathbf{1},\mathbf{2})_{1/4}$	41.98	0	0	0	0.23	0.70	2.40	0	0	0.08	0	0	0	0.14	0
$(\mathbf{1},\mathbf{2})_{3/4}$	2.83	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{2})_{1/6}$	0	0	8.98	0	0	0	26.72	0	0	0	0	96.55	10.58	0	0.35
$(\mathbf{1},\mathbf{2})_{5/6}$	0	0	0	0	0	0	0	0	0	0	0	0.21	0	0	0
$(\mathbf{1},\mathbf{2})_{1/14}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{2})_{5/14}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{1/2}$	100	100	92.09	0	100	100	76.96	100	95.79	99.16	99.36	0	90.28	99.99	99.65
$(\mathbf{1},\mathbf{1})_{1/3}$	0	0	8.98	0	0	0	29.20	0	0	0	0	99.52	10.58	0	0.35
$(\mathbf{1},\mathbf{1})_{2/3}$	0	0	8.98	0	0	0	28.72	0	0	0	0	95.62	10.58	0	0.35
$(\mathbf{1},\mathbf{1})_{1/4}$	52.36	0	0	0	0.23	0.95	3.20	0	0	0.19	0	0	0	0.17	0
$(\mathbf{1},\mathbf{1})_{3/4}$	42.92	0	0	0	0.23	0.42	1.92	0	0	0.12	0	0	0	0.14	0
$(\mathbf{1},\mathbf{1})_{1/6}$	0	0	1.07	0	0	0	6.32	0	0	0	0	0	0.85	0	0
$(\mathbf{1},\mathbf{1})_{5/6}$	0	0	0	0	0	0	1.12	0	0	0	0	0	0.04	0	0
$(\mathbf{1},\mathbf{1})_{1/7}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{2/7}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{3/7}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{4/7}$	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0

Table D.1: *Percentage of MSSM-like orbifold models with certain types of vector-like exotics. Hypercharge is normalized such that* $(\mathbf{3},\mathbf{2})_{1/6}$ *is a left-chiral quark-doublet. The row "# MSSM" lists the number of inequivalent MSSM-like orbifold models in our dataset. A complex representation has to be amended by its complex conjugate, e.g.* $(\mathbf{3},\mathbf{2})_{1/6}$ *stands for* $(\mathbf{3},\mathbf{2})_{1/6} \oplus (\bar{\mathbf{3}},\mathbf{2})_{-1/6}$.

	$\mathbb{Z}_4$	$\mathbb{Z}_6$-I	$\mathbb{Z}_6$-II	$\mathbb{Z}_7$	$\mathbb{Z}_8$-I	$\mathbb{Z}_8$-II	$\mathbb{Z}_{12}$-I	$\mathbb{Z}_{12}$-II	$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\mathbb{Z}_2 \times \mathbb{Z}_4$	$\mathbb{Z}_2 \times \mathbb{Z}_6$-I	$\mathbb{Z}_3 \times \mathbb{Z}_3$	$\mathbb{Z}_3 \times \mathbb{Z}_6$	$\mathbb{Z}_4 \times \mathbb{Z}_4$	$\mathbb{Z}_6 \times \mathbb{Z}_6$
# MSSM	212	62	1,870	1	865	2,844	1,250	435	1,711	55,429	1,095	3,337	5,153	48,812	3,707
$(\mathbf{3},\mathbf{2})_{1/6}$	0.04	0	0.48	0	0.05	0.31	0.09	0.46	0.09	0.19	0.51	0.17	0.23	0.18	0.46
$(\bar{\mathbf{3}},\mathbf{1})_{-2/3}$	0.52	0	0.41	0	0.08	0.49	0.06	0.71	0.09	0.21	0.35	0.23	0.23	0.18	0.36
$(\bar{\mathbf{3}},\mathbf{1})_{1/3}$	3.52	3.52	5.22	4.00	4.41	4.36	4.21	7.58	7.31	5.66	7.96	3.64	6.57	5.72	6.25
$(\mathbf{1},\mathbf{2})_{-1/2}$	2.64	1.68	4.03	3.00	3.41	3.06	2.92	4.28	3.34	3.97	5.23	2.86	5.83	4.46	5.50
$(\mathbf{1},\mathbf{1})_{1}$	0.04	0	0.41	0	0.08	0.49	0.06	0.71	0.09	0.21	0.35	0.15	0.23	0.18	0.36
$(\mathbf{1},\mathbf{1})_{0}$	67.37	136.58	106.06	32.00	108.74	89.78	85.99	101.78	122.54	132.64	139.59	106.48	177.56	166.38	186.95
$(\mathbf{3},\mathbf{2})_{-1/2}$	0	0	0	0	0	0	0	0	0	0	0	0.00	0	0	0
$(\mathbf{3},\mathbf{2})_{1/2}$	0	0	0	0	0	0	0	0	0	0	0	0.00	0	0	0
$(\mathbf{3},\mathbf{2})_{-1/3}$	0	0	0	0	0.00	0.01	0.00	0	0	0.03	0.00	0	0.01	0.01	0.00
$(\mathbf{3},\mathbf{2})_{-1/6}$	0	0	0	0	0	0	0.01	0	0	0	0	0.08	0.00	0	0
$(\mathbf{3},\mathbf{2})_{-1/12}$	0.03	0	0	0	0	0	0	0	0	0.00	0	0	0	0.00	0
$(\mathbf{3},\mathbf{1})_{0}$	0	0	0.19	0	0	0	0.69	0	0	0	0	3.32	0.36	0	0.01
$(\mathbf{3},\mathbf{1})_{-1/2}$	0	0	0	0	0	0	0.01	0	0	0	0	0	0.00	0	0
$(\mathbf{3},\mathbf{1})_{1/2}$	0	0	0	0	0	0	0.01	0	0	0	0	0	0.00	0	0
$(\mathbf{3},\mathbf{1})_{-2/3}$	0	0	0	0	0	0	0.00	0	0	0	0	0.06	0.00	0	0
$(\mathbf{3},\mathbf{1})_{1/3}$	0	0	0.13	0	0	0	0.36	0	0	0	0	1.95	0.14	0	0.01
$(\mathbf{3},\mathbf{1})_{-5/6}$	0	0	0.04	0	0	0.01	0	0	0	0.00	0	0	0.01	0.01	0.01
$(\mathbf{3},\mathbf{1})_{-1/6}$	0	0	0.00	0	0	0	0.06	0	0	0	0	0	0.01	0	0
$(\mathbf{3},\mathbf{1})_{1/6}$	1.99	3.29	2.38	0	3.14	2.55	2.00	1.69	3.83	4.59	4.87	0	3.47	5.37	6.57
$(\mathbf{3},\mathbf{1})_{-7/12}$	0.03	0	0	0	0	0	0.00	0	0	0	0	0	0	0.00	0
$(\mathbf{3},\mathbf{1})_{-1/12}$	1.10	0	0	0	0.00	0.01	0.04	0	0	0.00	0	0	0	0.00	0
$(\mathbf{3},\mathbf{1})_{5/12}$	0.08	0	0	0	0	0.00	0.01	0	0	0.00	0	0	0	0.00	0
$(\mathbf{3},\mathbf{1})_{2/21}$	0	0	0	3.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{3},\mathbf{1})_{5/21}$	0	0	0	4.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{2})_{0}$	7.78	8.58	7.99	0	9.49	8.23	5.84	6.10	7.62	13.58	14.25	0	10.97	15.88	19.27
$(\mathbf{1},\mathbf{2})_{1/3}$	0	0	0.01	0	0	0	0.08	0	0	0	0	0	0.01	0	0
$(\mathbf{1},\mathbf{2})_{2/3}$	0	0	0	0	0	0	0.01	0	0	0	0	0	0.00	0	0
$(\mathbf{1},\mathbf{2})_{1/4}$	1.19	0	0	0	0.00	0.02	0.07	0	0	0.00	0	0	0	0.01	0
$(\mathbf{1},\mathbf{2})_{3/4}$	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{2})_{1/6}$	0	0	0.44	0	0	0	1.18	0	0	0	0	7.75	0.81	0	0.02
$(\mathbf{1},\mathbf{2})_{5/6}$	0	0	0	0	0	0	0	0	0	0	0	0.00	0	0	0
$(\mathbf{1},\mathbf{2})_{1/14}$	0	0	0	5.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{2})_{5/14}$	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{1/2}$	18.79	23.81	14.60	0	23.19	20.92	12.45	13.59	18.80	29.37	29.20	0	25.33	37.34	41.21
$(\mathbf{1},\mathbf{1})_{1/3}$	0	0	3.14	0	0	0	7.72	0	0	0	0	42.99	5.22	0	0.23
$(\mathbf{1},\mathbf{1})_{2/3}$	0	0	0.43	0	0	0	1.25	0	0	0	0	8.30	0.84	0	0.03
$(\mathbf{1},\mathbf{1})_{1/4}$	14.90	0	0	0	0.07	0.17	0.68	0	0	0.03	0	0	0	0.10	0
$(\mathbf{1},\mathbf{1})_{3/4}$	1.18	0	0	0	0.00	0.01	0.04	0	0	0.00	0	0	0	0.01	0
$(\mathbf{1},\mathbf{1})_{1/6}$	0	0	0.13	0	0	0	0.63	0	0	0	0	0	0.17	0	0
$(\mathbf{1},\mathbf{1})_{5/6}$	0	0	0	0	0	0	0.01	0	0	0	0	0	0.00	0	0
$(\mathbf{1},\mathbf{1})_{1/7}$	0	0	0	22.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{2/7}$	0	0	0	16.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{3/7}$	0	0	0	11.00	0	0	0	0	0	0	0	0	0	0	0
$(\mathbf{1},\mathbf{1})_{4/7}$	0	0	0	9.00	0	0	0	0	0	0	0	0	0	0	0

Table D.2: *Average numbers of vector-like exotics for MSSM-like orbifold models. Hypercharge is normalized such that $(\mathbf{3},\mathbf{2})_{1/6}$ is a left-chiral quark-doublet. The row "# MSSM" lists the number of inequivalent MSSM-like orbifold models in our dataset. A complex representation has to be amended by its complex conjugate, e.g. $(\mathbf{3},\mathbf{2})_{1/6}$ stands for $(\mathbf{3},\mathbf{2})_{1/6} \oplus (\bar{\mathbf{3}},\mathbf{2})_{-1/6}$.*

Physics bibliography

[1] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.*, vol. B716, pp. 1–29, 2012. DOI: `10.1016/j.physletb.2012.08.020`. arXiv: `1207.7214 [hep-ex]`.

[2] S. Chatrchyan *et al.*, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC,” *Phys. Lett.*, vol. B716, pp. 30–61, 2012. DOI: `10.1016/j.physletb.2012.08.021`. arXiv: `1207.7235 [hep-ex]`.

[3] B. P. Abbott *et al.*, “Observation of Gravitational Waves from a Binary Black Hole Merger,” *Phys. Rev. Lett.*, vol. 116, no. 6, p. 061 102, 2016. DOI: `10.1103/PhysRevLett.116.061102`. arXiv: `1602.03837 [gr-qc]`.

[4] B. P. Abbott *et al.*, “GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence,” *Phys. Rev. Lett.*, vol. 116, no. 24, p. 241 103, 2016. DOI: `10.1103/PhysRevLett.116.241103`. arXiv: `1606.04855 [gr-qc]`.

[5] B. P. Abbott *et al.*, “GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2,” *Phys. Rev. Lett.*, vol. 118, no. 22, p. 221 101, 2017, [Erratum: Phys. Rev. Lett.121,no.12,129901(2018)]. DOI: `10.1103/PhysRevLett.118.221101,10.1103/PhysRevLett.121.129901`. arXiv: `1706.01812 [gr-qc]`.

[6] B. P. Abbott *et al.*, “GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence,” *Astrophys. J.*, vol. 851, no. 2, p. L35, 2017. DOI: `10.3847/2041-8213/aa9f0c`. arXiv: `1711.05578 [astro-ph.HE]`.

[7] B. P. Abbott *et al.*, “GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence,” *Phys. Rev. Lett.*, vol. 119, no. 14, p. 141 101, 2017. DOI: `10.1103/PhysRevLett.119.141101`. arXiv: `1709.09660 [gr-qc]`.

[8] B. P. Abbott *et al.*, “GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral,” *Phys. Rev. Lett.*, vol. 119, no. 16, p. 161 101, 2017. DOI: `10.1103/PhysRevLett.119.161101`. arXiv: `1710.05832 [gr-qc]`.

[9] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964. DOI: `10.1103/PhysRevLett.13.508`.

[10] P. W. Higgs, “Broken symmetries, massless particles and gauge fields,” *Phys. Lett.*, vol. 12, pp. 132–133, 1964. DOI: `10.1016/0031-9163(64)91136-9`.

[11] S. L. Glashow, “Partial Symmetries of Weak Interactions,” *Nucl. Phys.*, vol. 22, pp. 579–588, 1961. DOI: `10.1016/0029-5582(61)90469-2`.

[12] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. DOI: `10.1103/PhysRevLett.19.1264`.

[13] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc.*, vol. C680519, pp. 367–377, 1968. DOI: `10.1142/9789812795915_0034`.

[14] A. Einstein, "The Foundation of the General Theory of Relativity," *Annalen Phys.*, vol. 49, no. 7, pp. 769–822, 1916. DOI: `10.1002/andp.200590044, 10.1002/andp.19163540702`.

[15] J. F. Donoghue, E. Golowich, and B. R. Holstein, "Dynamics of the standard model," *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, vol. 2, pp. 1–540, 1992, [Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.35(2014)]. DOI: `10.1017/CBO9780511524370`.

[16] A. Eichhorn, "An asymptotically safe guide to quantum gravity and matter," *Front. Astron. Space Sci.*, vol. 5, p. 47, 2019. DOI: `10.3389/fspas.2018.00047`. arXiv: `1810.07615 [hep-th]`.

[17] C. Rovelli, "Loop quantum gravity," *Living Rev. Rel.*, vol. 1, p. 1, 1998. DOI: `10.12942/lrr-1998-1`. arXiv: `gr-qc/9710008 [gr-qc]`.

[18] W. Lerche, D. Lüst, and A. N. Schellekens, "Chiral Four-Dimensional Heterotic Strings from Selfdual Lattices," *Nucl. Phys.*, vol. B287, p. 477, 1987. DOI: `10.1016/0550-3213(87)90115-5`.

[19] *GAP – Groups, Algorithms, and Programming, Version 4.8.8*, The GAP Group, 2017.

[20] E. Parr, P. K. Vaudrevange, and M. Wimmer, "Predicting the orbifold origin of the MSSM," *Fortsch. Phys.*, vol. 68, no. 5, p. 2 000 032, 2020. DOI: `10.1002/prop.202000032`. arXiv: `2003.01732 [hep-th]`.

[21] E. Parr and P. K. S. Vaudrevange, "Contrast data mining for the MSSM from strings," *Nucl. Phys.*, vol. B952, p. 114 922, 2020. DOI: `10.1016/j.nuclphysb.2020.114922`. arXiv: `1910.13473 [hep-th]`.

[22] S. Biermann, A. Mütter, E. Parr, M. Ratz, and P. K. S. Vaudrevange, "Discrete remnants of orbifolding," *Phys. Rev.*, vol. D100, no. 6, p. 066 030, 2019. DOI: `10.1103/PhysRevD.100.066030`. arXiv: `1906.10276 [hep-ph]`.

[23] A. Mütter, E. Parr, and P. K. S. Vaudrevange, "Deep learning in the heterotic orbifold landscape," *Nucl. Phys.*, vol. B940, pp. 113–129, 2019. DOI: `10.1016/j.nuclphysb.2019.01.013`. arXiv: `1811.05993 [hep-th]`.

[24] S. Groot Nibbelink, O. Loukas, A. Mütter, E. Parr, and P. K. S. Vaudrevange, "Tension Between a Vanishing Cosmological Constant and Non-Supersymmetric Heterotic Orbifolds," 2017. arXiv: `1710.09237 [hep-th]`.

[25] K. Becker, M. Becker, and J. Schwarz, "String theory and M-theory: A modern introduction," 2007.

[26] R. Blumenhagen, D. Lüst, and S. Theisen, *Basic concepts of string theory*, ser. Theoretical and Mathematical Physics. Heidelberg, Germany: Springer, 2013, ISBN: 9783642294969. DOI: `10.1007/978-3-642-29497-6`.

[27] M. B. Green, J. H. Schwarz, and E. Witten, "Superstring theory. vol. 1: Introduction," Cambridge, Uk: Univ. Pr. (1987).

[28] M. B. Green, J. H. Schwarz, and E. Witten, "Superstring theory. vol. 2: Loop amplitudes, anomalies and phenomenology," Cambridge, Uk: Univ. Pr. (1987).

[29] L. E. Ibanez and A. M. Uranga, *String theory and particle physics: An introduction to string phenomenology.* Cambridge University Press, 2012, ISBN: 9781139227421.

[30] K.-S. Choi and J. E. Kim, "Quarks and leptons from orbifolded superstring," *Lect. Notes Phys.*, vol. 696, pp. 1–406, 2006. DOI: `10.1007/b11681670`.

[31] S. Ramos-Sanchez, "Towards Low Energy Physics from the Heterotic String," *Fortsch. Phys.*, vol. 10, pp. 907–1036, 2009. DOI: `10.1002/prop.200900073`. arXiv: `0812.3560 [hep-th]`.

[32] P. K. S. Vaudrevange, "Grand Unification in the Heterotic Brane World," PhD thesis, Bonn U., 2008. arXiv: `0812.3503 [hep-th]`.

[33] J. Scherk and J. H. Schwarz, "Dual Models for Nonhadrons," *Nucl. Phys.*, vol. B81, pp. 118–144, 1974. DOI: `10.1016/0550-3213(74)90010-8`.

[34] J. Scherk and J. H. Schwarz, "Dual Models and the Geometry of Space-Time," *Phys. Lett.*, vol. 52B, pp. 347–350, 1974. DOI: `10.1016/0370-2693(74)90059-8`.

[35] D. Lust and S. Theisen, "Lectures on string theory," *Lect. Notes Phys.*, vol. 346, pp. 1–346, 1989. DOI: `10.1007/BFb0113507`.

[36] F. Gliozzi, J. Scherk, and D. I. Olive, "Supersymmetry, Supergravity Theories and the Dual Spinor Model," *Nucl. Phys.*, vol. B122, pp. 253–290, 1977. DOI: `10.1016/0550-3213(77)90206-1`.

[37] Y. Kawamura, "Gauge symmetry reduction from the extra space s(1)/z(2)," *Prog. Theor. Phys.*, vol. 103, pp. 613–619, 2000. eprint: `hep-ph/9902423`.

[38] R. Barbieri, L. J. Hall, and Y. Nomura, "A constrained standard model from a compact extra dimension," *Phys. Rev.*, vol. D63, p. 105 007, 2001. eprint: `hep-ph/0011311`.

[39] Y. Kawamura, "Triplet-doublet splitting, proton stability and extra dimension," *Prog. Theor. Phys.*, vol. 105, pp. 999–1006, 2001. eprint: `hep-ph/0012125`.

[40] G. Altarelli and F. Feruglio, "Su(5) grand unification in extra dimensions and proton decay," *Phys. Lett.*, vol. B511, pp. 257–264, 2001. eprint: `hep-ph/0102301`.

[41] A. Hebecker and J. March-Russell, "A minimal s(1)/(z(2) x z'(2)) orbifold gut," *Nucl. Phys.*, vol. B613, pp. 3–16, 2001. eprint: `hep-ph/0106166`.

[42] A. Hebecker and J. March-Russell, "The structure of GUT breaking by orbifolding," *Nucl. Phys.*, vol. B625, pp. 128–150, 2002. eprint: `hep-ph/0107039`.

[43] T. Asaka, W. Buchmüller, and L. Covi, "Gauge unification in six dimensions," *Phys. Lett.*, vol. B523, pp. 199–204, 2001. eprint: `hep-ph/0108021`.

[44] L. J. Dixon, J. A. Harvey, C. Vafa, and E. Witten, "Strings on Orbifolds. 2.," *Nucl. Phys.*, vol. B274, pp. 285–314, 1986. DOI: `10.1016/0550-3213(86)90287-7`.

[45] L. J. Dixon, J. A. Harvey, C. Vafa, and E. Witten, "Strings on Orbifolds," *Nucl. Phys.*, vol. B261, pp. 678–686, 1985, [,678(1985)]. DOI: `10.1016/0550-3213(85)90593-0`.

[46] A. E. Faraggi, "Construction of realistic standard - like models in the free fermionic superstring formulation," *Nucl. Phys.*, vol. B387, pp. 239–262, 1992. DOI: `10.1016/0550-3213(92)90160-D`. arXiv: `hep-th/9208024 [hep-th]`.

[47] T. P. T. Dijkstra, L. R. Huiszoon, and A. N. Schellekens, "Supersymmetric standard model spectra from RCFT orientifolds," *Nucl. Phys.*, vol. B710, pp. 3–57, 2005. DOI: `10.1016/j.nuclphysb.2004.12.032`. arXiv: `hep-th/0411129 [hep-th]`.

[48] V. Braun, Y.-H. He, B. A. Ovrut, and T. Pantev, "A Heterotic standard model," *Phys. Lett.*, vol. B618, pp. 252–258, 2005. DOI: `10.1016/j.physletb.2005.05.007`. arXiv: `hep-th/0501070 [hep-th]`.

[49] F. Gmeiner, R. Blumenhagen, G. Honecker, D. Lüst, and T. Weigand, “One in a billion: MSSM-like D-brane statistics,” *JHEP*, vol. 01, p. 004, 2006. doi: `10.1088/1126-6708/2006/01/004`. arXiv: `hep-th/0510170 [hep-th]`.

[50] K. R. Dienes, “Statistics on the heterotic landscape: Gauge groups and cosmological constants of four-dimensional heterotic strings,” *Phys. Rev.*, vol. D73, p. 106 010, 2006. doi: `10.1103/PhysRevD.73.106010`. arXiv: `hep-th/0602286 [hep-th]`.

[51] R. Blumenhagen, B. Körs, D. Lüst, and S. Stieberger, “Four-dimensional String Compactifications with D-Branes, Orientifolds and Fluxes,” *Phys. Rept.*, vol. 445, pp. 1–193, 2007. doi: `10.1016/j.physrep.2007.04.003`. arXiv: `hep-th/0610327 [hep-th]`.

[52] L. B. Anderson, J. Gray, A. Lukas, and E. Palti, “Two Hundred Heterotic Standard Models on Smooth Calabi-Yau Threefolds,” *Phys. Rev.*, vol. D84, p. 106 005, 2011. doi: `10.1103/PhysRevD.84.106005`. arXiv: `1106.4804 [hep-th]`.

[53] L. B. Anderson, J. Gray, A. Lukas, and E. Palti, “Heterotic Line Bundle Standard Models,” *JHEP*, vol. 06, p. 113, 2012. doi: `10.1007/JHEP06(2012)113`. arXiv: `1202.1757 [hep-th]`.

[54] M. Cvetič, D. Klevers, D. K. Mayorga Peña, P.-K. Oehlmann, and J. Reuter, “Three-Family Particle Physics Models from Global F-theory Compactifications,” *JHEP*, vol. 08, p. 087, 2015. doi: `10.1007/JHEP08(2015)087`. arXiv: `1503.02068 [hep-th]`.

[55] M. Cvetič, L. Lin, M. Liu, and P.-K. Oehlmann, “An F-theory Realization of the Chiral MSSM with $\mathbb{Z}_2$-Parity,” *JHEP*, vol. 09, p. 089, 2018. doi: `10.1007/JHEP09(2018)089`. arXiv: `1807.01320 [hep-th]`.

[56] M. Cvetič, J. Halverson, L. Lin, M. Liu, and J. Tian, “Quadrillion F-Theory Compactifications with the Exact Chiral Spectrum of the Standard Model,” *Phys. Rev. Lett.*, vol. 123, no. 10, p. 101 601, 2019. doi: `10.1103/PhysRevLett.123.101601`. arXiv: `1903.00009 [hep-th]`.

[57] M. Fischer, M. Ratz, J. Torrado, and P. K. S. Vaudrevange, “Classification of symmetric toroidal orbifolds,” *JHEP*, vol. 01, p. 084, 2013. doi: `10.1007/JHEP01(2013)084`. arXiv: `1209.3906 [hep-th]`.

[58] J. Opgenorth, W. Plesken, and T. Schulz, “Crystallographic Algorithms and Tables,” *Acta Crystallographica Section A*, vol. 54, no. 5, pp. 517–531, Sep. 1998. doi: `10.1107/S010876739701547X`.

[59] W. Plesken and T. Schulz, “Counting crystallographic groups in low dimensions,” *Experimental Mathematics*, vol. 9, no. 3, pp. 407–411, 2000. doi: `10.1080/10586458.2000.10504417`. eprint: `http://dx.doi.org/10.1080/10586458.2000.10504417`.

[60] D. Bailin, A. Love, and S. Thomas, “Orbifold compactified superstring models with tetrahedral and octahedral point groups,” *Nucl.Phys.B*, vol. 298, pp. 75–91, 1988. doi: `10.1016/0550-3213(88)90304-5`.

[61] Y. Katsuki, Y. Kawamura, T. Kobayashi, N. Ohtsubo, Y. Ono, and K. Tanioka, “Z_N orbifold models,” *Nucl. Phys.*, vol. B341, pp. 611–640, 1990.

[62] A. Font, L. E. Ibàñez, and F. Quevedo, “Z(n) x z(m) orbifolds and discrete torsion,” *Phys. Lett.*, vol. B217, p. 272, 1989.

[63] L. E. Ibáñez, H. P. Nilles, and F. Quevedo, “Orbifolds and wilson lines,” *Phys. Lett.*, vol. B187, pp. 25–32, 1987.

[64] S. Ramos-Sánchez and P. K. S. Vaudrevange, "Note on the space group selection rule for closed strings on orbifolds," *JHEP*, vol. 01, p. 055, 2019. doi: `10.1007/JHEP01(2019)055`. arXiv: `1811.00580 [hep-th]`.

[65] T. Kobayashi and N. Ohtsubo, "Geometrical aspects of Z_N orbifold phenomenology," *Int. J. Mod. Phys.*, vol. A9, pp. 87–126, 1994.

[66] F. Plöger, S. Ramos-Sánchez, M. Ratz, and P. K. S. Vaudrevange, "Mirage Torsion," *JHEP*, vol. 04, p. 063, 2007. doi: `10.1088/1126-6708/2007/04/063`. arXiv: `hep-th/0702176 [hep-th]`.

[67] T. Kobayashi, S. Raby, and R.-J. Zhang, "Searching for realistic 4d string models with a pati-salam symmetry: Orbifold grand unified theories from heterotic string compactification on a Z(6) orbifold," *Nucl. Phys.*, vol. B704, pp. 3–55, 2005. eprint: `hep-ph/0409098`.

[68] S. Förste, H. P. Nilles, P. K. S. Vaudrevange, and A. Wingerter, "Heterotic brane world," *Phys. Rev.*, vol. D70, p. 106 008, 2004. doi: `10.1103/PhysRevD.70.106008`. arXiv: `hep-th/0406208 [hep-th]`.

[69] T. Kobayashi, S. Raby, and R.-J. Zhang, "Constructing 5d orbifold grand unified theories from heterotic strings," *Phys. Lett.*, vol. B593, pp. 262–270, 2004. eprint: `hep-ph/0403065`.

[70] W. Buchmüller, K. Hamaguchi, O. Lebedev, and M. Ratz, "Dual models of gauge unification in various dimensions," *Nucl. Phys.*, vol. B712, pp. 139–156, 2005. doi: `10.1016/j.nuclphysb.2005.01.038`. arXiv: `hep-ph/0412318 [hep-ph]`.

[71] H. P. Nilles, "Five golden rules for superstring phenomenology," 2004. eprint: `hep-th/0410160`.

[72] W. Buchmüller, K. Hamaguchi, O. Lebedev, and M. Ratz, "Supersymmetric standard model from the heterotic string," *Phys. Rev. Lett.*, vol. 96, p. 121 602, 2006. eprint: `hep-ph/0511035`.

[73] W. Buchmüller, K. Hamaguchi, O. Lebedev, and M. Ratz, "Local grand unification," in *Symposium GustavoFest*, 2005, pp. 143–156. arXiv: `hep-ph/0512326 [hep-ph]`.

[74] W. Buchmuller, K. Hamaguchi, O. Lebedev, and M. Ratz, "Supersymmetric Standard Model from the Heterotic String (II)," *Nucl. Phys. B*, vol. 785, pp. 149–209, 2007. doi: `10.1016/j.nuclphysb.2007.06.028`. arXiv: `hep-th/0606187`.

[75] O. Lebedev, H. P. Nilles, S. Raby, S. Ramos-Sánchez, M. Ratz, P. K. S. Vaudrevange, and A. Wingerter, "A Mini-landscape of exact MSSM spectra in heterotic orbifolds," *Phys. Lett.*, vol. B645, pp. 88–94, 2007. doi: `10.1016/j.physletb.2006.12.012`. arXiv: `hep-th/0611095 [hep-th]`.

[76] O. Lebedev, H.-P. Nilles, S. Raby, S. Ramos-Sánchez, M. Ratz, P. K. S. Vaudrevange, and A. Wingerter, "Low Energy Supersymmetry from the Heterotic Landscape," *Phys. Rev. Lett.*, vol. 98, p. 181 602, 2007. doi: `10.1103/PhysRevLett.98.181602`. arXiv: `hep-th/0611203 [hep-th]`.

[77] O. Lebedev, H. P. Nilles, S. Raby, S. Ramos-Sánchez, M. Ratz, and A. Vaudrevange Patrick K. S. and Wingerter, "The heterotic road to the MSSM with r parity," *Phys. Rev.*, vol. D77, p. 046 013, 2007. eprint: `arXiv:0708.2691[hep-th]`.

[78] H. P. Nilles, S. Ramos-Sánchez, P. K. S. Vaudrevange, and A. Wingerter, "Exploring the SO(32) heterotic string," *JHEP*, vol. 04, p. 050, 2006. eprint: `hep-th/0603086`.

[79] O. Lebedev, H. P. Nilles, S. Ramos-Sánchez, M. Ratz, and P. K. S. Vaudrevange, "Heterotic mini-landscape. (II). Completing the search for MSSM vacua in a $\mathbb{Z}_6$ orbifold," *Phys. Lett.*, vol. B668, pp. 331–335, 2008. doi: `10.1016/j.physletb.2008.08.054`. arXiv: `0807.4384 [hep-th]`.

[80] H. P. Nilles, S. Ramos-Sánchez, P. K. S. Vaudrevange, and A. Wingerter, "The Orbifolder: A Tool to study the Low Energy Effective Theory of Heterotic Orbifolds," *Comput. Phys. Commun.*, vol. 183, pp. 1363–1380, 2012, web page http://projects.hepforge.org/orbifolder/. doi: `10.1016/j.cpc.2012.01.026`. arXiv: `1110.5229 [hep-th]`.

[81] L. J. Dixon and J. A. Harvey, "String theories in ten-dimensions without space-time supersymmetry," *Nucl.Phys.*, vol. B274, pp. 93–105, 1986. doi: `10.1016/0550-3213(86)90619-X`.

[82] L. Alvarez-Gaumé, P. H. Ginsparg, G. W. Moore, and C. Vafa, "An O(16) x O(16) heterotic string," *Phys.Lett.*, vol. B171, p. 155, 1986. doi: `10.1016/0370-2693(86)91524-8`.

[83] S. Groot Nibbelink and E. Parr, "Twisted superspace: Non-renormalization and fermionic symmetries in certain heterotic-string-inspired non-supersymmetric field theories," *Phys. Rev.*, vol. D94, no. 4, p. 041 704, 2016. doi: `10.1103/PhysRevD.94.041704`. arXiv: `1605.07470 [hep-ph]`.

[84] V. Nair, A. D. Shapere, A. Strominger, and F. Wilczek, "Compactification of the twisted heterotic string," *Nucl.Phys.*, vol. B287, p. 402, 1987. doi: `10.1016/0550-3213(87)90112-X`.

[85] P. H. Ginsparg and C. Vafa, "Toroidal compactification of nonsupersymmetric heterotic strings," *Nucl.Phys.*, vol. B289, p. 414, 1987. doi: `10.1016/0550-3213(87)90387-7`.

[86] W. Lerche, D. Lüst, and A. Schellekens, "Ten-dimensional heterotic strings from Niemeier lattices," *Phys.Lett.*, vol. B181, p. 71, 1986. doi: `10.1016/0370-2693(86)91257-8`.

[87] T. Taylor, "Model building on asymmetric Z(3) orbifolds: Nonsupersymmetric models," *Nucl.Phys.*, vol. B303, p. 543, 1988. doi: `10.1016/0550-3213(88)90393-8`.

[88] A. Toon, "Nonsupersymmetric Z(4) orbifolds and Atkin-Lehner symmetry," *Phys.Lett.*, vol. B243, pp. 68–72, 1990. doi: `10.1016/0370-2693(90)90958-9`.

[89] T. Sasada, "Asymmetric orbifold models of nonsupersymmetric heterotic strings," *Prog. Theor. Phys.*, vol. 95, pp. 249–257, 1996. doi: `10.1143/PTP.95.249`. arXiv: `hep-th/9508098 [hep-th]`.

[90] A. Font and A. Hernández, "Nonsupersymmetric orbifolds," *Nucl.Phys.*, vol. B634, pp. 51–70, 2002. doi: `10.1016/S0550-3213(02)00336-X`. arXiv: `hep-th/0202057 [hep-th]`.

[91] K. R. Dienes, "Modular invariance, finiteness, and misaligned supersymmetry: New constraints on the numbers of physical string states," *Nucl. Phys.*, vol. B429, pp. 533–588, 1994. doi: `10.1016/0550-3213(94)90153-8`. arXiv: `hep-th/9402006 [hep-th]`.

[92] J. D. Blum and K. R. Dienes, "Duality without supersymmetry: The case of the SO(16) x SO(16) string," *Phys.Lett.*, vol. B414, pp. 260–268, 1997. doi: `10.1016/S0370-2693(97)01172-6`. arXiv: `hep-th/9707148 [hep-th]`.

[93] G. Shiu and S. H. H. Tye, "Bose-Fermi degeneracy and duality in nonsupersymmetric strings," *Nucl. Phys.*, vol. B542, pp. 45–72, 1999. doi: `10.1016/S0550-3213(98)00775-5`. arXiv: `hep-th/9808095 [hep-th]`.

[94] A. E. Faraggi and M. Tsulaia, "On the Low Energy Spectra of the Nonsupersymmetric Heterotic String Theories," *Eur. Phys. J.*, vol. C54, pp. 495–500, 2008. DOI: `10.1140/epjc/s10052-008-0545-2`. arXiv: `0706.1649 [hep-th]`.

[95] M. Blaszczyk, S. Groot Nibbelink, O. Loukas, and S. Ramos-Sánchez, "Non-supersymmetric heterotic model building," *JHEP*, vol. 10, p. 119, 2014. DOI: `10.1007/JHEP10(2014)119`. arXiv: `1407.6362 [hep-th]`.

[96] M. Blaszczyk, S. Groot Nibbelink, O. Loukas, and F. Ruehle, "Calabi-Yau compactifications of non-supersymmetric heterotic string theory," *JHEP*, vol. 10, p. 166, 2015. DOI: `10.1007/JHEP10(2015)166`. arXiv: `1507.06147 [hep-th]`.

[97] S. G. Nibbelink, O. Loukas, and F. Ruehle, "(MS)SM-like models on smooth Calabi-Yau manifolds from all three heterotic string theories," *Fortsch. Phys.*, vol. 63, pp. 609–632, 2015. DOI: `10.1002/prop.201500041`. arXiv: `1507.07559 [hep-th]`.

[98] S. Groot Nibbelink, "Model building with the non-supersymmetric heterotic SO(16)xSO(16) string," *J. Phys. Conf. Ser.*, vol. 631, no. 1, p. 012 077, 2015. DOI: `10.1088/1742-6596/631/1/012077`. arXiv: `1502.03604 [hep-th]`.

[99] S. Abel, K. R. Dienes, and E. Mavroudi, "Towards a nonsupersymmetric string phenomenology," *Phys. Rev.*, vol. D91, no. 12, p. 126 014, 2015. DOI: `10.1103/PhysRevD.91.126014`. arXiv: `1502.03087 [hep-th]`.

[100] J. M. Ashfaque, P. Athanasopoulos, A. E. Faraggi, and H. Sonmez, "Non-Tachyonic Semi-Realistic Non-Supersymmetric Heterotic String Vacua," *Eur. Phys. J.*, vol. C76, no. 4, p. 208, 2016. DOI: `10.1140/epjc/s10052-016-4056-2`. arXiv: `1506.03114 [hep-th]`.

[101] M. Green, J. Schwarz, and E. Witten, *Superstring Theory: Volume 2, Loop Amplitudes, Anomalies and Phenomenology*, ser. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1988, ISBN: 9780521357531.

[102] M. Dine, *String theory in four dimensions*, ser. Current physics. North Holland, 1988, ISBN: 9780444871008.

[103] M. Aaboud *et al.*, "Search for supersymmetry in events with four or more leptons in $\sqrt{s} = 13$ TeV pp collisions with ATLAS," *Phys. Rev.*, vol. D98, no. 3, p. 032 009, 2018. DOI: `10.1103/PhysRevD.98.032009`. arXiv: `1804.03602 [hep-ex]`.

[104] H. Itoyama and T. Taylor, "Supersymmetry Restoration in the Compactified O(16) x O(16)-prime Heterotic String Theory," *Phys.Lett.*, vol. B186, p. 129, 1987. DOI: `10.1016/0370-2693(87)90267-X`.

[105] K. R. Dienes, "New string partition functions with vanishing cosmological constant," *Phys.Rev.Lett.*, vol. 65, pp. 1979–1982, 1990. DOI: `10.1103/PhysRevLett.65.1979`.

[106] A. E. Faraggi and M. Tsulaia, "Interpolations Among NAHE-based Supersymmetric and Nonsupersymmetric String Vacua," *Phys.Lett.*, vol. B683, pp. 314–320, 2010. DOI: `10.1016/j.physletb.2009.12.039`. arXiv: `0911.5125 [hep-th]`.

[107] I. Florakis and J. Rizos, "Chiral Heterotic Strings with Positive Cosmological Constant," *Nucl. Phys.*, vol. B913, pp. 495–533, 2016. DOI: `10.1016/j.nuclphysb.2016.09.018`. arXiv: `1608.04582 [hep-th]`.

[108] A. E. Faraggi, C. Kounnas, and H. Partouche, "Large volume susy breaking with a solution to the decompactification problem," *Nucl. Phys.*, vol. B899, pp. 328–374, 2015. DOI: `10.1016/j.nuclphysb.2015.08.001`. arXiv: `1410.6147 [hep-th]`.

[109] S. Kachru, J. Kumar, and E. Silverstein, "Vacuum energy cancellation in a nonsupersymmetric string," *Phys. Rev.*, vol. D59, p. 106 004, 1999. DOI: `10.1103/PhysRevD.59.106004`. arXiv: `hep-th/9807076 [hep-th]`.

[110] S. Kachru and E. Silverstein, "On vanishing two loop cosmological constants in nonsupersymmetric strings," *JHEP*, vol. 01, p. 004, 1999. DOI: `10.1088/1126-6708/1999/01/004`. arXiv: `hep-th/9810129 [hep-th]`.

[111] R. Blumenhagen and L. Gorlich, "Orientifolds of nonsupersymmetric asymmetric orbifolds," *Nucl. Phys.*, vol. B551, pp. 601–616, 1999. DOI: `10.1016/S0550-3213(99)00241-2`. arXiv: `hep-th/9812158 [hep-th]`.

[112] Y. Satoh, Y. Sugawara, and T. Wada, "Non-supersymmetric Asymmetric Orbifolds with Vanishing Cosmological Constant," *JHEP*, vol. 02, p. 184, 2016. DOI: `10.1007/JHEP02(2016)184`. arXiv: `1512.05155 [hep-th]`.

[113] R. Iengo and C.-J. Zhu, "Evidence for nonvanishing cosmological constant in nonSUSY superstring models," *JHEP*, vol. 04, p. 028, 2000. DOI: `10.1088/1126-6708/2000/04/028`. arXiv: `hep-th/9912074 [hep-th]`.

[114] G. Moore, "Atkin-lehner symmetry," *Nuclear Physics B*, vol. 293, no. Supplement C, pp. 139–188, 1987, ISSN: 0550-3213. DOI: `https://doi.org/10.1016/0550-3213(87)90067-8`.

[115] K. R. Dienes, "Generalized atkin-lehner symmetry," *Phys. Rev. D*, vol. 42, pp. 2004–2021, 6 Sep. 1990. DOI: `10.1103/PhysRevD.42.2004`.

[116] J. Balog and M. P. Tuite, "The Failure of Atkin-lehner Symmetry for Lattice Compactified Strings," *Nucl. Phys.*, vol. B319, pp. 387–414, 1989. DOI: `10.1016/0550-3213(89)90083-7`.

[117] T. Taylor, "Model building on asymmetric z3 orbifolds: Non-supersymmetric models," *Nuclear Physics B*, vol. 303, pp. 543–556, Jul. 1988. DOI: `10.1016/0550-3213(88)90393-8`.

[118] M. Fischer, S. Ramos-Sánchez, and P. K. S. Vaudrevange, "Heterotic non-Abelian orbifolds," *JHEP*, vol. 1307, p. 080, 2013. DOI: `10.1007/JHEP07(2013)080`. arXiv: `1304.7742 [hep-th]`.

[119] I. Isaacs, *Character Theory of Finite Groups*, ser. Dover books on advanced mathematics. Dover, 1994, ISBN: 9780486680149.

[120] T. van Ritbergen, A. N. Schellekens, and J. A. M. Vermaseren, "Group theory factors for feynman diagrams," *Int. J. Mod. Phys.*, vol. A14, pp. 41–96, 1999. eprint: `hep-ph/9802376`.

[121] S. Groot Nibbelink, H. P. Nilles, M. Olechowski, and M. G. A. Walter, "Localized tadpoles of anomalous heterotic u(1)'s," *Nucl. Phys.*, vol. B665, pp. 236–272, 2003. eprint: `hep-th/0303101`.

[122] H. Ishimori, T. Kobayashi, H. Ohki, Y. Shimizu, H. Okada, *et al.*, "Non-Abelian Discrete Symmetries in Particle Physics," *Prog.Theor.Phys.Suppl.*, vol. 183, pp. 1–163, 2010. DOI: `10.1143/PTPS.183.1`. arXiv: `1003.3552 [hep-th]`.

[123] S. J. Konopka, "Non Abelian orbifold compactifications of the heterotic string," *JHEP*, vol. 1307, p. 023, 2013. DOI: `10.1007/JHEP07(2013)023`. arXiv: `1210.5040 [hep-th]`.

[124] M. R. Douglas, "The statistics of string / m theory vacua," *JHEP*, vol. 05, p. 046, 2003. eprint: `hep-th/0303194`.

[125] D. K. Mayorga Peña, H. P. Nilles, and P.-K. Oehlmann, "A Zip-code for Quarks, Leptons and Higgs Bosons," *JHEP*, vol. 12, p. 024, 2012. DOI: `10.1007/JHEP12(2012)024`. arXiv: `1209.6041 [hep-th]`.

[126] H. P. Nilles and P. K. S. Vaudrevange, "Geography of Fields in Extra Dimensions: String Theory Lessons for Particle Physics," *Mod. Phys. Lett.*, vol. A30, no. 10, p. 1 530 008, 2015. DOI: `10.1142/S0217732315300086`. arXiv: `1403.1597 [hep-th]`.

[127] Y. Olguin-Trejo, R. Pérez-Martínez, and S. Ramos-Sánchez, "Charting the flavor landscape of MSSM-like Abelian heterotic orbifolds," *Phys. Rev.*, vol. D98, no. 10, p. 106 020, 2018. DOI: `10.1103/PhysRevD.98.106020`. arXiv: `1808.06622 [hep-th]`.

[128] S. Groot Nibbelink and O. Loukas, "MSSM-like models on $\mathbb{Z}_8$ toroidal orbifolds," *JHEP*, vol. 12, p. 044, 2013. DOI: `10.1007/JHEP12(2013)044`. arXiv: `1308.5145 [hep-th]`.

[129] S. Abel and J. Rizos, "Genetic Algorithms and the Search for Viable String Vacua," *JHEP*, vol. 08, p. 010, 2014. DOI: `10.1007/JHEP08(2014)010`. arXiv: `1404.7359 [hep-th]`.

[130] J. Carifio, W. J. Cunningham, J. Halverson, D. Krioukov, C. Long, and B. D. Nelson, "Vacuum Selection from Cosmology on Networks of String Geometries," *Phys. Rev. Lett.*, vol. 121, no. 10, p. 101 602, 2018. DOI: `10.1103/PhysRevLett.121.101602`. arXiv: `1711.06685 [hep-th]`.

[131] Y.-H. He, "Deep-Learning the Landscape," 2017. arXiv: `1706.02714 [hep-th]`.

[132] D. Krefl and R.-K. Seong, "Machine Learning of Calabi-Yau Volumes," *Phys. Rev.*, vol. D96, no. 6, p. 066 014, 2017. DOI: `10.1103/PhysRevD.96.066014`. arXiv: `1706.03346 [hep-th]`.

[133] F. Ruehle, "Evolving neural networks with genetic algorithms to study the String Landscape," *JHEP*, vol. 08, p. 038, 2017. DOI: `10.1007/JHEP08(2017)038`. arXiv: `1706.07024 [hep-th]`.

[134] J. Carifio, J. Halverson, D. Krioukov, and B. D. Nelson, "Machine Learning in the String Landscape," *JHEP*, vol. 09, p. 157, 2017. DOI: `10.1007/JHEP09(2017)157`. arXiv: `1707.00655 [hep-th]`.

[135] Y.-N. Wang and Z. Zhang, "Learning non-Higgsable gauge groups in 4D F-theory," *JHEP*, vol. 08, p. 009, 2018. DOI: `10.1007/JHEP08(2018)009`. arXiv: `1804.07296 [hep-th]`.

[136] K. Bull, Y.-H. He, V. Jejjala, and C. Mishra, "Machine Learning CICY Threefolds," *Phys. Lett.*, vol. B785, pp. 65–72, 2018. DOI: `10.1016/j.physletb.2018.08.008`. arXiv: `1806.03121 [hep-th]`.

[137] D. Klaewer and L. Schlechter, "Machine Learning Line Bundle Cohomologies of Hypersurfaces in Toric Varieties," *Phys. Lett.*, vol. B789, pp. 438–443, 2019. DOI: `10.1016/j.physletb.2019.01.002`. arXiv: `1809.02547 [hep-th]`.

[138] J. Halverson, B. Nelson, and F. Ruehle, "Branes with Brains: Exploring String Vacua with Deep Reinforcement Learning," *JHEP*, vol. 06, p. 003, 2019. DOI: `10.1007/JHEP06(2019)003`. arXiv: `1903.11616 [hep-th]`.

[139] A. Cole and G. Shiu, "Topological Data Analysis for the String Landscape," *JHEP*, vol. 03, p. 054, 2019. DOI: `10.1007/JHEP03(2019)054`. arXiv: `1812.06960 [hep-th]`.

[140] A. Cole, A. Schachner, and G. Shiu, "Searching the Landscape of Flux Vacua with Genetic Algorithms," 2019. arXiv: `1907.10072 [hep-th]`.

[141] K. Bull, Y.-H. He, V. Jejjala, and C. Mishra, "Getting CICY High," *Phys. Lett.*, vol. B795, pp. 700–706, 2019. DOI: 10.1016/j.physletb.2019.06.067. arXiv: 1903.03113 [hep-th].

[142] A. Ashmore, Y.-H. He, and B. Ovrut, "Machine learning calabi-yau metrics," 2019. arXiv: 1910.08605 [hep-th].

[143] T. Kobayashi, H. P. Nilles, F. Plöger, S. Raby, and M. Ratz, "Stringy origin of non-Abelian discrete flavor symmetries," *Nucl. Phys.*, vol. B768, pp. 135–156, 2007. DOI: 10.1016/j.nuclphysb.2007.01.018. arXiv: hep-ph/0611020 [hep-ph].

[144] K. R. Dienes and M. Lennek, "Fighting the floating correlations: Expectations and complications in extracting statistical correlations from the string theory landscape," *Phys. Rev.*, vol. D75, p. 026 008, 2007. eprint: hep-th/0610319.

[145] J. A. Casas, M. Mondragon, and C. Muñoz, "Reducing the Number of Candidates to Standard Model in the $Z(3)$ Orbifold," *Phys. Lett.*, vol. B230, pp. 63–70, 1989. DOI: 10.1016/0370-2693(89)91654-7.

[146] J. Fuchs and C. Schweigert, *Symmetries, Lie algebras and representations: A graduate course for physicists*. Cambridge, UK: University Press, 1997, 438 p.

[147] S. Groot Nibbelink and P. K. Vaudrevange, "Schoen manifold with line bundles as resolved magnetized orbifolds," *JHEP*, vol. 1303, p. 142, 2013. DOI: 10.1007/JHEP03(2013)142. arXiv: 1212.4033 [hep-th].

[148] W. Buchmüller, R. Ruckl, and D. Wyler, "Leptoquarks in Lepton - Quark Collisions," *Phys. Lett.*, vol. B191, pp. 442–448, 1987, [Erratum: Phys. Lett.B448,320(1999)]. DOI: 10.1016/S0370-2693(99)00014-3,10.1016/0370-2693(87)90637-X.

[149] M. Bauer and M. Neubert, "Minimal Leptoquark Explanation for the $\mathrm{R}_{D^{(*)}}$, R_K , and $(g-2)_g$ Anomalies," *Phys. Rev. Lett.*, vol. 116, no. 14, p. 141 802, 2016. DOI: 10.1103/PhysRevLett.116.141802. arXiv: 1511.01900 [hep-ph].

[150] B. Diaz, M. Schmaltz, and Y.-M. Zhong, "The leptoquark Hunter's guide: Pair production," *JHEP*, vol. 10, p. 097, 2017. DOI: 10.1007/JHEP10(2017)097. arXiv: 1706.05033 [hep-ph].

[151] M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev.*, vol. D98, no. 3, p. 030 001, 2018. DOI: 10.1103/PhysRevD.98.030001.

[156] S. Hamidi and C. Vafa, "Interactions on Orbifolds," *Nucl. Phys.*, vol. B279, p. 465, 1987.

[157] L. J. Dixon, D. Friedan, E. J. Martinec, and S. H. Shenker, "The Conformal Field Theory of Orbifolds," *Nucl. Phys.*, vol. B282, pp. 13–73, 1987.

[158] H. P. Nilles, S. Ramos-Sánchez, M. Ratz, and P. K. Vaudrevange, "A note on discrete R symmetries in $\mathbb{Z}_6$-II orbifolds with Wilson lines," *Phys. Lett.*, vol. B726, pp. 876–881, 2013. DOI: 10.1016/j.physletb.2013.09.041. arXiv: 1308.3435 [hep-th].

[152] A. E. Faraggi, "Hierarchical top - bottom mass relation in a superstring derived standard - like model," *Phys. Lett.*, vol. B274, pp. 47–52, 1992. DOI: 10.1016/0370-2693(92)90302-K.

[153] A. E. Faraggi, "Top quark mass prediction in superstring derived standard - like model," *Phys.Lett.B*, vol. 377, pp. 43–47, 1996. DOI: 10.1016/0370-2693(96)00310-3. arXiv: hep-ph/9506388.

[154] G. Burdman and Y. Nomura, "Unification of Higgs and gauge fields in five dimensions," *Nucl. Phys.*, vol. B656, pp. 3–22, 2003. arXiv: hep-ph/0210257.

[155] P. Hosteins, R. Kappl, M. Ratz, and K. Schmidt-Hoberg, “Gauge-top unification,” *JHEP*, vol. 07, p. 029, 2009. DOI: `10.1088/1126-6708/2009/07/029`. arXiv: `0905.3323 [hep-ph]`.

[159] F. Feruglio and A. Romanino, “Neutrino Flavour Symmetries,” 2019. arXiv: `1912.06028 [hep-ph]`.

[160] M. B. Green and J. H. Schwarz, “Anomaly Cancellation in Supersymmetric $D = 10$ Gauge Theory and Superstring Theory,” *Phys. Lett.*, vol. B149, pp. 117–122, 1984. DOI: `10.1016/0370-2693(84)91565-X`.

[161] C. Froggatt and H. B. Nielsen, “Hierarchy of Quark Masses, Cabibbo Angles and CP Violation,” *Nucl. Phys.*, vol. B147, p. 277, 1979. DOI: `10.1016/0550-3213(79)90316-X`.

[162] H. P. Nilles, “Dynamically broken supergravity and the hierarchy problem,” *Phys. Lett.*, vol. B115, p. 193, 1982.

[163] S. Ferrara, L. Girardello, and H. P. Nilles, “Breakdown of local supersymmetry through gauge fermion condensates,” *Phys. Lett.*, vol. B125, p. 457, 1983.

[164] J. P. Derendinger, L. E. Ibáñez, and H. P. Nilles, “On the low-energy $d = 4$, $N = 1$ supergravity theory extracted from the $d = 10$, $N = 1$ superstring,” *Phys. Lett.*, vol. B155, p. 65, 1985.

[165] M. Dine, R. Rohm, N. Seiberg, and E. Witten, “Gluino condensation in superstring models,” *Phys. Lett.*, vol. B156, p. 55, 1985.

[166] V. S. Kaplunovsky, “One Loop Threshold Effects in String Unification,” *Nucl. Phys.*, vol. B307, p. 145, 1988, [Erratum: Nucl. Phys.B382,436(1992)]. DOI: `10.1016/0550-3213(88)90526-3`. arXiv: `hep-th/9205068 [hep-th]`.

[167] L. J. Dixon, V. Kaplunovsky, and J. Louis, “Moduli dependence of string loop corrections to gauge coupling constants,” *Nucl. Phys.*, vol. B355, pp. 649–688, 1991.

[168] V. Kaplunovsky and J. Louis, “On Gauge couplings in string theory,” *Nucl. Phys.*, vol. B444, pp. 191–244, 1995. DOI: `10.1016/0550-3213(95)00172-O`. arXiv: `hep-th/9502077 [hep-th]`.

[169] S. King and G. G. Ross, “Fermion masses and mixing angles from SU(3) family symmetry,” *Phys. Lett. B*, vol. 520, pp. 243–253, 2001. DOI: `10.1016/S0370-2693(01)01139-X`. arXiv: `hep-ph/0108112`.

[170] S. King and G. G. Ross, “Fermion masses and mixing angles from SU (3) family symmetry and unification,” *Phys. Lett. B*, vol. 574, pp. 239–252, 2003. DOI: `10.1016/j.physletb.2003.09.027`. arXiv: `hep-ph/0307190`.

[171] M. Schmitz, “R-Symmetries from the Orbifolded Heterotic String,” PhD thesis, Bonn U., 2014.

[172] N. G. Cabo Bizet, T. Kobayashi, D. K. Mayorga Pena, S. L. Parameswaran, M. Schmitz, and I. Zavala, “R-charge Conservation and More in Factorizable and Non-Factorizable Orbifolds,” *JHEP*, vol. 05, p. 076, 2013. DOI: `10.1007/JHEP05(2013)076`. arXiv: `1301.2322 [hep-th]`.

[173] N. G. Cabo Bizet, T. Kobayashi, D. K. Mayorga Pena, S. L. Parameswaran, M. Schmitz, and I. Zavala, “Discrete R-symmetries and Anomaly Universality in Heterotic Orbifolds,” *JHEP*, vol. 02, p. 098, 2014. DOI: `10.1007/JHEP02(2014)098`. arXiv: `1308.5669 [hep-th]`.

[174] R. Deen, Y.-H. He, S.-J. Lee, and A. Lukas, "Machine Learning String Standard Models," Mar. 2020. arXiv: `2003.13339 [hep-th]`.

[175] H. Otsuka and K. Takemoto, "Deep learning and k-means clustering in heterotic string vacua with line bundles," Mar. 2020. arXiv: `2003.11880 [hep-th]`.

[176] J. Halverson and C. Long, "Statistical Predictions in String Theory and Deep Generative Models," 2020. arXiv: `2001.00555 [hep-th]`.

[177] H. Erbin and S. Krippendorf, "GANs for generating EFT models," Sep. 2018. arXiv: `1809.02612 [cs.LG]`.

[178] R. Altman, J. Carifio, J. Halverson, and B. D. Nelson, "Estimating Calabi-Yau Hypersurface and Triangulation Counts with Equation Learners," *JHEP*, vol. 03, p. 186, 2019. DOI: `10.1007/JHEP03(2019)186`. arXiv: `1811.06490 [hep-th]`.

[179] D. Z. Freedman and A. Van Proeyen, *Supergravity*. Cambridge, UK: Cambridge Univ. Press, May 2012, ISBN: 978-1-139-36806-3, 978-0-521-19401-3.

[180] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Statist. Sci.*, vol. 16, no. 2, pp. 101–133, May 2001. DOI: `10.1214/ss/1009213286`.

[181] F. Ruehle, "Data science applications to string theory," *Phys. Rept.*, vol. 839, pp. 1–117, 2020. DOI: `10.1016/j.physrep.2019.09.005`.

Machine learning and statistics bibliography

[182] M. Waskom, O. Botvinnik, D. O'Kane, P. Hobson, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. de Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Yarkoni, M. L. Williams, C. Evans, C. Fitzgerald, Brian, C. Fonnesbeck, A. Lee, and A. Qalieh, "Mwaskom/seaborn: V0.8.1 (september 2017)," version v0.8.1, 2017. DOI: `10.5281/zenodo.883859`.

[183] M. Bastian, S. Heymann, and M. Jacomy, *Gephi: An open source software for exploring and manipulating networks*, 2009.

[184] J. Hare, "Dealing with Sparse Rewards in Reinforcement Learning," *arXiv e-prints*, arXiv: 1910.09281, 2019. arXiv: `1910.09281 [cs.LG]`.

[185] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," *CoRR*, vol. abs/1601.06759, 2016. arXiv: `1601.06759`.

[186] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006, ISSN: 0036-8075. DOI: `10.1126/science.1127647`. eprint: `http://science.sciencemag.org/content/313/5786/504.full.pdf`.

[187] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[188] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *CoRR*, vol. abs/1706.02515, 2017. arXiv: `1706.02515`.

[189] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[190] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[191] E. Parviainen, “Deep bottleneck classifiers in supervised dimension reduction,” in *Artificial Neural Networks – ICANN 2010*, K. Diamantaras, W. Duch, and L. S. Iliadis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–10, ISBN: 978-3-642-15825-4.

[192] C. Song, F. Liu, Y. Huang, L. Wang, and T. Tan, “Auto-encoder based data clustering,” in *Proceedings, Part I, of the 18th Iberoamerican Congress on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - Volume 8258*, ser. CIARP 2013, Havana, Cuba: Springer-Verlag, 2013, pp. 117–124, ISBN: 978-3-642-41821-1. DOI: `10.1007/978-3-642-41822-8_15`.

[193] G. Dong and J. Bailey, *Contrast Data Mining: Concepts, Algorithms, and Applications*, 1st. Chapman & Hall/CRC, 2012, ISBN: 1439854327.

[194] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006, ISBN: 0387310738.

[195] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *Journal of Big Data*, vol. 6, no. 1, p. 27, Mar. 2019, ISSN: 2196-1115. DOI: `10.1186/s40537-019-0192-5`.

[196] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, “Training deep neural networks on imbalanced data sets,” *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 4368–4374, 2016.

[197] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009.

[198] J. Friedman, T. Hastie, and R. Tibshirani, “Additive Logistic Regression: a Statistical View of Boosting,” *The Annals of Statistics*, vol. 38, no. 2, 2000.

[199] L. Mason, J. Baxter, P. Bartlett, and M. Frean, “Boosting algorithms as gradient descent,” in *In Advances in Neural Information Processing Systems 12*, MIT Press, 2000, pp. 512–518.

[200] G. Ke, Q. Meng, T. Finely, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Lightgbm: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems 30 (NIP 2017)*, Dec. 2017.

[201] F. Chollet *et al.*, *Keras*, `https://keras.io`, 2015.

[202] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” *CoRR*, 2016. arXiv: `1603.02754`.

[203] F. Lavancier and P. Rochet, *A tutorial on estimator averaging in spatial point process models*, 2016. arXiv: `1607.00864 [math.ST]`.

[204] C. Doersch, *Tutorial on variational autoencoders*, 2016. arXiv: `1606.05908 [stat.ML]`.

[205] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Networks,” *arXiv e-prints*, arXiv: 1406.2661, 2014. arXiv: `1406.2661 [stat.ML]`.

[206] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” *arXiv e-prints*, arXiv: 1701.07875, 2017. arXiv: `1701.07875 [stat.ML]`.

[207] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, “Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications,” *CoRR*, vol. abs/1701.05517, 2017. arXiv: `1701.05517`.

[208] F. Agostinelli, M. D. Hoffman, P. J. Sadowski, and P. Baldi, "Learning activation functions to improve deep neural networks," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[209] G. Martius and C. H. Lampert, "Extrapolation and learning equations," *CoRR*, vol. abs/1610.02995, 2016. arXiv: `1610.02995`.

[210] F. J. Samaniego, *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. Springer New York, 2010. doi: `10.1007/978-1-4419-5941-6`.

[211] H. H. Ku, "Notes on the use of propagation of error formulas," *Journal of Research of the National Bureau of Standards*, vol. 70C, No.4, 1966.

[212] J. A. Hanley and A. Lippman-Hand, "If nothing goes wrong, is everything all right? interpreting zero numerators.," *JAMA*, vol. 249 13, pp. 1743–5, 1983.

[213] E. Eypasch, R. Lefering, C. K. Kum, and H. Troidl, "Probability of adverse events that have not yet occurred: A statistical reminder," *BMJ*, vol. 311, no. 7005, pp. 619–620, 1995, issn: 0959-8138. doi: `10.1136/bmj.311.7005.619`. eprint: `https://www.bmj.com/content`.

[214] D. Erhan, Y. Bengio, A. Courville, and P. Vincent, "Visualizing higher-layer features of a deep network," University of Montreal, Tech. Rep. 1341, 2009, Also presented at the ICML 2009 Workshop on Learning Feature Hierarchies, Montréal, Canada.

[215] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 818–833, isbn: 978-3-319-10590-1.

[216] B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," *CoRR*, vol. abs/1512.04150, 2015. arXiv: `1512.04150`.

[217] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *CoRR*, vol. abs/1012.2599, 2010. arXiv: `1012.2599`.

[218] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.

[219] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, Dec. 1989, issn: 1435-568X. doi: `10.1007/BF02551274`.

[220] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993, issn: 0893-6080. doi: `https://doi.org/10.1016/S0893-6080(05)80131-5`.

[221] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, `http://www.deeplearningbook.org`.

[222] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. arXiv: `1412.6980 [cs.LG]`.