



# Satisfiability modulo theories and chiral heterotic string vacua with positive cosmological constant

Alon E. Faraggi<sup>a,\*</sup>, Benjamin Percival<sup>a</sup>, Sven Schewe<sup>b</sup>, Dominik Wojtczak<sup>b</sup>

<sup>a</sup> Dept. of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, UK

<sup>b</sup> Dept. of Computer Sciences, University of Liverpool, Liverpool L69 7ZL, UK

## ARTICLE INFO

### Article history:

Received 27 January 2021

Accepted 26 February 2021

Available online 10 March 2021

Editor: N. Lambert

## ABSTRACT

We apply Boolean Satisfiability (SAT) and Satisfiability Modulo Theories (SMT) solvers in the context of finding chiral heterotic string models with positive cosmological constant from  $\mathbb{Z}_2 \times \mathbb{Z}_2$  orbifolds. The power of using SAT/SMT solvers to sift large parameter spaces quickly to decide satisfiability, both to declare and prove unsatisfiability and to declare satisfiability, are demonstrated in this setting. These models are partly chosen to be small enough to plot the performance against exhaustive search, which takes around 2 hours 20 minutes to comb through the parameter space. We show that making use of SMT based techniques with integer encoding is rather simple and effective, while a more careful Boolean SAT encoding provides a significant speed-up – determining satisfiability or unsatisfiability has, in our experiments varied between 0.03 and 0.06 seconds, while determining all models (where models exist) took 19 seconds for a constraint system that allows for 2048 models and 8.4 seconds for a constraint system that admits 640 models. We thus gain several orders of magnitude in speed, and this advantage is set to grow with a growing parameter space. This holds the promise that the method scales well beyond the initial problem we have used it for in this paper.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Funded by SCOAP<sup>3</sup>.

## 1. Introduction

Although the Standard Model (SM) of particle physics successfully predicts all observational data from particle collider experiments to date, several mysteries are left unanswered. Among the most important of these mysteries are the origin of dark matter, how to stabilise the Higgs mass and how gravity may be incorporated into the framework. To address these issues we can turn to String theory, which is the leading candidate for a theory beyond the Standard Model and naturally allows for the simultaneous analysis of gauge and gravitational interactions.

The goal of identifying 4D string models that come as close as possible to the SM or Minimal Supersymmetric Standard Model (MSSM) has been pursued within a range of different approaches (see e.g. ref. [1] and references therein). The main obstacle to these searches is the vast space of consistent string vacua in 4D, known as the string landscape. From a geometric perspective, this vast space of string vacua is generated by the enormous number of consistent compactifications of the 10D superstring down to 4D.

In this paper we work within the free fermionic worldsheet description [2] of the heterotic string, where the large space of vacua comes from the freedom in choosing the boundary conditions of worldsheet fermions. This construction can be translated into the geometric context of  $\mathbb{Z}_2 \times \mathbb{Z}_2$  orbifolds as detailed in ref. [3].

The exploration of subspaces of the string landscape using Machine learning and Deep Learning has become a burgeoning area of research in string phenomenology, see, e.g., refs. [4] and for review see, e.g., ref. [5]. In this paper we make use of a different tool within this context: Boolean Satisfiability Checking (SAT) and Satisfiability Modulo Theories (SMT). SAT and SMT solvers use powerful algorithms for determining the satisfiability of constraints as well as for finding solutions where the constraints are satisfiable. We expect SAT and SMT solving to have a wide range of applications in string phenomenology as they have, indeed, proven to have in a range of other fields, such as computational biology [6], artificial intelligence [7], combinatorics [8], and mathematical geometry [9].

Typically, the input variables needed to specify a string model are positive integers and *a priori* there are approximately 100 of them. They may, for example, specify the geometry of the compactified six-dimensional space or, in our case, generate boundary conditions for world-sheet free fermions. These inputs are required to meet certain consistency constraints, such as modular invariance, which reduces the number of independent input variables.

\* Corresponding author.

E-mail addresses: [alon.faraggi@liv.ac.uk](mailto:alon.faraggi@liv.ac.uk) (A.E. Faraggi), [benjamin.percival@liv.ac.uk](mailto:benjamin.percival@liv.ac.uk) (B. Percival), [sven.schewe@liverpool.ac.uk](mailto:sven.schewe@liverpool.ac.uk) (S. Schewe), [d.wojtczak@liv.ac.uk](mailto:d.wojtczak@liv.ac.uk) (D. Wojtczak).

In the case of free fermionic models, the boundary conditions are typically either Ramond or Neveu-Schwarz, and so the input variables can be represented as Boolean variables. Classes of models can be analysed in terms of relations of these variables, and constraints related to desirable physical properties can be imposed. Some examples of typical constraints are the presence of three particle generations, Higgs content and the absence of chiral exotic states. In recent works we have also been interested in unphysical, but interesting, string vacua, which are free of massless fermions or massless twisted bosons that are called Type 0 [10] and Type  $\bar{0}$  [11] models, respectively. These models have potential applications within cosmological scenarios of the early universe.

As more and more phenomenological constraints are imposed, the frequency of viable models can become low enough to evade random searches that take many weeks. This was found, for example, within the classification of Standard-Like Models and Left-Right Symmetric models [12,13] where, in the latter case, models would satisfy all imposed phenomenological criteria with a probability in the order of  $10^{-11}$ . This scarcity of viable models has motivated the introduction of the fertility methodology within these classes of vacua [12,14]. Furthermore, in some classes of vacua, important distinct characteristics of string models have been found to be incompatible. For example, within the classification of  $10^{12}$  Flipped  $SU(5)$  vacua [15], there were no exophobic models with three generations, whereas such vacua were found within the classification of Pati-Salam models [16]. A similar result was found in recent work on Type 0 vacua [10], where it was shown analytically for a minimal class of models that the absence of massless fermions necessitates the presence of tachyons, and the result persisted within a generalised class of models. As the construction of models becomes more complex and realistic, it is not straightforward to provide an analytic proof of apparent incompatibilities for phenomenological constraints that appear in classification statistics. Scenarios of this kind are the strong suit of SAT and SMT, as they are powerful and proven tools in both finding rare solutions (and thus uncovering models with interesting properties) and proving unsatisfiability for constraints and isolating the origin of this unsatisfiability.

In this paper we first of all test the efficiency of reducing the problem of finding tachyon-free Type  $\bar{0}$  models of the sort found in [11] to SAT or SMT. We first employ a standard SMT solver, Z3 [17], in an intuitive direct encoding of the input variables as integers, where all parameters are either 0 (corresponding to NS boundary conditions) or 1 (corresponding to R boundary conditions), which results in an efficiency improvement compared with a random classification: within the class of models we explore, the random classification approach finds approximately 500 tachyon-free Type  $\bar{0}$  vacua in 40 minutes, which is the time it takes the SMT with integer encoding to find all 2048 solutions and to complete the search. We then go further to translate the constraints into a Boolean encoding such that the system is simply a SAT problem and the full power of the SAT solver, also part of Z3 [17], is demonstrated. In this case, establishing satisfiability takes 0.04 seconds while constructing all models takes 19 seconds, which is 126 times faster than with integer encoding, and around 450 times faster than enumerating and testing all candidate models.

Having established this, we then employ the SAT and SMT solvers to confirm a contradiction between the presence of spinorial  $\mathbf{16}$ 's for (tachyon-free) Type  $\bar{0}$  models. The SAT solver not only confirms the unsatisfiability of these constraints in 0.06 seconds in the Boolean encoding (163 seconds for the SMT solver using integer encoding), which is more than 100,000 times faster than an exhausting search. The SAT solver can also be used to identify a minimal 'unsatisfiable core' [18], which isolates where the contradiction arises by giving a (locally) minimal subset of constraints, where dropping either of them results in a satisfiable

constraint system. From this, it is straightforward to apply Optimisation Modulo Theory (OMT), offered by many SMT solvers, to find models with a minimal number of massless twisted bosons that also contain  $\mathbf{16}/\bar{\mathbf{16}}$ 's. Where OMT is not offered, one can also add a minimisation/maximisation constraint to manually hone in on contradictions and optimal configurations. Due to the abundance of massless fermions in their spectra, such models are expected to have a positive cosmological constant at the free fermionic point and could be candidates for having a necessarily positive one-loop potential once the string moduli are incorporated, as discussed in ref. [19].

The models we explore can be regarded as compactifications of the non-supersymmetric 10D tachyonic heterotic string vacua [20–23]. By construction, the class of models we analyse here cannot give rise to (quasi-)realistic string vacua but rather is chosen as a simplified set-up, which is perfect for illustrating some key applications of SAT and SMT solvers. We expect the advantages of the SAT/SMT approach to magnify for more realistic classes of models with larger input spaces.

The structure of the paper is as follows: in Section 2 we provide an introduction to SAT/SMT. In Section 3 we define the class of models we will explore and the constraints we evaluate. In Section 4 we detail how these can be evaluated using an SMT and SAT solvers and present the results. Section 5 concludes the paper.

## 2. SMT and SAT

Satisfiability Modulo Theories (SMTs) are powerful algorithms used for deciding whether a set of constraints describing a problem is satisfiable. In other words, SMTs determine whether there exists a 'satisfying assignment' of a set of input variables to a system of constraints. These constraint formulae are constructed by defining operations over, what are referred to as, theory variables, and combining them with logical connectives. SMT problems are more expressive and powerful than the Boolean Satisfiability (SAT) problems that restrict all variables to be true or false and operators to be logical connectives. In particular, SMTs allow for operations over non-Boolean types such as integers, reals, bitvectors, and arrays. Both SMT and SAT are canonical NP-complete problems [24], which is a class of computational problems for which checking whether a given variable assignment satisfies the constraints can be done in polynomial time, but finding such an assignment is believed to be hard. Despite a lot of effort, no polynomial time algorithm was shown for any NP-complete problem since the class was defined in 1971. More than that, a widely believed conjecture [25] states that in the worst-case one cannot significantly improve over an exhaustive search of all possible assignments. Nevertheless, there has been a tremendous progress over the last decades in efficiency of algorithms solving these problems in practice, solving instances with hundreds and even thousands of variables, which shows that truly hard instances are few and far between.

A key aspect of how SMT-solvers work so effectively is through following the DPLL or conflict-driven clause learning (CDCL) class of algorithms. These algorithms implement a decision procedure for each theory by adding or subtracting constraints and querying for satisfiability as it goes. More detail on DPLL(T) and other decision procedures may be found in e.g. [26].

One of the most efficient and easy to use SMT solvers is Z3, which can be found open source on Github at <https://github.com/Z3Prover/Z3>. Z3 was developed by Microsoft primarily for software verification purposes. It also implements an efficient SAT solver that we use for the Boolean encoding of our problem. It has bindings for most common programming languages and in our case we used the Python front-end as a means of interfacing with Z3.

### 3. Minimal tachyon-free $SO(10)$ $\tilde{S}$ -models

We will utilise the free fermionic construction [2] to define a class of non-supersymmetric  $\mathbb{Z}_2 \times \mathbb{Z}_2$  orbifold models with an unbroken  $SO(10)$  observable group. Only the key aspects of the free fermionic construction will be described here, as our main purpose is the application of the SMT solver within this setting. We will be adopting the conventional notation used in the free fermionic literature [27–31,16,15,19,12–14]. Over the past two decades systematic methods to classify large numbers of free fermionic heterotic-string models were developed [32,33,16,15,12,13,22,23]. The initial method was developed for the classification of spinorial and anti-spinorial representations of an unbroken  $SO(10)$  GUT group [32], and extended to include its vectorial representations [33], which led to the discovery of spinor-vector duality over the space of vacua [33,34]. It was extended to include the entire massless twisted spectrum in models with  $SO(6) \times SO(4)$  [16],  $SU(5) \times U(1)$  [15],  $SU(3) \times SU(2) \times U(1)^2$  [12],  $SU(3) \times U(1) \times SU(2)^2$  [13,14], unbroken  $SO(10)$  subgroups. Exophobic three generation models, in which exotic fractionally charged states only appear in the massive string spectrum were discovered in the case of  $SO(6) \times SO(4)$  models, whereas all other cases contained non-chiral massless exotic states in the spectra of three generation models. Over the past year the classification methodology was extended to non-supersymmetric heterotic-string vacua [22,10,23,11], in which case the existence of physical tachyonic states in the physical spectrum is of particular interest. In the process of such classifications, we are particularly interested in the presence, or lack thereof, of some specific states in the spectrum of the models. It is evident that SAT and SMT solving are particularly well suited to address such questions.

The construction of a free fermionic string model is defined at the free fermionic point in the moduli space and is generated by specifying two ingredients. The first is a set of  $N$  boundary condition basis vectors,  $v_i \in \mathcal{B}$ ,  $i = 1, \dots, N$ , and the second is a set of one-loop Generalised GSO (GGSO) phases,  $C \begin{bmatrix} v_i \\ v_j \end{bmatrix}$ .

We will use the basis explored in [11], where Type  $\bar{0}$  string vacua, i.e. models without any twisted massless bosons, were uncovered. This basis is written as

$$\begin{aligned} \mathbb{1} &= \{\psi^\mu, \chi^{1,\dots,6}, y^{1,\dots,6}, w^{1,\dots,6} | \bar{y}^{1,\dots,6}, \\ &\quad \bar{w}^{1,\dots,6}, \bar{\psi}^{1,\dots,5}, \bar{\eta}^{1,2,3}, \bar{\phi}^{1,\dots,8}\}, \\ \tilde{S} &= \{\psi^\mu, \chi^{1,\dots,6} | \bar{\phi}^{3,4,5,6}\}, \\ T_1 &= \{y^{1,2}, w^{1,2} | \bar{y}^{1,2}, \bar{w}^{1,2}\}, \\ T_2 &= \{y^{3,4}, w^{3,4} | \bar{y}^{3,4}, \bar{w}^{3,4}\}, \\ T_3 &= \{y^{5,6}, w^{5,6} | \bar{y}^{5,6}, \bar{w}^{5,6}\}, \\ b_1 &= \{\psi^\mu, \chi^{12}, y^{34}, y^{56} | \bar{y}^{34}, \bar{y}^{56}, \bar{\eta}^1, \bar{\psi}^{1,\dots,5}\}, \\ b_2 &= \{\psi^\mu, \chi^{34}, y^{12}, w^{56} | \bar{y}^{12}, \bar{w}^{56}, \bar{\eta}^2, \bar{\psi}^{1,\dots,5}\}, \\ b_3 &= \{\psi^\mu, \chi^{56}, w^{12}, w^{34} | \bar{w}^{12}, \bar{w}^{34}, \bar{\eta}^3, \bar{\psi}^{1,\dots,5}\}, \\ z_1 &= \{\bar{\phi}^{1,\dots,4}\}. \end{aligned} \quad (3.1)$$

We note that the vectors  $T_i$ ,  $i = 1, 2, 3$  allow for internal symmetric shifts around the 3 internal  $T^2$  tori. The inclusion of  $\tilde{S}$  in the basis means we have no supersymmetry and these vacua can be considered as compactifications of the tachyonic 10D heterotic string, as discussed in [20–23].

The NS sector vector gauge bosons give rise to a gauge group

$$SO(10) \times U(1)^3 \times SO(4)^3 \times SU(2)^8 \quad (3.2)$$

and may receive additional vector boson enhancements from the sectors

$$\left\{ \begin{array}{l} \psi^\mu |z_1\rangle_L \otimes \{\bar{\lambda}^i\} |z_1\rangle_R \\ \psi^\mu |z_2\rangle_L \otimes \{\bar{\lambda}^i\} |z_2\rangle_R \\ \psi^\mu |z_1 + z_2\rangle_L \otimes |z_1 + z_2\rangle_R \end{array} \right\} \quad (3.3)$$

where  $\bar{\lambda}^i$  are all possible right moving Neveu-Schwarz oscillators and  $z_2$  is the important linear combination

$$z_2 = 1 + b_1 + b_2 + b_3 + z_1 = \{\bar{\phi}^{5,6,7,8}\}. \quad (3.4)$$

Having defined these 9 basis vectors, the other ingredient for defining a model is the GGSO coefficients,  $C \begin{bmatrix} v_i \\ v_j \end{bmatrix} = \pm 1$ , which will generate a  $9 \times 9$  matrix for this basis but, due to modular invariance constraints, only the upper triangle of 36 GGSO phases is independent. This leaves a space of  $2^{36} \sim 6.87 \times 10^{10}$  independent GGSO phase configurations.

With an assignment of GGSO phases the Hilbert space of states  $|S_\xi\rangle$  can be specified through

$$\mathcal{H} = \bigoplus_{\xi \in \Xi} \prod_{i=1}^k \left\{ e^{i\pi v_i \cdot F_\xi} |S_\xi\rangle = \delta_\xi C \begin{bmatrix} \xi \\ v_i \end{bmatrix}^* |S_\xi\rangle \right\}, \quad (3.5)$$

where  $\xi$  are sectors (linear combinations of basis vectors) in the additive space  $\Xi$  and  $\delta_\xi$  is the spin statistic index. The GGSO projection equation inside the curly brackets of (3.5) will give us the constraints implemented within the SMT solver described in the next section.

The sectors in the model can be characterised according to their left and right moving vacuum separately

$$\begin{aligned} M_L^2 &= -\frac{1}{2} + \frac{\xi_L \cdot \xi_L}{8} + N_L \\ M_R^2 &= -1 + \frac{\xi_R \cdot \xi_R}{8} + N_R \end{aligned} \quad (3.6)$$

where  $N_L$  and  $N_R$  are sums over left and right moving oscillators, respectively. Physical states must additionally satisfy the Virasoro matching condition,  $M_L^2 = M_R^2$  such that massless states are those with  $M_L^2 = M_R^2 = 0$  and physical tachyons arise for sectors with  $M_L^2 = M_R^2 < 0$ .

#### 3.1. Massless bosonic sector analysis

The first aspect of these models we wish to explore is the massless twisted bosons. In [11], we have described the conditions on the absence of twisted massless bosons within this class of models and found tachyon-free GGSO configurations that corresponded to two distinct partition functions. The details are repeated here as we will apply the SMT solver to these conditions on the absence of massless twisted bosons.

For this class of models, there are 15 vectorial bosonic sectors of the form

$$\begin{aligned} V_{pq}^1 &= b_2 + b_3 + T_1 + pT_2 + qT_3 \\ V_{pq}^2 &= b_1 + b_3 + T_2 + pT_1 + qT_3 \\ V_{pq}^3 &= b_1 + b_2 + T_3 + pT_1 + qT_2 \\ V^4 &= T_1 + T_2 \\ V^5 &= T_1 + T_3 \\ V^6 &= T_2 + T_3 \end{aligned} \quad (3.7)$$

where  $p, q = 0, 1$  and these sectors all come with a right-moving NS oscillator. Additionally, there are 30 fermionic spinorial sectors

$$\begin{aligned}
B_{pq}^1 &= b_2 + b_3 + z_1 + T_1 + pT_2 + qT_3 \\
B_{pq}^2 &= b_1 + b_3 + z_1 + T_2 + pT_1 + qT_3 \\
B_{pq}^3 &= b_1 + b_2 + z_1 + T_3 + pT_1 + qT_2 \\
B_{pq}^4 &= \mathbb{1} + b_1 + z_1 + T_1 + pT_2 + qT_3 \\
B_{pq}^5 &= \mathbb{1} + b_2 + z_1 + T_2 + pT_1 + qT_3 \\
B_{pq}^6 &= \mathbb{1} + b_3 + z_1 + T_3 + pT_1 + qT_2. \\
B^7 &= T_1 + T_2 + z_1 \\
B^8 &= T_1 + T_3 + z_1 \\
B^9 &= T_2 + T_3 + z_1 \\
B^{10} &= T_1 + T_2 + z_2 \\
B^{11} &= T_1 + T_3 + z_2 \\
B^{12} &= T_2 + T_3 + z_2
\end{aligned} \tag{3.8}$$

A Type  $\bar{0}$  model arises when all of these sectors are projected for an assignment of GGSO phases. For example, taking sector  $B_{pq}^1$ , a projector is constructed of the form

$$P_{pq}^1 = \frac{1}{2^3} \left( 1 + C \begin{bmatrix} B_{pq}^1 \\ T_1 \end{bmatrix} \right) \left( 1 + C \begin{bmatrix} B_{pq}^1 \\ z_2 \end{bmatrix} \right) \left( 1 + C \begin{bmatrix} B_{pq}^1 \\ b_1 + pT_2 + qT_3 \end{bmatrix} \right) \tag{3.9}$$

and requiring that this is zero ensures its absence from the massless spectrum. Repeating this for all massless twisted bosonic sectors allows for the identification of Type  $\bar{0}$  configurations.

### 3.2. Tachyon sector analysis

The next constraint to impose after the projection of massless bosonic sectors, is the absence of tachyonic sectors. Since our models are non-supersymmetric this is an important check for determining the stability of our models for a 4D Minkowski background. The same procedure of encoding the GGSO projections applies to the tachyonic sectors. The tachyonic sectors in this construction are:  $\{\bar{\lambda}\}T_i, z_1, z_2, z_1 + T_i$  and  $z_2 + T_i, i = 1, 2, 3$ , where  $i = 1, 2, 3$  and  $\bar{\lambda}$  is some right-moving NS oscillator and we note that the untwisted tachyon from  $|0\rangle_L \otimes \{\bar{\lambda}\}|0\rangle_R$  is projected regardless of the GGSO phase choices in this basis.

As an example, we can delineate the condition for the projection of the  $\{\bar{\lambda}\}T_1$  tachyonic sector

$$\#(x \in S | x = -1) > 1$$

where  $S = \left\{ C \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}, C \begin{bmatrix} T_1 \\ T_3 \end{bmatrix}, C \begin{bmatrix} T_1 \\ z_1 \end{bmatrix}, C \begin{bmatrix} T_1 \\ z_2 \end{bmatrix} \right\}$ ,  $(3.10)$

which ensures all oscillator cases are projected.

### 3.3. Spinorial $\mathbf{16}/\overline{\mathbf{16}}$ sectors

The fermion generations transforming in the spinorial  $\mathbf{16}/\overline{\mathbf{16}}$  of  $SO(10)$  arise from the twisted sectors

$$\begin{aligned}
F_{pq}^1 &= b_1 + pT_2 + qT_3 \\
F_{pq}^2 &= b_2 + pT_1 + qT_3 \\
F_{pq}^3 &= b_3 + pT_1 + qT_2,
\end{aligned} \tag{3.11}$$

which have projectors

$$\begin{aligned}
R_{pq}^1 &= \frac{1}{2^3} \left( 1 - C \begin{bmatrix} F_{pq}^1 \\ T_1 \end{bmatrix} \right) \left( 1 - C \begin{bmatrix} F_{pq}^1 \\ z_1 \end{bmatrix} \right) \left( 1 - C \begin{bmatrix} F_{pq}^1 \\ z_2 \end{bmatrix} \right) \\
R_{pq}^2 &= \frac{1}{2^3} \left( 1 - C \begin{bmatrix} F_{pq}^2 \\ T_2 \end{bmatrix} \right) \left( 1 - C \begin{bmatrix} F_{pq}^2 \\ z_1 \end{bmatrix} \right) \left( 1 - C \begin{bmatrix} F_{pq}^2 \\ z_2 \end{bmatrix} \right) \\
R_{pq}^3 &= \frac{1}{2^3} \left( 1 - C \begin{bmatrix} F_{pq}^3 \\ T_3 \end{bmatrix} \right) \left( 1 - C \begin{bmatrix} F_{pq}^3 \\ z_1 \end{bmatrix} \right) \left( 1 - C \begin{bmatrix} F_{pq}^3 \\ z_2 \end{bmatrix} \right)
\end{aligned} \tag{3.12}$$

which can be used to tell us  $\#(\mathbf{16} + \overline{\mathbf{16}})$  for any model. Knowing this is sufficient for our purposes here with the SMT solver analysis and we will not implement the chirality projection distinguishing the  $\mathbf{16}$  and  $\overline{\mathbf{16}}$ .

## 4. Application of SMT

Now we turn our attention to the analysis of these model characteristics with an SMT written using Z3 in Python. It is convenient to introduce the notation

$$C \begin{bmatrix} v_i \\ v_j \end{bmatrix} = \exp[i\pi(v_i|v_j)] \tag{4.1}$$

and use the 36 independent phases  $(v_i|v_j) \in \{0, 1\}$  as the input variables for our SMT solver. In terms of these variables, the GGSO projection equation for each sector can be written in terms of sums of  $(v_i|v_j)$  modulo 2. We call this representation the integer encoding for our system of constraints.

As mentioned in Section 2, we use the general purpose SMT solver Z3, which allows us to use mathematical expressions so we do not need to reduce the constraints to purely propositional logic as in this integer encoding of the GGSO equations. This comes at the expense of performance because the SMT solver needs to include reasoning for mathematical theories (such as integer arithmetic). Therefore we will also detail a Boolean encoding of our constraint system, where we rewrite the GGSO projections purely as Boolean propositions.

It turns out that the conditions for the projection of massless twisted bosons and tachyonic sectors do not involve the following 9 of the 36  $(v_i|v_j)$  phases:

$$(\mathbb{1}|\tilde{S}), (\mathbb{1}|b_1), (\mathbb{1}|b_2), (\mathbb{1}|b_3), (\mathbb{1}|z_1), (\tilde{S}|b_1), (\tilde{S}|b_2), (\tilde{S}|b_3), (\tilde{S}|z_1) \tag{4.2}$$

and so our space of models to explore is reduced to  $2^{27}$  (ca.  $1.34 \times 10^8$ ). This is well within the reach of a complete enumeration of possible GGSO configurations but our purpose here is testing the features and efficiency of the SMT solver within this simple class of models. Moreover, as it does not adversely affect performance, we have re-introduced these 9 variables as input to our Boolean encoding with the expectation that it will not significantly impact the running time (in face, it has proven to actually cut the running time by 4%.)

As an example of how we can write the GGSO projections in both the integer representation and the Boolean representation we will take the massless bosonic sector  $B_{00}^1$ . The GGSO projection equation is written for generic  $p, q = 0, 1$  in terms of  $C[\cdot]$ 's in eq. (3.9). In the integer encoding we can write the  $B_{00}^1$  projection condition by first defining

$$\begin{aligned}
P &= [(T_1|b_2) + (T_1|b_3) + (T_1|z_1) + 1 + (\mathbb{1}|T_1)] \bmod 2 \\
Q &= [(\mathbb{1}|T_1) + (T_1|b_1) + (T_1|b_2) + (T_1|b_3) + (T_1|z_1) + (b_1|b_2) \\
&\quad + (b_1|b_3) + (b_1|z_1)] \bmod 2 \\
R &= [1 + (b_1|b_2) + (b_1|b_3) + (b_1|z_1) + (T_1|b_1)] \bmod 2
\end{aligned} \tag{4.3}$$

and then constructing the conjunction

$$(P = 1 \vee Q = 1 \vee R = 1), \quad (4.4)$$

to impose the projection constraint. This can then be translated into the Boolean representation using the `Xor()` operator,  $\vee$ , as follows:

$$\begin{aligned} \tilde{P} &= (T_1|b_2) \vee (T_1|b_3) \vee (T_1|z_1) \vee \text{True} \vee \mathbb{1}|T_1 \\ \tilde{Q} &= (\mathbb{1}|T_1) \vee (T_1|b_1) \vee (T_1|b_2) \vee (T_1|b_3) \vee (T_1|z_1) \\ &\quad \vee (b_1|b_2) \vee (b_1|b_3) \vee (b_1|z_1) \\ \tilde{R} &= \text{True} \vee (b_1|b_2) \vee (b_1|b_3) \vee (b_1|z_1) \vee (T_1|b_1) \end{aligned} \quad (4.5)$$

where  $(v_i|v_j)$  are now Booleans rather than 0 or 1. The constraint for ensuring the projection of this sector is then

$$(\tilde{P} \vee \tilde{Q} \vee \tilde{R}). \quad (4.6)$$

For both representations we can schematically write the steps for the construction of the Z3 solver in the case of finding tachyon-free Type  $\bar{0}$  string models.<sup>1</sup>

- 
- 1: Define the 27 input variables  $c_0, \dots, c_{26}$
  - 2: Add constraints on input variable domain ( $c_i = 0 \vee c_i = 1$ )  $\forall i = 0, \dots, 26$ .
  - 3: Add constraints for GGSO projection of all twisted massless bosons
  - 4: Add constraints for GGSO projection of all tachyonic sectors.
  - 5: Check satisfiability OR
  - 6: Find satisfying assignments (print all solutions)
- 

Step 5 is perhaps the most fundamental application of SAT/SMT solvers, which can be used to quickly identify whether the set of constraints permits a solution or not. In this case, we can utilise the prior analysis in ref. [11], which found tachyon-free Type  $\bar{0}$  models, to realise that the SMT solver should return `sat` in this case.

#### 4.1. Results of SMT search for tachyon-free type $\bar{0}$ models

Enumerating all tachyon-free Type  $\bar{0}$  models within the class of models under consideration is a good testing ground for the efficiency of the SMT/SAT solver. As mentioned earlier, the space of models is  $2^{27}$  (ca.  $1.34 \times 10^8$ ), which is within the grasp of a complete enumeration approach. We have run a random search in comparison, which is able to analyse ca. 16,000 sample points per second. An exhaustive enumeration of the model thus takes around 2 hours 20 minutes, and a random search (with repetitions) needs on average around 19 hours to find all solutions and 4 seconds to find the first model.

Within the integer representation, Z3's SMT solver determines satisfiability in 12 seconds and finds all 2048 solutions in the full  $2^{27}$  space in 2405 seconds (ca. 40 minutes). A random search found just under a quarter of these models in the same amount of time (500 models). While this is a useful speed up, it is not that impressive. This is expected since, in this representation, the SMT solver deals with mathematical theories that create a significant overhead. Fig. 1 depicts the accumulation of solutions over time for the SMT solver within the integer representation.

The performance is significantly improved when implementing a Boolean encoding. In this case, Z3's SAT solver determines satisfiability in 0.04 seconds. It can find and print all 2048 solutions in 19 seconds. This is 126 times faster than for the integer representation, and 450 times faster than exhaustively enumerating and checking the statespace. Fig. 2 depicts the accumulation of solutions over time for the SAT solver, where we see that the solver

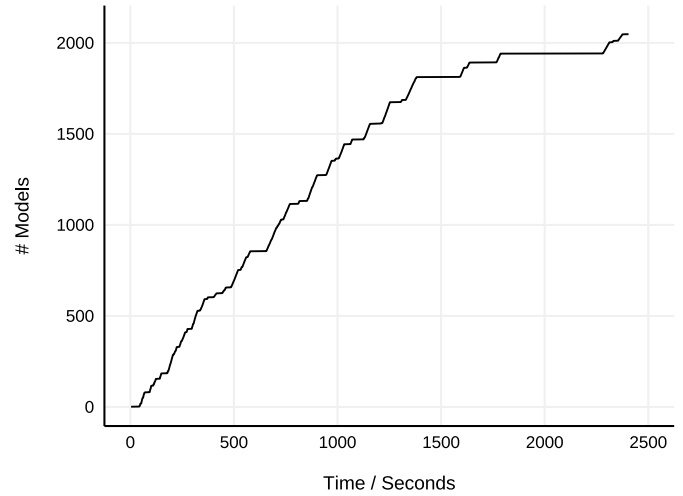


Fig. 1. Rate at which the Integer representation finds all tachyon-free Type  $\bar{0}$  models.

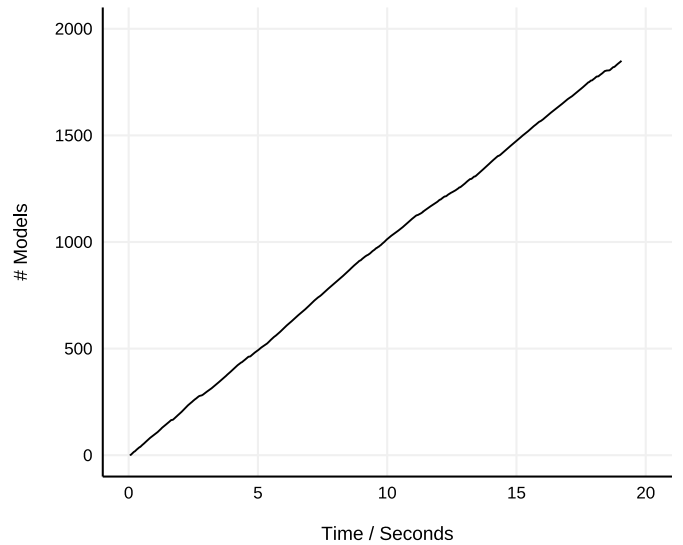


Fig. 2. Rate at which the Boolean representation finds all tachyon-free Type  $\bar{0}$  models, depicting the number of compact solutions found; the 1850 compact solutions contain all 2048 explicit solutions (exactly once).

is not slowed down by the creation of new lemmas as solutions are enumerated. It can be seen that only 1850 compact solutions are recorded on the graph in Fig. 2. This is because some of the compact solutions include variables labelled with `None`, meaning that this variable/s can be set to true or false; they can be trivially expanded to the 2048 solutions (without omissions and without multiple occurrences of individual solutions).

We have repeated the experiment using all 36 original input variables. Our expectation was that the SAT solver would essentially ignore them, as they do not enter into the reasoning at any point, though it will need to output a few additional `None` in the compact representation. We expected an insignificant increase in the overall running time. We found that the solver worked 4% faster, which is likely due to different decisions made by the heuristics. But it shows an interesting effect: overlooking the variables that do not matter did not slow the SAT solver down, whereas an exhaustive search would have taken 6 orders of magnitude longer.

#### 4.2. Identifying chiral, tachyon-free type $\bar{0}$ vacua

In the analysis from ref. [11] it was found that no Type  $\bar{0}$  vacua include the fermion generations from the spinorial  $\mathbf{16}/\bar{\mathbf{16}}$  sectors

<sup>1</sup> Both the integer and Boolean Z3 codes are available at <https://github.com/thePlumbaked/SMTsType0bar>.

(3.11). Since these sectors are phenomenologically desirable, we aim to ensure at least one remains in the Hilbert space after GGSO projections.

The sectors giving rise to the  $\mathbf{16}/\overline{\mathbf{16}}$  were given in eq. (3.11) and the projectors of eq. (3.12) can be rewritten in the integer and Boolean representations in a similar way as delineated above for  $B_{00}^1$ .

With the addition of this condition the SMT structure summary can be updated to

- 
- 1: Define the 27 input variables  $c_0, \dots, c_{26}$
  - 2: Add constraint on input variable domain ( $c_i = 0 \vee c_i = 1$ )  $\forall i = 0, \dots, 26$ .
  - 3: Add constraints for GGSO projection of all twisted massless bosons
  - 4: Add constraints for GGSO projection of all tachyonic sectors.
  - 5: Add constraint on presence of at least 1  $\mathbf{16}/\overline{\mathbf{16}}$  sector
  - 6: Check satisfiability OR
  - 7: Find satisfying assignments (solutions)
- 

As expected from the findings of ref. [11], with this added constraint on the spinorial  $\mathbf{16}/\overline{\mathbf{16}}$  the Z3 solver returns `unsat`. This takes only 0.06 seconds in the Boolean encoding (163 seconds in the integer encoding), underlining the feasibility of combing large parameter spaces.

In such cases where constraint systems are unsatisfiable there are several tools within Z3 that are helpful to understand and isolate where the inconsistency arises from. Using its `proof()` method, Z3 will output a proof of inconsistency. Unsatisfiability proofs can be long and tedious because, while satisfiability can be shown by providing a model, unsatisfiability needs to make a mathematical argument that establishes the contradiction. Such proofs are often long and tedious—though this is no comparison to the tedious work of manually distilling the contradiction, but it is fair to say that they often do not provide much accessible further insight.

There are, however, several other helpful tools that can be used to isolate inconsistencies. In particular, a minimal ‘unsatisfiable core’ of constraints can be returned by most SMT solvers. Additionally, a more manual approach of using push and pop methods on constraints allows for pinning down the source of an inconsistency. Using this approach, the presence of the spinorial  $\mathbf{16}/\overline{\mathbf{16}}$ ’s from  $F_{pq}^i$  is found to contradict the projection of the vectorial  $V_{pq}^i$  (under the presence of the remaining constraints). Since the Higgs bidoublet representation would reside within the vectorial  $\mathbf{10}$  of  $SO(10)$  coming from the  $V_{pq}^i$ ,  $i = 1, 2, 3$ , there is physical motivation to keep at least one of these twisted bosonic sectors in the Hilbert space. Demanding that at least one sector from  $F_{pq}^i$  remains, and at least one of  $V_{pq}^i$  makes the SMT return `sat`, which takes 0.03 seconds for the Boolean constraint system and 8.4 seconds for the integer encoding. Creating all 640 satisfying assignments in 7.5 seconds for the Boolean constraint system (1661 seconds for the integer encoding) of the 27 input variables with these conditions. One such model is defined by

$$C \begin{bmatrix} v_i \\ v_j \end{bmatrix} = \begin{pmatrix} \mathbf{1} & \tilde{S} & T_1 & T_2 & T_3 & b_1 & b_2 & b_3 & z_1 \\ \mathbf{1} & 1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ \tilde{S} & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ T_1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 \\ T_2 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 \\ T_3 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ b_1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ b_2 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 \\ b_3 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ z_1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 \end{pmatrix} \quad (4.7)$$

which has partition function

$$Z = 2\bar{q}^{-1} + 1280q^{-1/2}\bar{q}^{1/4} + 48q^{1/4}\bar{q}^{-3/4} - 1016 - 16288q^{1/4}\bar{q}^{-1/4} + \dots \quad (4.8)$$

where we can see that  $N_b^0 - N_f^0 = -1016$ , which is a large abundance of massless fermions as expected. This model generates a worldsheet cosmological constant of  $\Lambda = -886.43$  which corresponds to a large positive cosmological constant,  $\lambda$ , via the rescaling  $\lambda = -\frac{1}{2}\mathcal{M}^4\Lambda$ , where  $\mathcal{M} = M_{String}/2\pi$ .

## 5. Conclusion

The application of Machine Learning techniques within the string landscape is already a burgeoning field. In this work, we open the door to the application of SAT and SMT solvers within this context and have demonstrated their power and efficiency within a simplified class of string models and expect their benefits to only increase as more generalised and (quasi-)realistic classes of vacua are studied.

We have demonstrated how SAT and SMT solvers can be used to help isolate inconsistent constraints and be used to optimise desired characteristics. This method was then employed to find string vacua with positive cosmological constant and desirable  $SO(10)$  representations. Furthermore, this approach essentially meant minimising the number of massless twisted bosons, which is generally enough to ensure a large abundance of massless fermions resulting in the largest contribution to cosmological constant (the massless level) being positive. Although the massive and off-shell contributions also need accounting for, a large abundance of massless fermions can effectively guarantee a positive cosmological constant. Such vacua are then ripe for analysis away from the free fermionic point in the moduli space by exploring the one-loop potential in terms of the moduli fields as was pursued in [19], where a model with  $N_f > N_b$  at the free fermionic point and  $N_b = N_f$  at a generic point in the moduli space resulted in guaranteeing a positivity of the one-loop potential. Using the SAT and SMT solvers described in the current work makes it very quick and easy to find string models at the free fermionic point with desired properties at the massless level, even when the models are rare in the full space.

There are a couple of obvious limitations, from a phenomenological perspective, in the class of models that we examined in this paper. First of all, a  $F_{pq}^1$  and  $V_{pq}^1$  could not give rise to desired coupling as they reside on the same orbifold plane. Secondly, due to only employing the  $T_i$ ,  $i = 1, 2, 3$ , in the basis we do not allow for shifts around the 6 circles of the internal  $T^6$  and there is a multiplicity factor of 4 attached to each sector, making 3 generation models impossible. However, our focus here is on the introduction of SAT and SMT solving and the illustration of its application, while the application of SAT and SMT solving to more realistic constructions is left for future work. It is evident, however, that as we seek to construct satisfiability criteria for non-supersymmetric [35] and more detailed phenomenological string models, for which the SAT and SMT algorithms are particularly well suited, these algorithmic approaches are of immense interest and utility.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We would like to thank Viktor Matyas for helpful discussions. The work of BP is supported in part by STFC grant ST/N504130/1.

The work of SS and DW is supported in part by EPSRC grant EP/P020909/1.

## References

- [1] L. Ibáñez, A. Uranga, Cambridge University Press, 2012;  
K.S. Choi, J.E. Kim, Lect. Notes Phys. 696 (2006) 1–406, <https://doi.org/10.1007/b11681670>.
- [2] I. Antoniadis, C. Bachas, C. Kounnas, Nucl. Phys. B 289 (1987) 87;  
H. Kawai, D.C. Lewellen, S.H.-H. Tye, Nucl. Phys. B 288 (1987) 1;  
I. Antoniadis, C. Bachas, Nucl. Phys. B 298 (1988) 586.
- [3] A.E. Faraggi, Phys. Lett. B 326 (1994) 62;  
E. Kiritsis, C. Kounnas, Nucl. Phys. B 503 (1997) 117;  
A.E. Faraggi, S. Forste, C. Timirgaziu, J. High Energy Phys. 0608 (2006) 057;  
R. Donagi, K. Wendland, J. Geom. Phys. 59 (2009) 942;  
P. Athanasopoulos, A.E. Faraggi, S. Groot Nibbelink, V.M. Mehta, J. High Energy Phys. 1604 (2016) 038.
- [4] See e.g. and references therein: Lara B. Anderson, Mathis Gerdes, James Gray, et al., arXiv:2012.04656, 2020;  
J. Carifio, J. Halverson, D. Krioukov, B.D. Nelson, J. High Energy Phys. 1709 (2017) 157;  
A. Mutter, E. Parr, P.K.S. Vaudrevange, Nucl. Phys. B 940 (2019) 113;  
Y.H. He, Phys. Lett. B 774 (2017) 564;  
S. Abel, J. Rizos, J. High Energy Phys. 08 (2014) 010.
- [5] For a comprehensive review see e.g. and references therein: F. Ruehle, Phys. Rep. 839 (2020) 1.
- [6] Nicola Paoletti, et al., Computer Aided Verification, Springer International Publishing, 2014.
- [7] Luca Pulina, Armando Tacchella, AI Commun. 25 (2) (2012) 117–135;  
H. Kautz, D. McAllester, B. Selman, in: Proc. of KR-96, Boston, MA, 1996.
- [8] B. Konev, A. Lisitsa, [https://doi.org/10.1007/978-3-319-09284-3\\_17](https://doi.org/10.1007/978-3-319-09284-3_17), 2014.
- [9] C. Lüders, arXiv:2004.07058, 2020.
- [10] A.E. Faraggi, V.G. Matyas, B. Percival, arXiv:2010.06637.
- [11] A.E. Faraggi, V.G. Matyas, B. Percival, arXiv:2011.12630.
- [12] A.E. Faraggi, J. Rizos, H. Sonmez, Nucl. Phys. B 927 (2018) 1.
- [13] A.E. Faraggi, G. Harries, J. Rizos, Nucl. Phys. B 936 (2018) 472.
- [14] A.E. Faraggi, G. Harries, B. Percival, J. Rizos, Nucl. Phys. B 953 (2020) 114969.
- [15] A.E. Faraggi, J. Rizos, H. Sonmez, Nucl. Phys. B 886 (2014) 202;  
H. Sonmez, Phys. Rev. D 93 (2016) 125002.
- [16] B. Assel, et al., Phys. Lett. B 683 (2010) 306;  
Nucl. Phys. B 844 (2011) 365;  
C. Christodoulides, A.E. Faraggi, J. Rizos, Phys. Lett. B 702 (2011) 81.
- [17] L. de Moura, N. Bjørner, in: Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS '08), 2008.
- [18] A. Cimatti, A. Griggio, R. Sebastiani, J. Artif. Intell. Res. 40 (1) (2001) 701–728.
- [19] I. Florakis, J. Rizos, Nucl. Phys. B 913 (2016) 495.
- [20] A.E. Faraggi, Eur. Phys. J. C 79 (2019) 703.
- [21] A.E. Faraggi, V.G. Matyas, B. Percival, Eur. Phys. J. C 80 (2020) 337.
- [22] A.E. Faraggi, V.G. Matyas, B. Percival, Nucl. Phys. B 961 (2020) 115231.
- [23] A.E. Faraggi, V.G. Matyas, B. Percival, arXiv:2011.04113.
- [24] S. Cook, <https://doi.org/10.1145/800157.805047>, 1971;  
L. Levin, Problems of information transmission 9 (3) (1973) 115–116 (in Russian).
- [25] R. Impagliazzo, R. Paturi, <https://doi.org/10.1109/CCC.1999.766282>, 1999.
- [26] A.R. Bradley, Z. Manna, The Calculus of Computation: Decision Procedures with Applications to Verification, Springer, Oct. 2007, isbn: 3-540-74112-7;  
D. Kroening, O. Strichman, Decision Procedures, Springer, 2008, isbn: 978-3-540-74104-6.
- [27] I. Antoniadis, J. Ellis, J. Hagelin, D.V. Nanopoulos, Phys. Lett. B 231 (1989) 65.
- [28] A.E. Faraggi, D.V. Nanopoulos, K. Yuan, Nucl. Phys. B 335 (1990) 347;  
A.E. Faraggi, Phys. Rev. D 46 (1992) 3204;  
G.B. Cleaver, A.E. Faraggi, D.V. Nanopoulos, Phys. Lett. B 455 (1999) 135.
- [29] I. Antoniadis, G.K. Leontaris, J. Rizos, Phys. Lett. B 245 (1990) 161;  
G.K. Leontaris, J. Rizos, Nucl. Phys. B 554 (1999) 3.
- [30] A.E. Faraggi, Phys. Lett. B 278 (1992) 131;  
Nucl. Phys. B 387 (1992) 239;  
A.E. Faraggi, E. Manno, C.M. Timirgaziu, Eur. Phys. J. C 50 (2007) 701.
- [31] G.B. Cleaver, A.E. Faraggi, C. Savage, Phys. Rev. D 63 (2001) 066001;  
G.B. Cleaver, D.J. Clements, A.E. Faraggi, Phys. Rev. D 65 (2002) 106003.
- [32] A.E. Faraggi, C. Kounnas, S.E.M. Nooij, J. Rizos, arXiv:hep-th/0311058;  
Nucl. Phys. B 695 (2004) 41.
- [33] A.E. Faraggi, C. Kounnas, J. Rizos, Phys. Lett. B 648 (2007) 84;  
Nucl. Phys. B 774 (2007) 208;  
Nucl. Phys. B 799 (2008) 19.
- [34] T. Catelin-Julian, et al., Nucl. Phys. B 812 (2009) 103;  
C. Angelantonj, A.E. Faraggi, M. Tsulaia, J. High Energy Phys. 1007 (2010) 314;  
A.E. Faraggi, I. Florakis, T. Mohaupt, M. Tsulaia, Nucl. Phys. B 848 (2011) 332.
- [35] L.J. Dixon, J.A. Harvey, Nucl. Phys. B 274 (1986) 93;  
L. Alvarez-Gaume, P.H. Ginsparg, G.W. Moore, C. Vafa, Phys. Lett. B 171 (1986) 155;  
H. Kawai, D.C. Lewellen, S.H.H. Tye, Phys. Rev. D 34 (1986) 3794;  
H. Itoyama, T.R. Taylor, Phys. Lett. B 186 (1987) 129;  
K.R. Dienes, Phys. Rev. Lett. 65 (1990) 1979;  
Phys. Rev. D 42 (1990) 2004;  
S. Abel, K.R. Dienes, E. Mavroudi, Phys. Rev. D 91 (2015) 126014;  
M. Blaszczyk, et al., J. High Energy Phys. 1510 (2015) 166;  
J.M. Ashfaque, et al., Eur. Phys. J. C 76 (2016) 208;  
H. Itoyama, S. Nakajima, Nucl. Phys. B 958 (2020) 115111;  
S. Parameswaran, F. Tonioni, arXiv:2007.11333;  
I. Basile, S. Lanza, arXiv:2007.13757.