



Testing swampland conjectures with machine learning

Nana Cabo Bizet^{1,a}, Cesar Damian^{2,b} , Oscar Loaiza-Brito^{1,c}, Damián Kaloni Mayorga Peña^{3,d},
J. A. Montañez-Barrera^{2,e}

¹ Departamento de Física, Universidad de Guanajuato, Loma del Bosque No. 103 Col. Lomas del Campestre, C.P 37150 Leon, Guanajuato, Mexico

² Departamento de Ingeniería Mecánica, Universidad de Guanajuato, Carretera Salamanca-Valle de Santiago Km 3.5 + 1.8 Comunidad de Palo Blanco, Salamanca, Mexico

³ Mandelstam Institute for Theoretical Physics, School of Physics, NITheP, and CoE-MaSS, University of the Witwatersrand, WITS, Johannesburg 2050, South Africa

Received: 8 July 2020 / Accepted: 6 August 2020 / Published online: 24 August 2020

© The Author(s) 2020

Abstract We consider Type IIB compactifications on an isotropic torus T^6 threaded by geometric and non geometric fluxes. For this particular setup we apply supervised machine learning techniques, namely an artificial neural network coupled to a genetic algorithm, in order to obtain more than sixty thousand flux configurations yielding to a scalar potential with at least one critical point. We observe that both stable AdS vacua with large moduli masses and small vacuum energy as well as unstable dS vacua with small tachyonic mass and large energy are absent, in accordance to the refined de Sitter conjecture. Moreover, by considering a hierarchy among fluxes, we observe that perturbative solutions with small values for the vacuum energy and moduli masses are favored, as well as scenarios in which the lightest modulus mass is much smaller than the corresponding AdS vacuum scale. Finally we apply some results on random matrix theory to conclude that the most probable mass spectrum derived from this string setup is that satisfying the Refined de Sitter and AdS scale conjectures.

1 Introduction

One of the main aims of string theory is the construction of realistic effective theories with a small cosmological constant Λ within the perturbative regime. Motivated by the recent series of conjectures around the construction of de Sitter (dS) vacua and inflationary conditions [1–8] (see also [9–17]), the

question about a possible microscopic origin of Λ has lately received an increasing attention [18]. It is then worthwhile to focus on specific flux configurations which can be related to effective models with small energy values at extremal points in moduli space.

In this context, one would be tracing back the origin of a small Λ to some well-identified features of flux configurations. This would certainly be very interesting since fluxes drive many important physical phenomena, such as: supersymmetry breakdown, symmetry breaking, axion monodromy inflation and F-term monodromies. As it was observed in [19–22], all these expected and desirable features naturally arise in the so called flux-scaling scenarios, where fluxes play a role in fixing the values of the vacuum energy at extrema of the potential.

A promising scenario as they are, flux compactifications must obey the quantum gravity conjectures if one hopes to complete these models in the UV regime. In this work we focus on the so-called refined de Sitter conjecture (RdSC) which states that the construction of a stable dS vacuum is excluded from a consistent quantum gravity theory (including string compactifications). More specifically the RdSC establishes a bound of the form

$$\frac{\min \nabla_i \nabla_j V}{V} \leq -c', \quad (1)$$

where V is a given effective scalar potential and i, j represent index coordinates in field space and c' is a given constant parameter. Besides the exclusion of stable de Sitter, the bound also implies that some apparently plausible AdS vacua must be discarded as well, depending on the actual value of the constant c' as shown in Fig. 1. The bound defines a line with a slope determining the value of c' for some specific model, i.e., the upper bound on the quotient between the minimum

^a e-mail: nana@fisica.ugto.mx

^b e-mail: cesar.damian@ugto.mx (corresponding author)

^c e-mail: oloaiza@fisica.ugto.mx

^d e-mail: damian.mayorgapena@wits.ac.za

^e e-mail: ja.montanezbarrera@ugto.mx

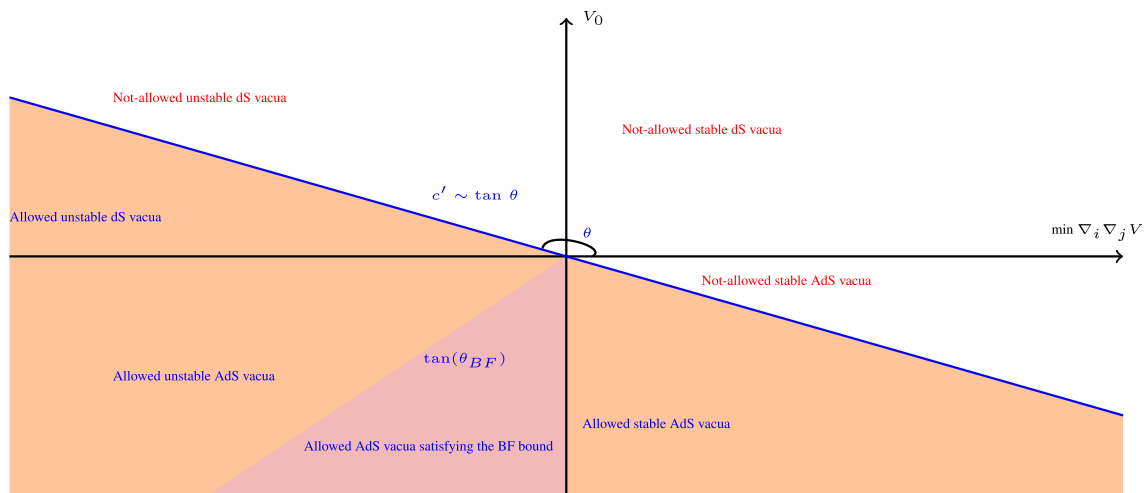


Fig. 1 Agreement of given effective model with the refined de Sitter conjecture defines 6 different zones. Three of them belong to the Swampland. The angle θ defines the slope of the line dividing the regions

of the Swampland from the Landscape. In pink we represent AdS vacua satisfying the BF bound defined by $\tan(\theta_{BF}) = -2/3$

mass squared and the value of V at an extremum for the potential. In Fig. 1 we can distinguish six different zones depending on whether the corresponding vacuum energy is positive or negative and on whether the vacuum is stable or not. As it was mentioned already, some AdS regions are excluded as well, in particular, stable AdS vacua with small energy and large moduli masses. The same is true for unstable dS regions with a large vacuum energy and a small tachyonic mass. Throughout this work we refer to unstable and stable vacua as solutions with or without tachyons respectively, for both dS and AdS. Notice however, that for AdS vacua, the instability can be alleviated provided the tachyon mass squared is above the Breitenlohner–Friedmann (BF) bound [23], i.e., for $m_{\text{tachyon}}^2 \geq -3|V_0|/2$. The region satisfying the BF bound is highlighted in Fig. 1.

We concentrate on a simple well studied model consisting on a Type IIB compactification on an isotropic torus in presence of orientifold 3-planes, threaded by the usual Ramond–Ramond (RR) and Neveu–Schwarz–Neveu–Schwarz (NS–NS) 3-form fluxes and by non-geometric (nG) fluxes as well [24–30] (see Appendix A). The scalar potential has three complex scalar fields: the complex structure (U), the axio-dilaton (S) and the Kähler modulus (T). The simplicity of this model lets us implement an algorithm to find as many extrema as possible for the scalar potential. One of the goals of the present work is to produce consistent and adequately quantized flux configurations. This, in order to obtain a reasonable sample of scenarios where one would be able to test whether or not the stable AdS and non-stable dS zones are excluded, in accordance or disagreement with the RdS conjecture.

We classify different flux configurations according to the features of the scalar potential at the extremum under consideration. For that purpose we use an artificial neural network¹ (ANN), by means of which we are able to classify more than sixty thousand different flux configurations and some relevant features of the corresponding vacua. There is however an important caveat here. It is necessary to provide the ANN with concrete examples to be able to identify certain patterns among the different fluxes, which in turn would lead to some stable or unstable extremal point in moduli space. This is the reason to use genetic algorithms previous to adapting the neural network [9, 10, 31, 39, 42, 43]. Since there is not a single example of a stable dS, it is possible that the network does not identify such cases and in consequence it will not learn how to construct them. So, we expect not to find dS stable extrema. Observe that this fact is only a consequence of our algorithm and it is not reflecting a general feature of our compactification model. However, we are not restricting the possible AdS vacua to encounter since there are plenty of examples of unstable and stable AdS extrema. By looking for them employing the neural network, we expect to reproduce all possible situations. Therefore, this is a fruitful zone to check for consistency with the RdS conjecture, and we find that those zones excluded by it are indeed absent in our classification, suggesting the validity of the conjecture or the quantum gravitational consistency of the considered setup.

Based on recent results [13, 25, 44], in which the presence of hierarchical values on fluxes induces a natural hierarchy on moduli masses, for which there are concrete (supersym-

¹ Implementation of different computational methods in high-energy physics research has increased notably in the last few years. See e.g. [31–41].

metric and non-supersymmetric) vacuum solutions with a small value for the cosmological constant Λ , we contemplate the possibility that hierarchical flux configurations lead to scenarios with small values for the vacuum energy. We observe that indeed, the values of the scalar potential at its minimum are smaller than one when the flux configuration possesses a hierarchy among their integer values. In this sense we suggest that a possible microscopic explanation for a small Λ in a quantum gravity theory such as string theory, might rely on specific features of the flux configuration. Moreover, we find that the smaller the string coupling, the higher the probability to find a vacuum solution with a small vacuum energy, suggesting that for the most probable scenarios, $\Lambda \sim \exp(-Re(S))$. This is another highlight of the use of hierarchical flux configurations.

We also report that, by considering hierarchical fluxes, the ANN classification shows that there is a higher probability for the vacuum solutions to show a spectrum in which the minimal stable modulus mass is greater than the scale of the AdS vacuum. These vacua, in accordance to the AdS scale conjecture cannot be uplifted to a stable dS vacuum.

In order to sustain the above observations on a more solid basis, we compare the spectra of critical values obtained from the mass matrix, with the spectra of a Gaussian orthogonal ensemble (GOE) with a mean-value μ and standard deviation σ . We observe that the mass matrix possesses similar characteristics as a GOE namely, the probability for the mass matrix eigenvalues to be non-negative coincides with that derived from a GOE. Thus, we use the spectral results obtained from random matrix theory applied to the squared mass eigenvalues to find that:

- Probability to find an unstable critical point is 10^6 times higher than finding a stable one.
- 80% of all generated flux configuration fulfilling string constraints as Tadpole cancelation and Bianchi identities do not exhibit a hierarchy among their values, pointing out the fact that it is not likely to obtain a hierarchical flux configuration from random selection.

Although the last point suggests that it is very unlikely to encounter a flux configuration with a hierarchy, if one departs from a hierarchical flux configuration, the probability to obtain an effective theory at the extremum of the scalar potential with some desired physical properties increases. This is:

- 70% of the constructed vacua are within the perturbative regime.
- Among all vacua (stable critical points), 40% of them have an (absolute) energy value smaller than unit.
- In 80% of all AdS vacua, the lightest moduli mass is larger than the (absolute value of) vacuum scale.

Therefore, although all generated vacua seems to satisfy the RdS conjecture we find that by restricting the construction of these simple models to hierarchical flux configurations, we increase the probability for the effective models to be in the perturbative regime and to fulfill scale conjecture as well. This suggests, at least for these simple toroidal models, that the source of the Swampland constraints could rely on specific features of flux configurations as the hierarchical values among them.

Our work is organized as follows: In Sect. 2 we describe generically and in simple terms the implementation of the artificial neural network coupled to the genetic algorithm. Technical issues concerning the structure of an ANN as well as a basic example are given in Appendix A. In Sect. 3 we discuss the numerical results obtained by implementing the scan over random and hierarchical flux configurations. Finally in Sect. 4 we present our concluding remarks. The physical description of the Type IIB flux compactification setup is presented in Appendix B. Similarly, a toy example illustrating the possibility to have small vacuum energy values at an extremum of the scalar potential is presented in Appendix B as well.

2 Classification of vacua and search for extrema of the potential

We are interested in classifying vacua constructed from different flux configurations. This is done in order to identify flux patterns which could lead to some desired particularities, such as: stability, a small value for the cosmological constant or the existence of a dS critical point. For that we shall use and implement an artificial neural network (ANN).²

The ANN architecture proposed in this paper is that of a pattern recognition feedforward network organized in three clusters of neurons: The input layer with 10 neurons, the hidden layer consisting of 12 neurons for the case of free-tachyon classification and 23 for the case of positive vacua classification, and the output with 1 neuron. The activation functions are chosen to be the hyperbolic tangent sigmoid transfer function. In the input we encode the integral values for the flux parameters consisting on a set of fluxes satisfying all string constraints, namely the Tadpole cancellation condition and Bianchi identities. In our case we consider *non-geometric fluxes* as well.

As previously mentioned, we concentrate on an isotropic toroidal flux compactification (see Appendix B). Hence we consider 4 integers parameterizing the R–R sector fluxes f

² This section deals with some technical details and numerical analysis (see Appendix A for details). For the reader interested in our conclusions on the construction of string derived effective models we suggest to go directly to Sect. 3.

(with components f_i , $i = 1 \cdots 4$), 4 integers for the NS–NS sector fluxes as well h (with components h_j , $j = 1 \cdots 4$), and 6 for the non-geometric (nG) fluxes b (with components b_k , $k = 1 \cdots 6$), adding up to the 14 nodes of the input. The output is made of those vacuum solutions of the scalar potential constructed from the corresponding flux compactification. Extra criteria must be added to stimulate the ANN searching. In our case we shall analyze two different criteria to stimulate the ANN, namely by looking for stable or dS critical points.

The use of the ANN requires a controlled training as a first step. The training consists on feeding the ANN with different flux configurations for which we know the existence of critical points as well as their corresponding features, such as vacuum stability and the value of the scalar potential at the critical point. The training data is obtained by randomly generating different flux configurations satisfying the Tadpole cancellation condition and Bianchi identities. We were able to generate about 40,000 different configurations using Mathematica codes. After that, we implemented a genetic algorithm (GA) in order to compute the moduli VEVs at which the scalar potential has a critical point, the corresponding scalar potential value at that point as well as its corresponding Hessian matrix (determining the stability).

The training process serves to optimize the network parameters (weights and biases) upon stepwise minimization of a certain objective function, which we have chosen to be the mean standard error (MSE, see Eq. A.2). For this purpose, the training data is divided into three randomly selected groups as follows: 80% of the data is used for the ANN training, 10% for validation (to avoid overfitting on the training data), and 10% for a posterior test (to avoid overfitting on the validation data).³ Thus it is expected from it to perform well beyond the training data (it might even be able to identify possible patterns relating the flux configuration with the existence of specific extrema of the resulting potential as well as the features of the potential at those critical points).

Once the ANN is trained we proceed to feed it with a variety of flux configurations. The ANN tells us which of them allow or not for the existence of some critical point with some required feature, i.e., it classifies the flux configurations into two groups according whether they fulfill the selected criteria or not. We confirm the results given by the ANN

by implementing the GA and calculating specific values at the critical points in case we have them. A flow map of our approach is shown in Fig. 2. More in detail, the sketch of our procedure is as follows:

1. We collect the training data. These are flux configurations fulfilling Tadpole cancellation condition and Bianchi identities. We generate nearly 40,000 different configurations. There are two training processes depending on the type of training data:
 - (a) CASE A: training with random fluxes. The NS-NS, RR and nG fluxes are picked at random.
 - (b) CASE B: training with hierarchical fluxes. Fluxes used for training are no longer chosen at random. Instead, the flux values in one of the closed sectors are higher than the rest, e.g. integer valued NS–NS fluxes are between one and four orders of magnitude larger than R–R and nG fluxes.
2. We use our trained network as a classifier for nearly 1.4 million flux configurations. In order to find some interesting statistics we have also selected two different criteria for the outcome data:
 - (a) Criterion I: a stable critical point for the scalar potential. This means that the ANN looks for patterns on the flux configuration such that the scalar potential has a minimum. This can be either AdS or dS.⁴
 - (b) Criterion II: a dS critical point. This means that the network is asked to determine whether a given flux configuration exhibits a dS critical point, regardless of whether it is a maximum, a minimum or a saddle point.
3. We implement a genetic algorithm (GA) to compute specific values for the vacua on the classified flux configurations.

In the following we describe our results by dividing them in terms of the flux configuration input set.

2.1 Case A: random fluxes

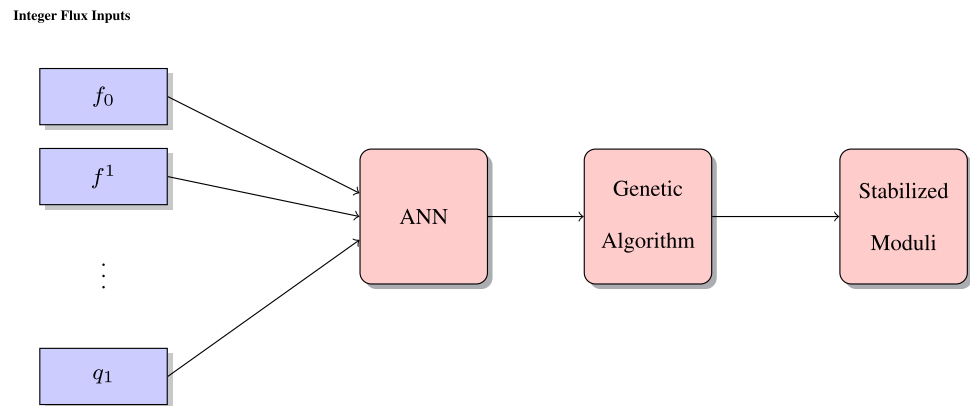
ANN training

After randomly generating 40,000 sets of fluxes satisfying the tadpoles and the Bianchi identities, we implement a GA to determine which of them contain critical points. We find 4034 critical points out of which there are 298 AdS solutions without tachyons, 139 dS with Tachyons and the remainder

³ Once the network is trained, the confusion matrix shows us that the ANN was able to correctly characterize the data in 98.6% of the cases. For 75.2% of the correct classification the output was positive (no tachyons in the spectra) and for 23.3% of the correct classification the output was negative (there was at least one tachyon). Besides, the ANN made a wrong classification of the positive answer by 0.8% (the ANN predicted at least one tachyon where there was no tachyon in the spectra) and it made a wrong classification with negative answer with an error of 0.6% (the ANN predicted no tachyons where there was at least one tachyon in the spectrum).

⁴ Minkowski vacua are excluded since, by construction the ANN is not trained to obtain such vacua.

Fig. 2 Flow chart of the vacua search procedure. One starts with a given flux configuration as an input for the neural network. The outcome is whether or not the fluxes under consideration lead to a scalar potential in the effective theory with critical points. If the outcome is positive, then one employs the genetic algorithm in order to find the critical point(s) and the corresponding field values at which the various moduli get fixed



are tachyonic AdS. The results are used to train a neural network classification which assigns a value, e.g., 1 or 0 as an output, depending on whether or not a given property is satisfied by the flux under consideration.

As mentioned above we have selected two different cases according to the feature we want the ANN to find: (1) A stable critical point, this is, a minimum regardless the value of the vacuum energy or (2) A critical point with a positive value of the scalar potential at such point. This would be a dS critical point, regardless its stability. For the first case, the ANN classifies flux configurations into three groups: Those generating a scalar potential with a stable critical point, those generating a scalar potential with unstable critical points and finally those generating a scalar potential without critical points. Similarly, for the second criterion, the classification of fluxes after feeding the ANN consists on a group of fluxes generating a dS extremal point, those with an AdS critical point, and finally, those generating a scalar potential without a critical point.

Results

After training the ANN we feed it with nearly a million different flux configurations satisfying the Tadpole and Bianchi constraints. In the following we summarize our findings.

Criterion I. Stable critical points Out of the roughly one million cases in the input, the ANN selects 66,000 sets of fluxes as candidates to generate a scalar potential with a minimum. In order to verify this, we use the GA and find that out of the 66,000 configurations, there are 20,779 with critical points and only 9872 without tachyons (see footnote 3). It is interesting to compare with the original training data, out of 40,000 flux configurations we obtained 298 stable critical points, a naive estimate can lead us to the expectation of 7450 stable critical points had we simply run the AG over one million flux configurations. Employing the ANN coupled to the GA we obtain an amount of minima in the same order of magnitude (slightly higher). From this observation we conclude that besides the advantage of the ANN + GA being much

less time consuming than the GA alone, we obtain roughly the same quality in the final outcome, therefore making this approach very suited for Landscape studies. The distribution of minima is presented in Fig. 3a. Finally let us recall that no dS minimum was found, although there are many unstable dS extremal points.

Criterion II. dS extremal points For this case the ANN favored a total of 50,000 sets of fluxes as possible candidates to contain a dS extremum. The GA confirms that out of those 50,000, only 4944 different flux configurations generate a scalar potential with an extremal point. Moreover, only 140 of them lead to a minimum, i.e., an extremal point free of tachyons. For all of the stable minima we find that they occur at negative values of the scalar potential, i.e., they are AdS minima. The rest of them correspond to unstable 2744 dS and 2200 AdS extremal points. The results of this classification are shown in Fig. 3b.⁵ Notice that contrary to training with Criterion I (9872 cases without tachyons), the number of stable vacua found using Criterion II (140 cases without tachyons) is less than the one obtained by the use of GA on aleatory fluxes (298 cases without tachyons).

It is important to emphasize that from the total set of critical points, no stable dS vacua was found no matter what criterion we have used. For instance, with Criterion I, the number of dS (180 cases) is considerably smaller than those obtained in Criterion II (2744 cases). Also with Criterion II the number of dS critical points increases as expected, in spite of an observed overall decrease in the number of stable points. This numerical analysis shows a correlation between the presence of tachyons and the number of actual dS critical points as suggested by the RdS Swampland Conjecture, at least for the isotropic torus with fluxes.

The ANN flux classification improves our capacity to find vacua and in consequence to explore the String Landscape or the Swampland. This follows from the analysis plotted in Fig. 4 where we show the number of vacua, stable or not,

⁵ Here we are analyzing only stable critical points out of those generated by the ANN through Criterion II.

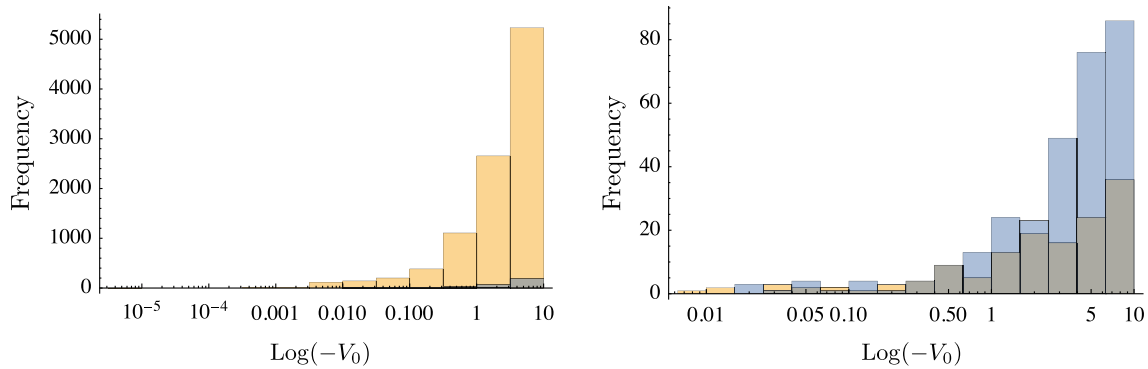


Fig. 3 Histogram of the stable vacua with random fluxes generated by the GA (blue bars) and by the ANN + GA (yellow bars). Intersection of blue and yellow bars appears as gray bars. Left: criterion I. Right:

criterion II. This data represents the first step in the training of the ANN and corresponds only to the free tachyon vacua

versus the value of the scalar potential at the critical point. We notice that for the case of AdS, the number of vacua is increased by the use of the ANN compared to those obtained by GA for the Case I. However the same is not true for Case II. On the other hand, the number of dS vacua increases by the use of the ANN in both cases, although neither of them contain a stable dS vacuum. See Fig. 4 for more details. By looking at the order of magnitude on the number of vacua found by the use of the ANN, we conclude that Case I is much more efficient than Case II.

2.2 Case B: hierarchy on fluxes

ANN training

In this case the ANN is trained by an input of flux configurations with a clear hierarchy on their integer values. This hierarchy means that the integer values parameterizing one of the sectors, e.g. NS–NS, R–R or nG are between one and four orders of magnitude bigger than the fluxes in the other sectors. As in Case A, all flux configurations satisfy the usual constraints of tadpole cancellation and Bianchi identities with no D-branes. We explore 3 hierarchies among the fluxes: $f, h \gg b, h, b \gg f$ and $f, b \gg h$. The inequalities imply that all the flux components of one kind differ by at least one order of magnitude from all the flux components of the other kind (i.e. for the first type $\forall_{i,j,k} f_i \gg b_k, h_j \gg b_k$).

A hierarchy on the integer values associated to all fluxes in turn establishes a hierarchy on the masses associated to the modulus. This is, if we take for example the R–R fluxes to be larger than the others $f \gg h, b$, we expect in this model, that the complex structure modulus would be the heaviest modulus $M_U \gg M_S, M_T$ [44]. Next we write the expected hierarchies between the moduli masses that are obtained by setting one of the explored hierarchies among the fluxes:

$$\text{Case K : } f, h \gg b \rightarrow M_U, M_S \gg M_T,$$

$$\text{Case CS : } h, b \gg f \rightarrow M_S, M_T \gg M_U,$$

$$\text{Case AD : } f, b \gg h \rightarrow M_T, M_U \gg M_S. \quad (2)$$

The classification as in the previous case is done by demanding the ANN to identify flux configurations which generate a scalar potential with a stable critical point.⁶ Since the flux configuration presents a hierarchy, all the critical points are also related to a spectrum with a lightest moduli. Notice that for this case *we are not training the ANN to find critical points with a positive value for the scalar potential*. This follows from our experience in case A in which the dS criterion (Criterion II) did not produce much more vacua, as desired. In Fig. 5 and Fig. 6 stable and critical points are analyzed.

Results

The histograms obtained after ANN's classification are shown in Fig. 5. As observed, selecting a specific hierarchy on the flux configuration affects the distribution of vacua:

- If we take, for instance, both R–R and NS–NS larger than nG fluxes ($f, h \gg b$), we obtain the lightest mass to be that of the Kähler modulus M_T . In this case we notice a clustering of the number of stable vacua around a given value for the cosmological constant well below the peak obtained for randomly selected fluxes, with a mean value of the cosmological constant lower than its value on the randomly selected vacua.
- If we take the complex structure as the lightest modulus ($h, b \gg f$), we observe an increase in the number of stable AdS vacua with a greater dispersion. However, for the case in which the lightest modulus is the axio-dilaton ($f, b \gg h$), we do not notice an improvement on the

⁶ We have discarded an analysis for Criterion II once we conclude it is not efficient when applied to hierarchical fluxes.

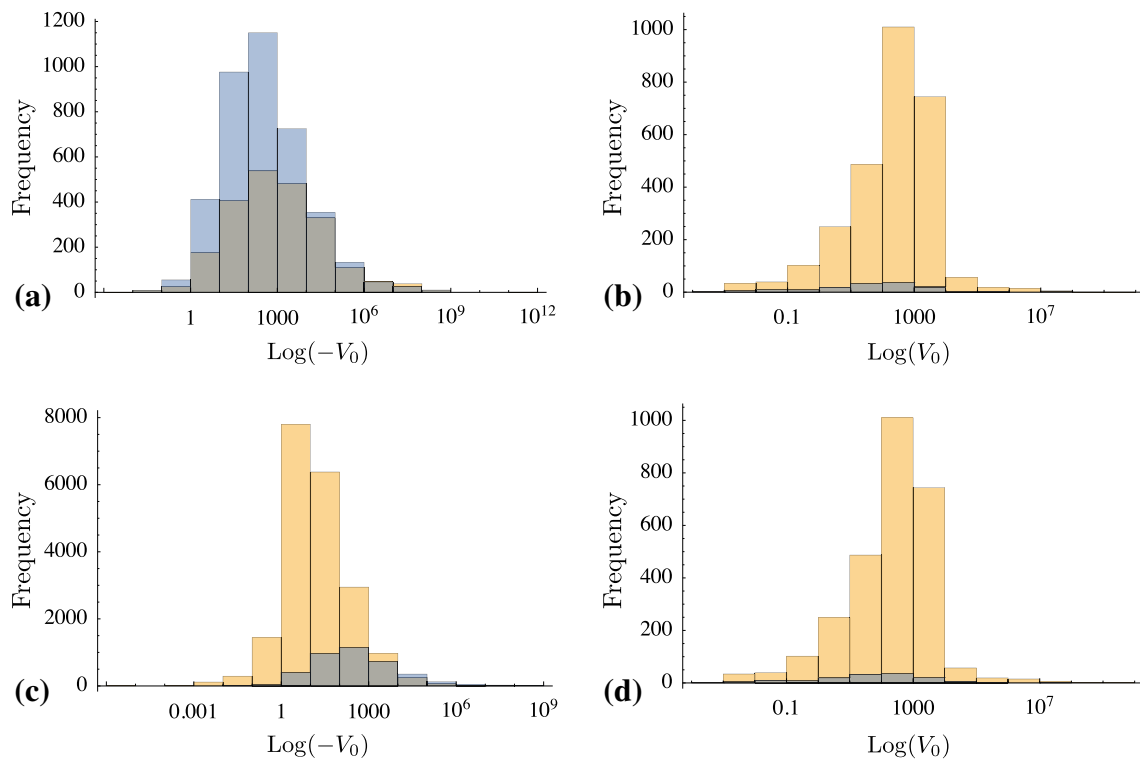


Fig. 4 Histogram of critical points with random fluxes obtained by the ANN + GA (yellow bars) and those randomly found by the GA (blue bars). Intersection appear as grey bars. **a** AdS vacua, criterion II, **b** dS vacua, criterion II, **c** AdS vacua, criterion I, **d** dS vacua, criterion I. Notice that the **a** differs from the others in that the amount of solutions

found by the GA alone and the ANN + GA are comparable, this reflects the fact that in this case the training set for the ANN was smaller. In contrast with Fig. 3, these histograms include the cases where at least exists one tachyonic state

amount of stable vacua in relation with a random flux configuration input.

In Fig. 6 we present the corresponding histograms related to different hierarchies on the moduli masses. Notice that for all cases the histograms seem to follow a normal distribution. Figures (a) and (b) indicate the distribution of vacua for the case in which the Kähler modulus is the lightest one $M_S, M_U \gg M_T$ (case K) against the value –negative or positive– of the scalar potential at that point. Figures (c), and (d) correspond to the case in which the axio-dilaton modulus is the lightest one $M_T, M_U \gg M_S$ (AD case); whereas Figures (e) and (f) refer to the case in which the complex-structure moduli is the lightest one $M_T, M_S \gg M_U$ (CS case).

The ANN classification shows a greater abundance of AdS critical points for this Case B than for Case A. Besides, the critical points for the K and CS cases respectively, have a mean value for the scalar potential lower than the value on the AD case. Conversely, the abundance of dS critical points is reduced in the K and CS cases in comparison with Case A.

3 Surveying the landscape of vacua

Upon correlation of different features for the vacua we obtained, we draw three important observations, which we present in order.

3.1 Perturbative regime is associated to a small minima of the scalar potential

A careful comparison of critical points shows that the largest values of the scalar potential at the corresponding critical point are related to non-perturbative regime ($Re S \ll 1$), and thus cannot be trusted. This can be seen in Fig. 7 where we have plotted all AdS and dS vacua (not necessarily stable) obtained by the ANN against the string coupling value (real part of the axio-dilaton at the critical point). We therefore observe that those flux configurations associated with very small values for the string coupling, i.e., describing an effective perturbative model, are related to small values for the cosmological constant, suggesting a relation of the form $\Lambda = \pm \exp(-Re S)$.

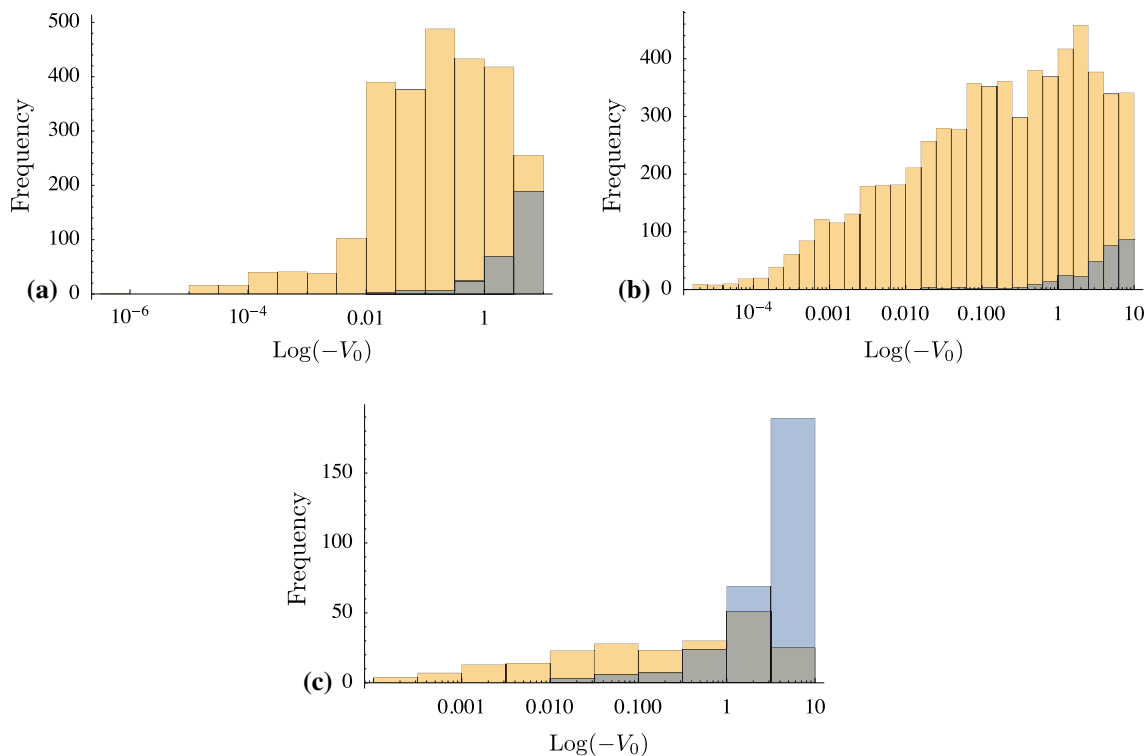


Fig. 5 Distributions of vacua obtained for hierarchical fluxes: **a** Kähler (K), **b** complex structure (CS) and **c** axio-dilaton (AD) as the lightest modulus. The critical points obtained by the ANN + GA are given in

yellow bars, and those randomly found by the GA are given in blue bars. This data represents the free tachyon spectrum classification of fluxes with hierarchy

3.2 Compatibility with the refined dS conjecture

The smallest eigenvalue of the $\nabla_i \nabla_j$ operator, denoted $\min \nabla_i \nabla_j V$, corresponds to the mass of the lightest modulus (which in the case of an unstable vacuum is tachyonic). Using the vacua distribution of the values of the potential at the critical point (Δ) versus the smallest modulus mass, we graphically observe that the vacua obtained populates only a half of that plane: essentially all the data lies below the line $V = -\frac{1}{c'} \min m^2 + c''$ for some for $c'' < 0$. As mentioned before, the slope of the line is related to c' parameter. In Fig. 8, vacua obtained in Case A are represented by red (Criterion I) and blue (Criterion II) points, while green (case K) and yellow (case CS) points represent those obtained in Case B.

From this analysis we conclude the following:

- The dispersion shows a structure in the vacua corresponding to straight lines. Different vacuum solutions in the same line belong to a set of fluxes related to a particular solution of Bianchi Identities and Tadpole conditions.
- Straight lines do not pass through the origin, instead they are displaced a small amount parametrized by c'' . This is

related to the fact that we look for solutions in which the second derivative for the scalar potential is different from zero.

- The hierarchies move the critical points towards the origin. This implies that, by demanding a hierarchy on the flux configuration input, the minima of the scalar potential becomes smaller, and according to our previous observation Sect. 3.1 a smaller string coupling is also obtained.
- Notice that this classification indeed reproduces the expected plot shown in Fig. 1, indicating not only the absence of stable dS vacua, but also the absence of some stable AdS and the presence of some unstable dS limited by a straight line.
- The vacuum points lie very close to the origin in Fig. 8 representing critical points with a small negative vacuum energy and with a small value for m_{ij}^2 , indicating that very close to the minimum there could be conditions on the scalar potential for which the AdS scale conjecture could be violated. It is then important to study how probable is to find such solutions.

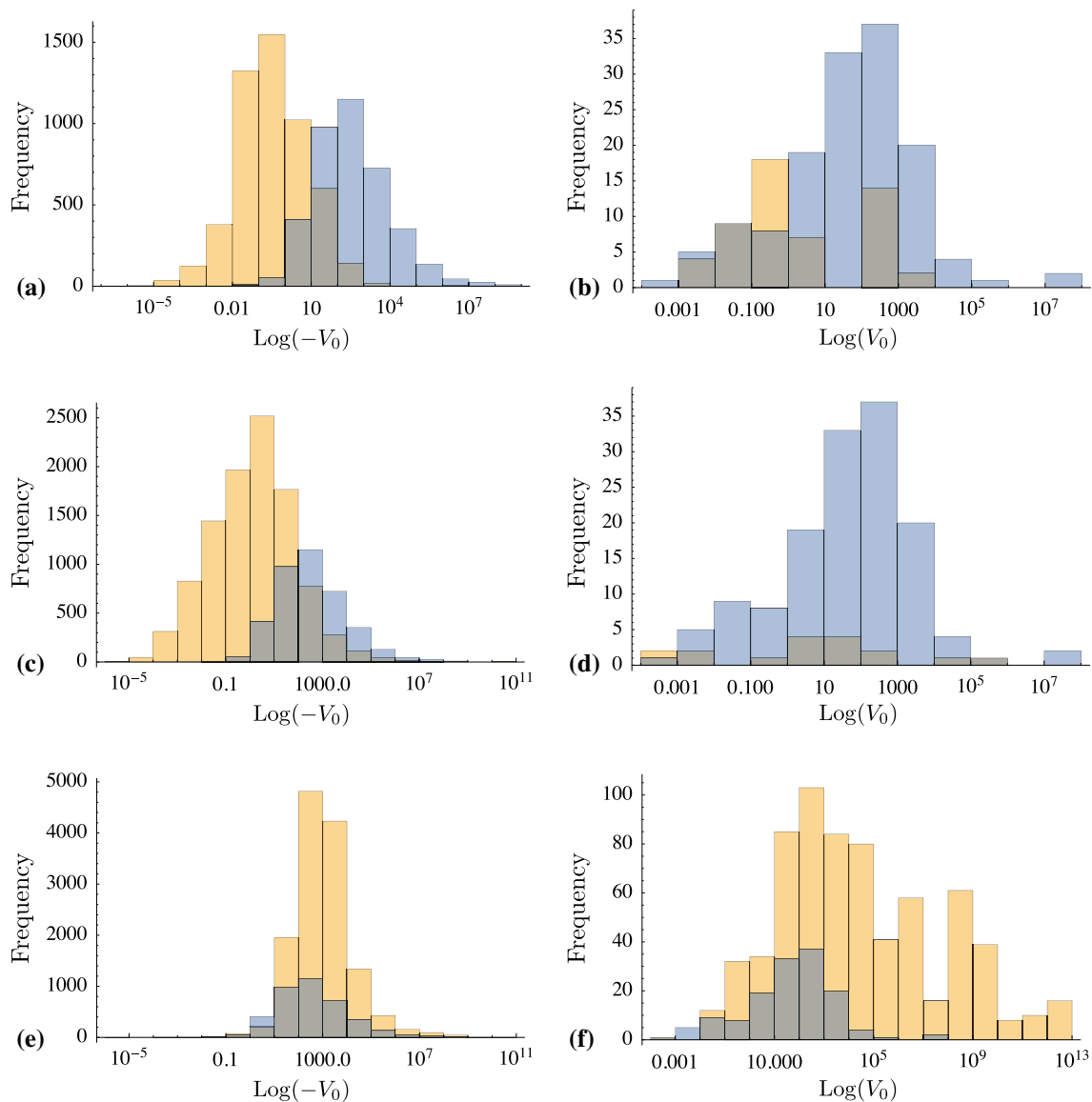


Fig. 6 Distributions of vacua for hierarchical fluxes. **a** AdS vacua, case K, **b** dS vacua, case K, **c** AdS vacua, case AD, **d** dS vacua, case AD, **e** AdS vacua case CS, **f** dS vacua, case CS. All dS vacua are unstable. The critical points obtained by the ANN+GA are given in yellow bars,

and those randomly found by the GA are given in blue bars. In contrast with Fig. 5 these histogram includes the vacua that contains at least one tachyon

3.3 AdS scale separation

Let us now classify the scale separation between stable AdS vacua Λ_{AdS} and the squared mass corresponding to the lightest modulus for all models constructed from a Case A configuration. This study allows us to directly see, as shown in Fig. 9, that by using a configuration of hierarchical fluxes it is more probable to find a hierarchy among moduli masses. Limited to our model we can say that the most probable scenario involves a maximum difference of masses of order of magnitude 3 where the difference is given by

$$\Delta m^2 = \max m^2 - \min m^2. \tag{3}$$

Notice that an exponential Δm^2 as present in a KKLT model is discarded in our case, probably due to the fact that we are considering a hierarchy among fluxes of an order of magnitude between 1 and 4 which in turn is a consequence of Bianchi and Tadpole constraints [25].

The AdS swampland scale conjecture asserts that it is not possible to separate the size of the AdS space and the mass of its lightest mode beyond a certain limit, this is

Fig. 7 Value of the scalar potential versus the string coupling at the critical point for all analyzed cases produced by the ANN + GA. Red and Blue points correspond to vacua classified in Case A, while yellow and green dots are related to Case B where a hierarchy of the flux configurations is assumed. It is observed that the smaller the string coupling the smaller the value for the cosmological constant Λ

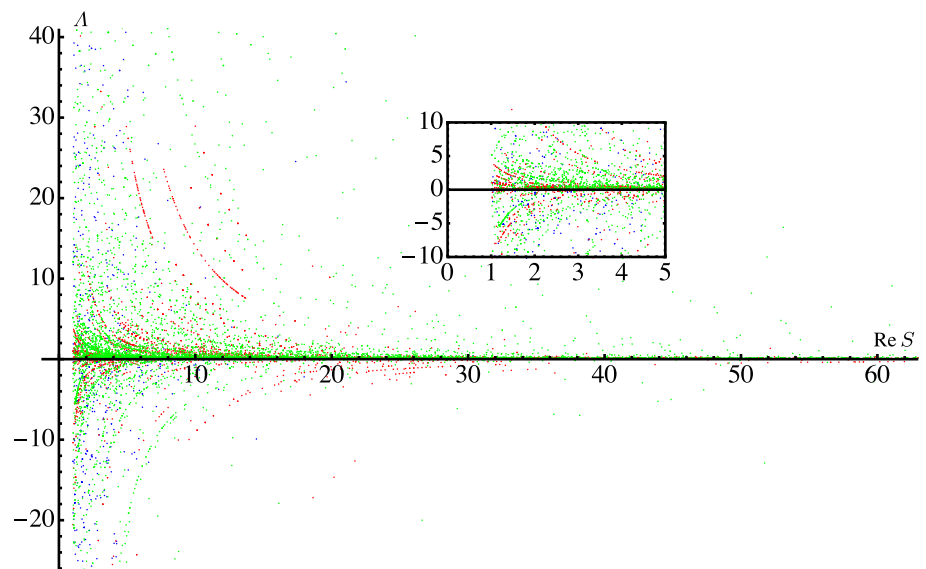
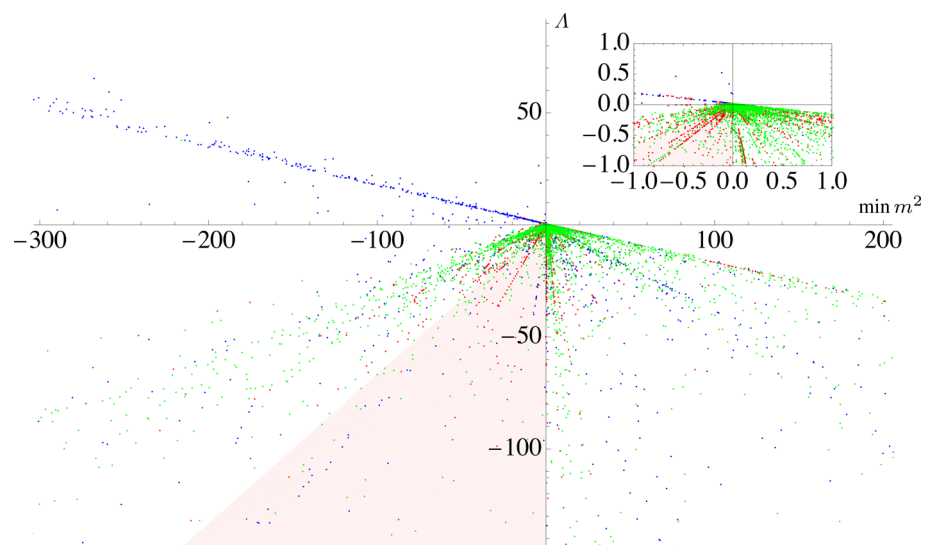


Fig. 8 Distribution diagram of the values of the extrema for the scalar potential RdSC. Red and blue points represent critical points obtained through the classification of randomly selected flux configurations (Case A) whereas green and yellow points correspond to critical points obtained by assuming hierarchical fluxes (Case B)



$$\left(\min m^2\right) L_{\text{AdS}}^2 \leq c, \quad (4)$$

where c is constant of order 1, and $L_{\text{AdS}}^2 \sim \Lambda_{\text{AdS}}^{-1}$. This conjecture is motivated from the point of view of the KKLT scenario, in the sense that any uplifting mechanism (from a supersymmetric stable vacua) does not destabilize the Kähler moduli as far as the potential well is parametrically narrow in comparison with the energy gap that needs to be filled by the uplifting mechanism. For the KKLT scenario, indeed this criteria is not fulfilled and thus it raises the question of its validity [45].

We analyze this conjecture (see Fig. 10) observing that both hierarchical and non-hierarchical fluxes lead to vacua with $\min m^2/\Lambda_{\text{AdS}}$ roughly of order 10 at most. Thus, as argued by [45] in most of the studied cases, any attempt to uplift the AdS vacua may destabilize the lightest modulus.

We also note a clustering of vacua for hierarchical fluxes for $\min m^2/\Lambda_{\text{AdS}} \leq 1$ compared to the non-hierarchical vacua which peak around $\min m^2/\Lambda_{\text{AdS}} \sim 7$. One can argue that all of the vacua obtained are in agreement with the AdS Scale Separation Conjecture, even when the ones with larger values of the quotient m^2/Λ_{AdS} might create some tension with it.

In summary, by assuming a hierarchy on the flux configuration among different sectors (NS–NS, R–R and NG) it is more probable for the generated vacua to have small values for the vacuum energy and a small value for the string coupling. Also, scenarios constructed with hierarchical fluxes exhibit a higher probability for the lightest modulus to be much smaller than the cosmological constant.

However, among all possible flux configurations, having a hierarchical one is not a likely scenario in a random set of flux configurations. By the use of random matrix theory we are in conditions to analyze this assertion.

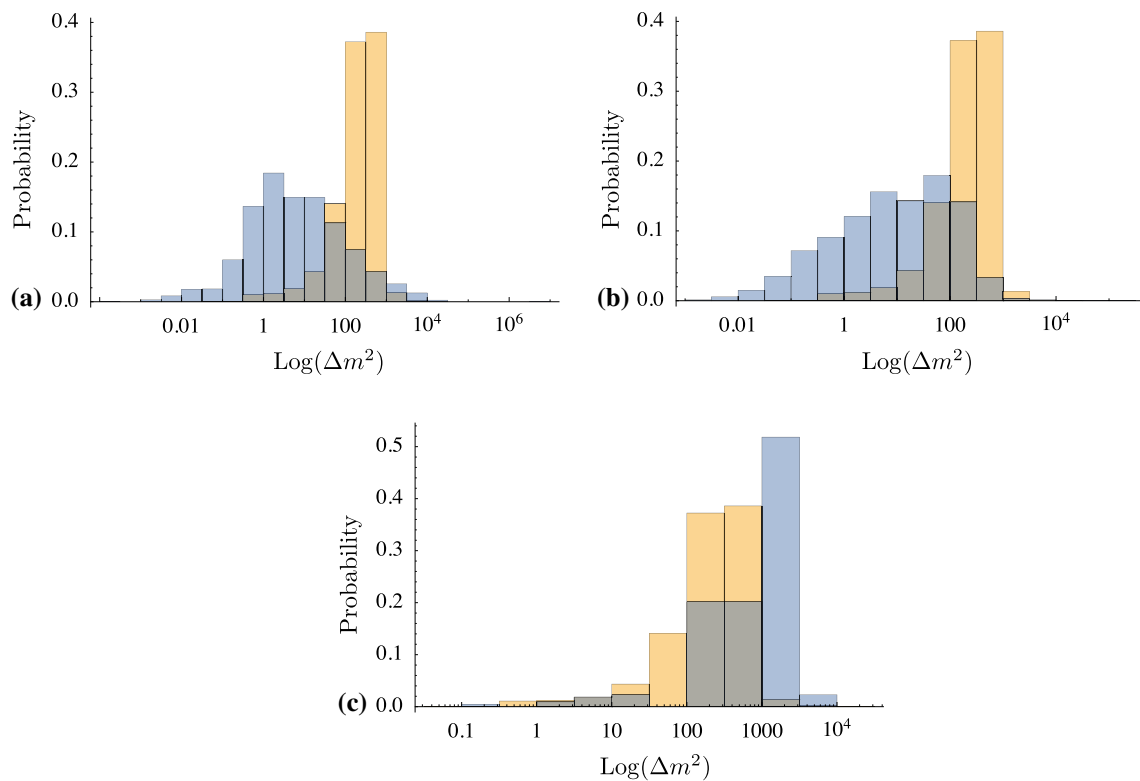


Fig. 9 Probability histograms for the scale separation between the heaviest and the lightest mode. Different histograms correspond to the case when a given modulus is the lightest: **a** Kähler (T), **b** complex

structure (U) and **c** axio-dilaton (S). Yellow bars refer to hierarchical flux configurations while blue bars refer to non-hierarchical ones

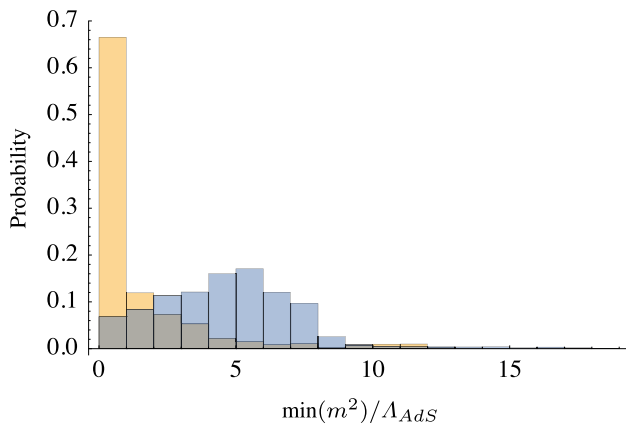


Fig. 10 Histogram of the scale separation between the lightest modulus and the corresponding value of the cosmological constant in Planck units $\min(m^2/\Lambda_{AdS})$. Yellow bars correspond to hierarchical fluxes while blue bars correspond to non-hierarchical fluxes

3.4 Relation to random matrix theory

The refined swampland criterion implies that for a dS vacuum the lowest eigenvalue of the mass matrix shall be negative and thus unstable. Indeed, if the RdSC is not satisfied, there exist an instability which leads to a breakdown of entropic

arguments [5]. This line of thought leads us to consider some sort of information/probabilistic feature of the dS conjecture and its refinement. Within this context, it was found [46] that using random functions as scalar potentials, the dS conjecture as well as the refined dS conjecture are the result of the most probable scenario. However, the connection with real vacua coming from dimensional reduction in string theory was not clear.

As already mentioned, after combining genetic algorithms and neural networks, we realize that there is a low probability of finding critical points. In Fig. 11 we present the histogram of the probability density distribution of the critical points obtained by all flux configurations. This distribution presents a mean value of 0 and a standard variation $\sigma = 0.35$. Besides, assuming identical and independent distributed (i.i.d.) entries coming from a Gaussian distribution, the probability density function (PDF) of the eigenvalue λ -spectrum of the mass matrix can be calculated by [47] (for a kindly check of the calculations see [48])

$$\rho(\lambda) = \frac{\mathcal{N}^{N/2-1}}{\sigma} \sum_{k=0} \exp\left[\frac{\lambda^2}{2\sigma^2}\right] (R_{2k}(\lambda)\Phi_{2k+1}(\lambda) - R_{2k+1}(\lambda)\Phi_{2k}(\lambda)), \tag{5}$$

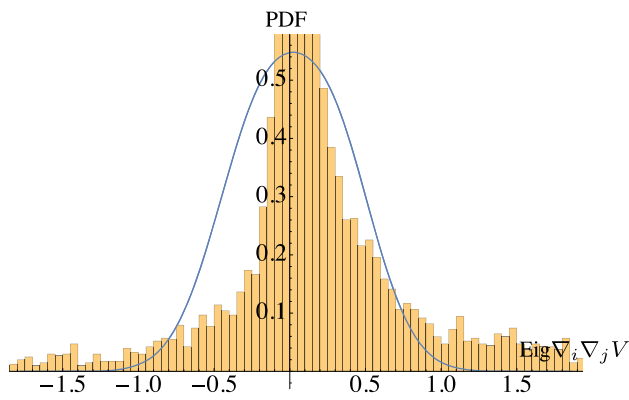


Fig. 11 Density distribution for all the eigenvalues of the mass matrix

where $\mathcal{N} = \frac{N! \hat{a}_N |2^{N/2-1}}{N \mathcal{Z}}$, \hat{a} is a constant that depends on N , \mathcal{Z} is a normalization factor analogous to the partition function (see [47]) and N is the rank of the mass matrix. The functions $\Phi_k(\lambda)$ are given by

$$\Phi_k(\lambda) = \frac{1}{\sigma} \int_{-\infty}^{\infty} d\lambda' R_k(\lambda') \exp\left[\frac{\lambda'^2}{2\sigma^2}\right] \text{sign}(\lambda - \lambda'), \quad (6)$$

with R_k being essentially Hermite polynomials:

$$\begin{aligned} R_{2k}(\lambda) &= \frac{\sqrt{2}}{\pi^{1/4} 2^k (2k)!!} H_{2k}(\lambda), \\ R_{2k+1}(\lambda) &= \frac{\sqrt{2}}{\pi^{1/4} 2^k (2k)!!} [-H_{2k+1}(\lambda) + 4k H_{2k+1}(\lambda)]. \end{aligned} \quad (7)$$

Thus, although we do not know to which probability density distribution the entries of the mass matrix belong, we shall assume that a Gaussian distributions comes as a good approximation, and it serves as a limiting case (see solid line of Fig. 11 which represents the PDF given by Eq. 5). We expect that a much amount of data would make closer our mass eigenvalues PDF comes from a GOE spectrum. Hence, the rest of our analysis relies on this assumption.

Now, if the mass matrix is interpreted as a random matrix with identically and independently distributed entries with Dyson index 1, this is a Gaussian orthogonal ensemble (GOE) with real entries, it is quite unlikely to get only positive eigenvalues. This well known result from random matrix theory (RMT) follows from the fact that extreme eigenvalues of a GOE obey the Tracy–Widow statistics and that any fluctuation in the lower limit is suppressed by a power $N^{-1/6}$ for N be the rank of the matrix [49] (as shown in Fig. 12). Thus let us put the RdSC in terms of a RMT.

The eigenvalues of a random matrix are expected to be distributed around zero, however, for large N it has been proved that the minimum eigenvalue tends to $-\sqrt{2N}$ while

the maximum to $\sqrt{2N}$. As we said, fluctuations of extreme eigenvalues falls as $N^{-1/6}$, and thus allowing a possibility for the minimum eigenvalue to acquire a value different from $-\sqrt{2N}$. The distribution of fluctuation around $-\sqrt{2N}$ is shown by the shadow region in Fig. 12. For a large value of N it seems that

$$(\min \nabla_i \nabla_j V)_{RM} \leq \alpha, \quad (8)$$

where the subindex RM stands for a random mass matrix and α a number to be determined. It is expected that in such scenarios (eigenvalue probability distribution), the probability for the minimum eigenvalue to be negative increases as N increases. Actually, as proved in [49], the probability for the minimum eigenvalue to be bounded by a number t is given by

$$\mathbf{P}(\min \lambda > t) = \exp\left[-\frac{1}{24} \left|\sqrt{2N}^{1/6}(t + \sqrt{2N})\right|^3\right]. \quad (9)$$

Notice that for $t > 0 (< 0)$ \mathbf{P} reduces (increases). In our case in which the eigenvalues λ are related to the mass eigenvalues, i.e. $\lambda \rightarrow \text{Eig}(\nabla_i \nabla_j V)$ we can chose t to be the proportional to the potential at the minimum. In that case we see that for $N = 6$,

$$\begin{aligned} \mathbf{P}(\min \nabla_i \nabla_j V > c'V) \\ = \exp\left[-\frac{1}{24} \left|\sqrt{2} \cdot 6^{1/6}(c'V + 2\sqrt{3})\right|^3\right], \end{aligned} \quad (10)$$

Thus for a dS vacua, \mathbf{P} is very small and the larger the value for V at the minimum, the smaller the probability for the lightest moduli to be positive. dS vacua seem to be very less favored than unstable critical dS points. Similarly, for an AdS vacuum, the probability for having all positive eigenvalues is much higher than the corresponding for a dS extreme point and it raises as the absolute value of the vacuum energy grows (see Fig. 12). We then conclude that the most probable configurations satisfy the bound

$$\min \nabla_i \nabla_j V \leq -c'V, \quad (11)$$

in agreement with the RdSC.

Notice as well that the probability expression also asserts that the ratio between the minimum squared mass in a stable AdS vacuum and the AdS scale larger than one, this is, $\min m^2/\Lambda_{AdS} < 1$ is more favored. Hence the AdS scale conjecture is also encoded in this probabilistic interpretation. Taking all our observations together, we conclude that:

In an effective model constructed from a perturbative flux compactification (at least for an isotropic toroidal one) the probability for the minimum mass eigenvalue to be larger than the corresponding vacuum energy Λ is given by

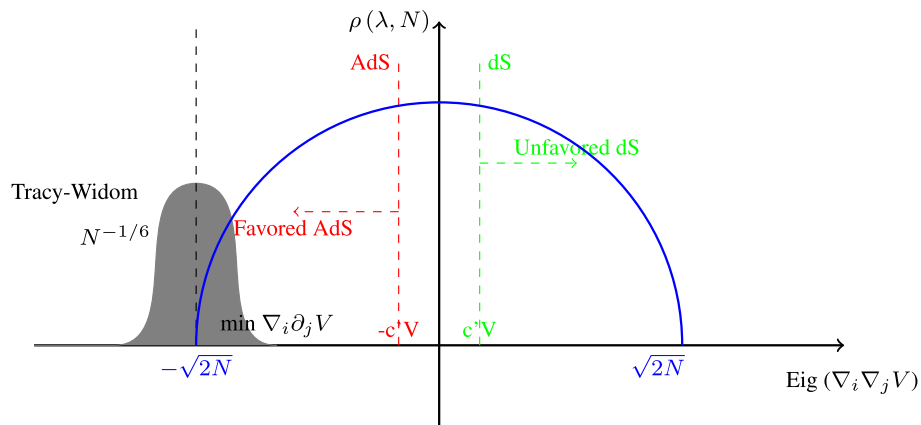


Fig. 12 Statistics for the extreme value statistics for the GOE. The blue line represents the probability density function for the eigenvalues. The horizontal axis represents the eigenvalues of the mass matrix, whereas the vertical axis represents the probability density function of those eigenvalues. For the case of AdS, the RdS conjecture is interpreted as the

probability to find the minimum below zero, i.e., $\mathbf{P}(\min \nabla_i \nabla_j V < c'V)$ (red lines) which is easily achieved. For the case of dS vacua, the RdS conjecture translates into the probability of all eigenvalues to be positive and above the $c'V$ vertical line, i.e., $\mathbf{P}(\min \nabla_i \nabla_j V < c'V)$ which is in general hard to be achieved (green lines)

$$\exp \left[-\frac{1}{24} \left| \sqrt{2} \cdot 6^{1/6} (\Lambda + \sqrt{12}) \right|^3 \right].$$

This implies that the most probable mass configurations with positive value of the cosmological constant are those which contain negative mass states in its spectrum. For the case of a negative value of the cosmological constant, the most probable scenario implies the presence of Tachyons. Notice that this implies that the most probable effective models are those precisely satisfying the RdS and the AdS conjectures.

4 Final comments

In this work we have implemented a vacuum search through an Artificial Neural Network coupled to a Genetic Algorithm. We report more than 60,000 flux configurations yielding to a scalar potential with at least one critical point. We use a simple model consisting on type IIB string theory flux compactification on an isotropic torus including non-geometric fluxes. With the data obtained by this classification we can test – in terms of probabilities – some of our model’s features in the light of recent Swampland conjectures.

Our main conclusion is that, at least for the studied model, generic flux configurations produce different vacua with two clear features:

- The refined dS conjecture is fulfilled and the relation $\min \nabla_i \nabla_j V \leq -c'V$ with c' of order 1 is graphically proved in Fig. 1. Notice the absence of certain stable AdS as well as some unstable dS vacua.

- A statistical correlation is observed favoring a small value for the cosmological constant in models exhibiting a small string coupling.

Our results show a clear increase in probability to find vacua with a smaller than unit cosmological constant (and in consequence within the perturbative regime) if they are constructed from a hierarchical flux configuration, meaning a flux configuration in which the integer quantized values for the different sectors, including non-geometrical fluxes, differ by at least one order of magnitude. The construction of different vacua, stable or not, from a hierarchical flux compactification leads to the following facts:

- The value of the corresponding cosmological constant is small and in consequence within the range of a perturbative effective theory. The probability to obtain such vacua increases by selecting the RR sector with the highest flux values, which in turn makes the complex structure moduli the heaviest.
- The probability to have an AdS stable vacuum in which the lightest modulus is much smaller than the corresponding cosmological constant increases.

We also observe by the use of random matrix theory that stable vacua are much less probable than unstable ones. Actually, in a random selection of fluxes which present a Gaussian distribution of mass eigenvalues, the more probable vacuum solutions are those which precisely fulfill the Swampland conjectures, namely the Refined de Sitter and the AdS scale ones. This suggests that the origin of the Swampland constraints, at least for the models we have studied, is probabilistic.

Finally we notice that the possibility to select a hierarchical flux configuration from a random set of different flux configurations, is very low, indicating that for a hierarchical flux configuration to be the source of effective models, a high-energy process must be the cause of fixing values for the fluxes. We leave this important issue for a future work.

Acknowledgements We thank Alejandro Cabo, Anamaria Font, Vishnu Jejjala, Albrecht Klemm, Challenger Mishra and Fernando Quevedo for useful discussions and comments. We thank Miguel Sabido for kind support through the Data Lab at University of Guanajuato. N.C.B. and O.L.B. are supported by the project CIIC 290/2020 UGTO, CONACyT Project A1-S-37752 and CONACyT Project CB-2015-01-258982. C.D. is supported by CONACyT through the S.N.I. program, and D.K.M.P. by DAIP Universidad de Guanajuato under project CIIC 344/2019 and by the Simons Foundation Mathematical and Physical Sciences Targeted Grants to Institutes, Award ID: 509116. J.A.M.B. thanks the National Council of Science and Technology (CONACyT), Mexico, for his Assisanship No. CVU-736083.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: The data used in this work as well as the codes to generate it and analyze it are publicly available under the following link. <https://l.facebook.com/l.php?u=https%3A%2F%2Fgithub.com%2Fdkmayorga%2FTesting-Swampland-Conjectures-with-ML%3Ffbclid%3DIwAR2yvoA0q7Ec4EL3EUds0MUckpPb2MxVq3EOaQEvINDeONQqRyQ0F94-eA&h=AT2ONJdNHkFYx2k0VPJw3RNK3CbVpnTaeNaTHtxM8uD6qWM5h5t6myuB8dchdoFTdfBihgMxsno7-vZjVqIgZ2OuFNK2Ehwmv8jBqU3KEIJjeJaOG8fR1RNkrQrXefCykGpK0Nmivg-iRrP-ox5Uj22Ezw>. The codes also generate the data for the Figures 3 to 11 as well as Figure 13.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Funded by SCOAP³.

Appendix A: Artificial neural network

Artificial neural networks (ANNs) are algorithms inspired in the biological learning process. The use of machine learning techniques to solve classification problems has attracted more attention in the last years [33, 37, 39, 50–53]. The machine learning techniques allows to search in large amount of data for specific patterns and thus, it provides an exhaustive check in a short time.⁷

⁷ For instance, in [37] using data mining the authors are able to look for suitable heterotic compactifications that selects an appropriate line bun-

In the following we describe in simple terms, the structure of an ANN. Each neuron in the hidden layer is connected with all the neurons in the neighboring clusters through a weight factor. The weighted interconnection is quantified by

$$a_j^l = \sigma \left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l \right), \quad (\text{A.1})$$

where the sum is over all neurons in the k level in the $l - 1$ layer, the weights w^l connects each l -th layer of neurons, b^l is the bias factor and a_j^l is the response/entrance of the ANN, for instance at $l = 1$, a represents the input data and at $l = n$ (last layer) a represents the output of the ANN. The function $\sigma(\cdot)$ is the activation function which is a sigmoid function that introduces the non-linearities to the ANN. The weights and bias factors are determined in such a way that the mean relative error defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (a_n - t_n)^2, \quad (\text{A.2})$$

is diminished, where a_n is the response of the ANN an the t is the target value. The numerical values are determined using the Levenberg-Marquardt optimization, which is a deterministic algorithm for non-linear systems that is able to find local minima in a iterative manner. This optimization requires to minimize the regularized function

$$f(w) = \frac{1}{2} (r_i)^2 + \frac{1}{2} \gamma^k (w_i^{k+1} - w_i^k)^2, \quad (\text{A.3})$$

where $r_i = a_i(w) - t_i$ is the residual and γ^k is a regularization parameter which is chosen through the trust region approach [54]. Thus, at each step the weights are determined by solving the equation

$$\left(J_{il}(w^k) J_{lj}(w^k) - \gamma^k \delta_{ij} \right) (w_j^{k+1} - w_j^k) = -J_{il}(w^k) r_l(w^k), \quad (\text{A.4})$$

for w^{k+1} at each iteration, where the index k represents the k -iteration and $J_{ij} = \partial_i r_j$ is the Jacobian of the residual.

Appendix A.1: An example of an ANN

To clarify the algorithm to compute the output on a ANN, in this section we present an explicit example of how a single perceptron is trained in order to reproduce the boolean product operation. Recall that this binary operator requires as input data two bits and calculates an output binary data

dle to produces a phenomenological motivated extension of the standard model.

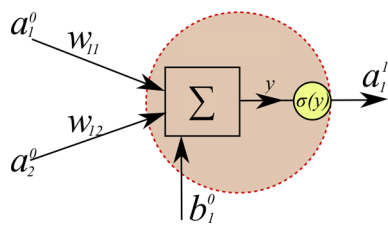


Fig. 13 Representation of a single Neuron. For this example the input is described by the vector (a_1^0, a_2^0) , associated to the neuron are the weights w_{11}, w_{12} and the bias b_1^0 , with $y = w^T \cdot a^0 + b_1^0$. With the neuron output $a_1^1 = \sigma(y)$ i.e. the activation function evaluated at y

(see Fig. 13). The training set employ the four possibilities for the boolean product, namely, $0 \cdot 0 = 0, 0 \cdot 1 = 0, 1 \cdot 0 = 0$ and $1 \cdot 1 = 1$.

Thus, for the single perceptron case, we initialize the weights of the network, for instance $w_{11}^0 = 0.5, w_{12}^0 = 1.5$ and for the bias $b_1^0 = 3$. For concreteness let us consider as activation function the logistic sigmoid function

$$\sigma(y) = \frac{1}{1 + \exp(-y)}, \tag{A.5}$$

and for the first case let us consider the boolean product $(0 \cdot 0 = 0)$, the argument of the activation function is

$$y_i = \sum_j w_{ij}^1 a_j^0 = 3, \tag{A.6}$$

and thus the activation function is $\sigma(3) = 0.95$. This result tell us that for the selected weight values the neuron gets activated with an answer of 1. However, the correct answer shall be 0 (since we are evaluating the $0 \cdot 0 = 0$ case, and we have obtained high error for the first boolean operation. Now, computing the remaining cases we get

$$\begin{aligned} y_0 &= w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 = 0.5 \cdot 0 + 1.5 \cdot 0 + 3 = 3.00 \\ &\rightarrow \sigma(3.00) = 0.95, \\ y_0 &= w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 = 0.5 \cdot 0 + 1.5 \cdot 1 + 3 = 4.50 \\ &\rightarrow \sigma(4.50) = 0.98, \\ y_0 &= w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 = 0.5 \cdot 1 + 1.5 \cdot 0 + 3 = 3.55 \\ &\rightarrow \sigma(3.55) = 0.97, \\ y_0 &= w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 = 0.5 \cdot 1 + 1.5 \cdot 1 + 3 = 5.05 \\ &\rightarrow \sigma(5.05) = 0.99, \end{aligned} \tag{A.7}$$

geometrically it is possible to see that in a a_0^0 vs a_1^0 plot, the region is divided into two regions, namely, if the value of the function $y_j > 0$ the neuron is activated and for $y_j < 0$ the neuron is de-activated as it is shown in Fig. 14. Thus, for the selected weights all the neuron is activated in all the training cases and the MSE given by Eq. A.2 is 0.698, which is too

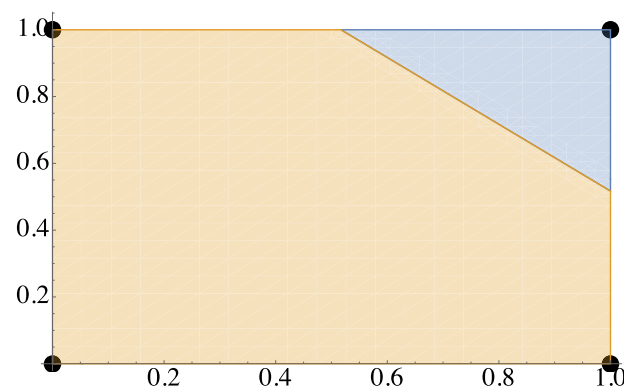


Fig. 14 Solution for the case of the AND gate. The x axis represents the first bit of the AND gate whereas the y axis the second bit. Thus, the yellow regions defines the region where the boolean product is 0 whereas the blue region represents the region where the boolean product is 1. The sharp changes between the region 0 and 1 is a consequence of the neuron activation

high to be acceptable. This step know as forward propagation allow us to compute the error for the randomly selected weights. The next step is to modify the random selected weights in order to reduce the MSE. The simplest way to do it is by implementing the batch gradient descent algorithm, which calculates the gradient of maximum descent for some objective function. Thus, we want to know how the gradient changes the objective function as a the weights changes, this is done by applying the rule of chain as

$$\frac{\partial \text{MSE}}{\partial w_{0j}} = \frac{\partial y_j}{\partial w_{0j}} \cdot \frac{\partial f}{\partial y_j} \cdot \frac{\partial \text{MSE}}{\partial f}, \tag{A.8}$$

where we shall use $w_{03} = b_0$. Thus, the new value of the weights, can be computed by resting the gradient from the previous value

$$w_j = w_j - \alpha \frac{\partial \text{MSE}}{\partial w_{0j}}, \tag{A.9}$$

where α is known as the learning rate and is an parameter between 0 and 1 (for this case we shall take 1). Thus, after the first iteration, this algorithm allows us to update the weights as, $w_{01}^0 = 0.48, w_{02}^0 = 1.48$ and for the bias $b_1^0 = 2.95$ with a MSE of 0.695. This algorithm usually has a low convergence ratio, since it takes a big number of steps to achieve a desired convergence criteria however for this toy model example a MSE < 0.002 is achieved at 2000 iterations. Thus, once the desired MSE is obtained, the updated weights take the values $w_{01}^0 = 5.49, w_{02}^0 = 6.45$ and the bias $b_1^0 = -8.28$ with an MSE of 0.002. Thus, a forward calculation shall make the ANN to reproduce the boolean product operation as

$$y_0 = w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 = 5.49 \cdot 0 + 6.45 \cdot 0 - 8.28$$

$$\begin{aligned}
 &= -8.28 \rightarrow \sigma(-8.28) = 0.00, \\
 y_0 = w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 &= 5.49 \cdot 0 + 6.45 \cdot 1 - 8.28 \\
 &= -1.83 \rightarrow \sigma(-1.83) = 0.05, \\
 y_0 = w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 &= 5.49 \cdot 1 + 6.45 \cdot 0 - 8.28 \\
 &= -2.79 \rightarrow \sigma(-2.79) = 0.05, \\
 y_0 = w_{01}^1 a_1^0 + w_{02}^1 a_1^0 + b_0 &= 5.49 \cdot 1 + 6.45 \cdot 1 - 8.28 \\
 &= 3.66 \rightarrow \sigma(3.66) = 0.93.
 \end{aligned}
 \tag{A.10}$$

which approximately is the desired result. Notice that for this toy model we employ all the available data to train the ANN and there are no remaining data for the validation. However, for a more complicated training set, a subset is not used in the training and a validation set is used.

Appendix B: Isotropic toroidal compactifications with non-geometric fluxes

In this appendix we describe a type IIB string theory compactification with non geometric fluxes on an isotropic T^6 torus. The effective 4D $\mathcal{N} = 1$ theory can be obtained in terms of a superpotential given by

$$W = P_1(U) - iSP_2(U) + iT P_3(U), \tag{B.11}$$

with U, S and T being the complex structure, axio-dilaton and Kähler moduli respectively. $P_1(U), P_2(U)$ and $P_3(U)$ are polynomials depending on U and their coefficients are given in terms of NS–NS h_j, R –R f_i and non geometric b_k fluxes respectively. The polynomials depend on the fluxes as follows

$$\begin{aligned}
 P_1 &= f_1 + 3If_2U - 3f_3U^2 - If_4U^3, \\
 P_2 &= h_1 + 3Ih_2U - 3h_3U^2 - Ih_4U^3, \\
 P_3 &= 3b_1 + I(2b_2 + b_3)U - (2b_4 + b_5)U^2 - Ib_6U^3.
 \end{aligned}
 \tag{B.12}$$

This structure of the superpotential, where the fluxes h and b determine the relevance of S, T to the scalar potential dependence respectively, suggests that the hierarchy between moduli masses can be reached by implementing hierarchies among the fluxes. The Kähler potential reads

$$K = -\ln(S + S^*) - 3\ln(U + U^*) - 3\ln(T + T^*). \tag{B.13}$$

We decompose the scalar fields in terms of their real and imaginary components as: $U = u + iv, S = s + ic$ and $T = t + i\tau$, where u, v, s, c, t, τ are real fields.

The corresponding scalar potential can be computed in terms of K and W and is given by

$$V = e^K \left(|D_I W|^2 K^{IJ} - 3|W|^2 \right). \tag{B.14}$$

This potential has extremal values when SUSY is preserved in an AdS or Minkowski vacuum. When SUSY is not preserved, it is possible to have extrema for all the different values of V_0 . The appearance of these extrema follows from the presence of fluxes including non-geometric fluxes, which could stabilize the Kähler moduli T . However, different flux configurations produce different type of vacua whose characteristics are also constrained by the tadpole on the NS–NS and RR fluxes. One important aspect in the construction of vacua solutions is to obtain physical consistency. This implies having a positive larger than one value for $s = 1/g_s$. This requirement ensures that the perturbative approximation for IIB string compactification on the isotropic torus is valid.

Since we search for SUSY vacua, it follows that all Kähler derivatives must vanish, i.e.

$$D_U W = D_S W = D_T W = 0. \tag{B.15}$$

Appendix B.1: Small cosmological constant: a toy example

We present an analysis concerning some SUSY solutions in order to elucidate the possible existence of some generic conditions on the flux configuration which leads us to effective models in which for an AdS vacuum not only $s > 1$ but also $|V_0| \ll 1$. Our goal is to obtain some insights about the characteristics of different flux configurations, which can assure the construction of such desirable vacua.

In the following we take a particular path in order to construct a supersymmetric solution with all moduli stabilized. First of all we observe that $D_T W = 0$ implies that Kähler moduli are fixed to

$$t = \frac{3}{2} \quad \text{and} \quad \tau = -\frac{1}{2} \frac{\tilde{p}_3}{p_3} = \frac{1}{2}, \tag{B.16}$$

where $P_2, P_3 \neq 0$ is assumed at any point of the complex structure at which the polynomial are evaluated with $p_3 = -\tilde{p}_3$. We are then forced to find some value U_0 for which this is valid for non-trivial polynomial. We shall come back to this point. Meanwhile, $D_S W = 0$ implies, once we take the above constraint fixed by $D_T W = 0$,

$$c = \frac{1}{2} \frac{p_2 + \tilde{p}_2}{p_2 - \tilde{p}_2}, \tag{B.17}$$

while s is kept unfixed. We shall fix it by minimizing the scalar potential with respect to s . Finally, $D_U W = 0$ fixes $U = U_0$ as a solution of the equation⁸

$$2u \frac{\partial P_2}{\partial U} - 3P_2 = 0. \tag{B.18}$$

⁸ More constraints would involve a sharing solution with a similar equation for $P_1(U)$. See Ref. [44].

However, as shown in [44], roots U_0 for the above polynomial implies that $P_3(U_0) = 0$. In order to keep a SUSY solution we shall then assume that

$$\lim_{U \rightarrow U_0} \frac{\tilde{P}_3(u_0, v_0)}{P_3(u_0, v_0)} = -1, \quad (\text{B.19})$$

with $\tau(u_0, v_0) = 1/2$. In this case the potential has the form

$$V(s) = -\frac{3}{2^7 u_0^3 v_0^3} \left(2\text{Im}(P_2 \omega^*) + s|P_2|^2 + \frac{|\omega|^2}{s} \right), \quad (\text{B.20})$$

where $\omega(u_0, v_0) = (p_1 + c_0 p_2) + i(\tilde{p}_1 + c_0 \tilde{p}_2)$. It follows that the string coupling is fixed at the value

$$s_0^2 = \frac{|\omega|^2}{|P_2|^2}(u_0, v_0), \quad (\text{B.21})$$

at the minimum of the scalar potential. At this point, our interest focuses in finding general conditions on the flux configuration upon which $|V_0| < 1$ while $s_0 > 1$. For the latter, observe that an easy and direct way to assure a small string coupling is to have $|\omega| > |P_2|$ which can easily be obtained by considering a hierarchy on the flux configuration among RR, NS-NS and non-geometric fluxes. Since these fluxes enter as the real coefficients on the polynomial $P_i(U_0)$ it follows that if RR fluxes are larger than NS-NS, which in turn are larger than non-geometric fluxes, one can obtain that $|\omega| > |P_2|$ for ω and P_2 evaluated at U_0 .

To obtain a small value for V at the minimum, observe that

$$V_0 = -\frac{1}{2^3 u_0^3 v_0^3} (\text{Im}(P_2 \omega^*) + |\omega||P_2|), \quad (\text{B.22})$$

which by taking same order fluxes among the same type of fluxes (e.g., all RR fluxes are the same order but larger than all NS-NS, which are of the same order among them), one can assure that $\text{Im}(P_2(U_0)\omega^*(U_0))$ be smaller than unit. Hence we can expect a small value for $|V_{min}|$ if ω and $|P_2|$ at $U = U_0$ are also smaller than one. As shown in [44], $\omega \sim \mathcal{O}(1)$ in flux units. This implies that

$$V_0 \sim -\frac{\mathcal{O}(1)}{u_0^3 s_0}. \quad (\text{B.23})$$

In effect, having a hierarchy on different type of fluxes, one can obtain in a generic form, at least for the SUSY solutions we have considered, a perturbative effective theory with a very small negative cosmological constant for $u_0 > 1/s_0$.

References

1. S.K. Garg, C. Krishnan, JHEP **11**, 075 (2019). [https://doi.org/10.1007/JHEP11\(2019\)075](https://doi.org/10.1007/JHEP11(2019)075)
2. S.K. Garg, C. Krishnan, M. Zaid Zaid, JHEP **03**, 029 (2019). [https://doi.org/10.1007/JHEP03\(2019\)029](https://doi.org/10.1007/JHEP03(2019)029)
3. G. Obied, H. Ooguri, L. Spodyneiko, C. Vafa, (2018). [arXiv:1806.08362](https://arxiv.org/abs/1806.08362) [hep-th]
4. P. Agrawal, G. Obied, P.J. Steinhardt, C. Vafa, Phys. Lett. B **784**, 271 (2018). <https://doi.org/10.1016/j.physletb.2018.07.040>
5. H. Ooguri, E. Palti, G. Shiu, C. Vafa, Phys. Lett. B **788**, 180 (2019). <https://doi.org/10.1016/j.physletb.2018.11.018>
6. U. Danielsson, JHEP **04**, 095 (2019). [https://doi.org/10.1007/JHEP04\(2019\)095](https://doi.org/10.1007/JHEP04(2019)095)
7. U.H. Danielsson, T. Van Riet, Int. J. Mod. Phys. D **27**(12), 1830007 (2018). <https://doi.org/10.1142/S0218271818300070>
8. R. Blumenhagen, D. Klawer, L. Schlechter, JHEP **05**, 152 (2019). [https://doi.org/10.1007/JHEP05\(2019\)152](https://doi.org/10.1007/JHEP05(2019)152)
9. C. Damian, L.R. Diaz-Barron, O. Loaiza-Brito, M. Sabido, JHEP **06**, 109 (2013). [https://doi.org/10.1007/JHEP06\(2013\)109](https://doi.org/10.1007/JHEP06(2013)109)
10. C. Damian, O. Loaiza-Brito, Phys. Rev. D **88**(4), 046008 (2013). <https://doi.org/10.1103/PhysRevD.88.046008>
11. N. Cabo Bizet, S. Hirano, (2016). [arXiv:1607.01139](https://arxiv.org/abs/1607.01139) [hep-th]
12. N. Cabo Bizet, O. Loaiza-Brito, I. Zavala, JHEP **10**, 082 (2016). [https://doi.org/10.1007/JHEP10\(2016\)082](https://doi.org/10.1007/JHEP10(2016)082)
13. C. Damian, O. Loaiza-Brito, Fortsch. Phys. **67**(1–2), 1800072 (2019). <https://doi.org/10.1002/prop.201800072>
14. A. Bedroya, C. Vafa, (2019). [arXiv:1909.11063](https://arxiv.org/abs/1909.11063) [hep-th]
15. D. Andriot, N. Cribiori, D. Erkiner, (2020). [arXiv:2004.00030](https://arxiv.org/abs/2004.00030) [hep-th]
16. D. Lüst, E. Palti, C. Vafa, Phys. Lett. B **797**, 134867 (2019). <https://doi.org/10.1016/j.physletb.2019.134867>
17. E. Palti, Fortsch. Phys. **67**(6), 1900037 (2019). <https://doi.org/10.1002/prop.201900037>
18. R. Blumenhagen, M. Brinkmann, A. Makridou, JHEP **20**, 064 (2020). [https://doi.org/10.1007/JHEP02\(2020\)064](https://doi.org/10.1007/JHEP02(2020)064)
19. R. Blumenhagen, E. Plauschinn, Phys. Lett. B **736**, 482 (2014). <https://doi.org/10.1016/j.physletb.2014.08.007>
20. R. Blumenhagen, C. Damian, A. Font, D. Herschmann, R. Sun, Fortsch. Phys. **64**(6–7), 536 (2016). <https://doi.org/10.1002/prop.201600030>
21. R. Blumenhagen, A. Font, M. Fuchs, D. Herschmann, E. Plauschinn, Y. Sekiguchi, F. Wolf, Nucl. Phys. B **897**, 500 (2015). <https://doi.org/10.1016/j.nuclphysb.2015.06.003>
22. R. Blumenhagen, I. Valenzuela, F. Wolf, JHEP **07**, 145 (2017). [https://doi.org/10.1007/JHEP07\(2017\)145](https://doi.org/10.1007/JHEP07(2017)145)
23. P. Breitenlohner, D.Z. Freedman, Phys. Lett. B **115**(3), 197 (1982)
24. N. Cribiori, R. Kallosh, A. Linde, C. Roupec, Phys. Rev. D **101**(4), 046018 (2020). <https://doi.org/10.1103/PhysRevD.101.046018>
25. P. Betzler, E. Plauschinn, Fortsch. Phys. **67**(11), 1900065 (2019). <https://doi.org/10.1002/prop.201900065>
26. J. Blabäck, U.H. Danielsson, G. Dibitetto, S.C. Vargas, JHEP **10**, 069 (2015). [https://doi.org/10.1007/JHEP10\(2015\)069](https://doi.org/10.1007/JHEP10(2015)069)
27. J. Blabäck, U. Danielsson, G. Dibitetto, JHEP **08**, 054 (2013). [https://doi.org/10.1007/JHEP08\(2013\)054](https://doi.org/10.1007/JHEP08(2013)054)
28. R. Blumenhagen, A. Deser, E. Plauschinn, F. Rennecke, C. Schmid, Fortsch. Phys. **61**, 893 (2013). <https://doi.org/10.1002/prop.201300013>
29. R. Blumenhagen, X. Gao, D. Herschmann, P. Shukla, JHEP **10**, 201 (2013). [https://doi.org/10.1007/JHEP10\(2013\)201](https://doi.org/10.1007/JHEP10(2013)201)
30. E. Plauschinn, Phys. Rep. **798**, 1 (2019). <https://doi.org/10.1016/j.physrep.2018.12.002>
31. F. Ruehle, JHEP **08**, 038 (2017). [https://doi.org/10.1007/JHEP08\(2017\)038](https://doi.org/10.1007/JHEP08(2017)038)

32. J. Carifio, J. Halverson, D. Krioukov, B.D. Nelson, *JHEP* **09**, 157 (2017). [https://doi.org/10.1007/JHEP09\(2017\)157](https://doi.org/10.1007/JHEP09(2017)157)
33. Y.H. He, (2018). [arXiv:1812.02893](https://arxiv.org/abs/1812.02893)
34. H. Erbin, S. Krippendorf, (2018). [arXiv:1809.02612](https://arxiv.org/abs/1809.02612) [cs.LG]
35. A. Mutter, E. Parr, P.K. Vaudrevange, *Nucl. Phys. B* **940**, 113 (2019). <https://doi.org/10.1016/j.nuclphysb.2019.01.013>
36. A. Ashmore, Y.H. He, B.A. Ovrut, (2019). [arXiv:1910.08605](https://arxiv.org/abs/1910.08605) [hep-th]
37. E. Parr, P.K.S. Vaudrevange, *Nucl. Phys. B* **952**, 114922 (2020). <https://doi.org/10.1016/j.nuclphysb.2020.114922>
38. J. Halverson, B. Nelson, F. Ruehle, *JHEP* **06**, 003 (2019). [https://doi.org/10.1007/JHEP06\(2019\)003](https://doi.org/10.1007/JHEP06(2019)003)
39. A. Cole, A. Schachner, G. Shiu, *JHEP* **11**, 045 (2019). [https://doi.org/10.1007/JHEP11\(2019\)045](https://doi.org/10.1007/JHEP11(2019)045)
40. F. Ruehle, *Phys. Rep.* **839**, 1 (2020). <https://doi.org/10.1016/j.physrep.2019.09.005>
41. Y. Gal, V. Jejjala, D.K. Mayorga Pena, C. Mishra, (2020). [arXiv:2003.10445](https://arxiv.org/abs/2003.10445) [hep-ph]
42. S. Abel, J. Rizos, *JHEP* **08**, 010 (2014). [https://doi.org/10.1007/JHEP08\(2014\)010](https://doi.org/10.1007/JHEP08(2014)010)
43. S. AbdusSalam, S. Abel, M. Cicoli, F. Quevedo, P. Shukla, (2020). [arXiv: 2005.11329](https://arxiv.org/abs/2005.11329) [hep-th]
44. N. Cabo Bizet, C. Damian, O. Loaiza-Brito, D.M. Peña, *JHEP* **09**, 123 (2019). [https://doi.org/10.1007/JHEP09\(2019\)123](https://doi.org/10.1007/JHEP09(2019)123)
45. F. Gautason, V. Van Hemelryck, T. Van Riet, *Fortsch. Phys.* **67**(1–2), 1800091 (2019). <https://doi.org/10.1002/prop.201800091>
46. L.F. Low, S. Hotchkiss, R. Easter, (2020). [arXiv:2004.04429](https://arxiv.org/abs/2004.04429) [hep-th]
47. M.L. Mehta, *Random Matrices* (Elsevier, Amsterdam, 2004)
48. G. Livan, M. Novaes, P. Vivo, *Introduction to Random Matrices: Theory and Practice*, vol. 26 (Springer, Berlin, 2018)
49. D.S. Dean, S.N. Majumdar, *Phys. Rev. Lett.* **97**(16), 160201 (2006)
50. J. Bao, Y.H. He, E. Hirst, S. Pietromonaco, (2020). [arXiv:2001.01212](https://arxiv.org/abs/2001.01212) [hep-th]
51. J. Halverson, C. Long, *Fortsch. Phys.* **68**(5), 2000005 (2020). <https://doi.org/10.1002/prop.202000005>
52. C.R. Brodie, A. Constantin, R. Deen, A. Lukas, *Fortsch. Phys.* **68**(1), 1900087 (2020). <https://doi.org/10.1002/prop.201900087>
53. Y.H. He, S.J. Lee, *Phys. Lett. B* **798**, 134889 (2019). <https://doi.org/10.1016/j.physletb.2019.134889>
54. A.A. Suratgar, M.B. Tavakoli, A. Hoseinabadi, *World Acad. Sci. Eng. Technol.* **6**(1), 46 (2005)