



City Research Online

## City, University of London Institutional Repository

---

**Citation:** Constantin, A., He, Y. & Lukas, A. (2019). Counting string theory standard models. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 792, pp. 258-262. doi: 10.1016/j.physletb.2019.03.048

This is the published version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/22130/>

**Link to published version:** <https://doi.org/10.1016/j.physletb.2019.03.048>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---

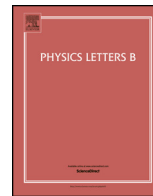
---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---



# Counting string theory standard models

Andrei Constantin<sup>a,b</sup>, Yang-Hui He<sup>c,d,e,\*</sup>, Andre Lukas<sup>f</sup>

<sup>a</sup> Pembroke College, Oxford University, OX1 1DW, Oxford, UK

<sup>b</sup> Mansfield College, Oxford University, OX1 3TF, Oxford, UK

<sup>c</sup> Merton College, University of Oxford, OX1 4JD, UK

<sup>d</sup> Department of Mathematics, City, University of London, EC1V0HB, UK

<sup>e</sup> School of Physics, NanKai University, Tianjin, 300071, China

<sup>f</sup> Rudolf Peierls Centre for Theoretical Physics, Oxford University, 1 Keble Road, Oxford, OX1 3NP, UK

## ARTICLE INFO

### Article history:

Received 13 December 2018

Received in revised form 22 March 2019

Accepted 25 March 2019

Available online 28 March 2019

Editor: M. Cvetič

## ABSTRACT

We derive an approximate analytic relation between the number of consistent heterotic Calabi-Yau compactifications of string theory with the exact charged matter content of the standard model of particle physics and the topological data of the internal manifold: the former scaling exponentially with the number of Kähler parameters. This is done by an estimate of the number of solutions to a set of Diophantine equations representing constraints satisfied by any consistent heterotic string vacuum with three chiral massless families, and has been computationally checked to hold for complete intersection Calabi-Yau threefolds (CICYs) with up to seven Kähler parameters. When extrapolated to the entire CICY list, the relation gives  $\sim 10^{23}$  string theory standard models; for the class of Calabi-Yau hypersurfaces in toric varieties, it gives  $\sim 10^{723}$  standard models.

© 2019 The Authors. 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 and summary

It is generally believed that string compactifications that have the exact charged matter content of the standard model of particle physics (and no other charged matter except moduli) are few in number. The purpose of this letter is to show that, although such compactifications may be rare and hard to find, their number is substantial. Admittedly, this bias has come in the past from the difficulty to construct phenomenologically viable compactifications. However, since the birth of string phenomenology in Ref. [1], from the advent of the first standard-like string model [2], to the first exact particle spectrum directly derived from a string compactification [3–6], to the first result [7] from algorithmic heterotic compactification [8], until the comprehensive computer scan of Refs. [9–11,15], as well as the various statistical perspectives on the heterotic landscape [12,38] (cf. [13] in Type II and beyond [14]), there has been much progress.

While it can be specified at different levels of sophistication, for this letter a “string standard model” is a model with a massless spectrum which is exactly that of the minimally supersymmet-

ric standard model (MSSM), plus any number of massless modes (moduli fields) *uncharged* under the standard model gauge group.

The general strategy of heterotic string phenomenology is to consider a smooth, compact Calabi-Yau threefold (CY), say  $X$ , with a non-trivial fundamental group  $\Gamma$ , together with a holomorphic, (poly-)stable vector bundle  $V$  over  $X$ , typically with structure group  $SU(5)$  or  $SU(4)$ . Subsequently, a  $\Gamma$ -Wilson line can break the Grand Unified Theory (GUT) group, typically  $SU(5)$  or  $SO(10)$ , to the MSSM group and the  $\Gamma$ -equivariant cohomology of  $V$  as well as its appropriate tensor powers correspond to the MSSM particles.

From an algorithmic point of view, one can (1) take manifolds  $\tilde{X}$  from existing databases, most of which are simply connected, then search for discrete, freely acting symmetry groups  $\Gamma$  on  $\tilde{X}$ , and consider the quotient  $X \simeq \tilde{X}/\Gamma$  with fundamental group  $\Gamma$ ; (2) construct and classify families of  $\Gamma$ -equivariant bundles  $V$  on  $X$ , ensure stability, and then compute the relevant cohomology groups; and (3) scan through the results to look for exact MSSM particle content. Much of these can be implemented on a computer.

The most extensively used databases of manifolds are the Complete Intersection Calabi-Yau three-folds (CICYs) embedded in products of projective spaces of around 8000 manifolds [16,17] as well as the Kreuzer-Skarke (KS) dataset of Calabi-Yau hypersurfaces embedded in around half-billion four-dimensional toric varieties [18–20] (the actual number of Calabi-Yau will vastly ex-

\* Corresponding author.

E-mail addresses: [andrei.constantin@physics.uu.se](mailto:andrei.constantin@physics.uu.se) (A. Constantin), [hey@maths.ox.ac.uk](mailto:hey@maths.ox.ac.uk) (Y.-H. He), [lukas@physics.ox.ac.uk](mailto:lukas@physics.ox.ac.uk) (A. Lukas).

ceed this number because of triangulations and a recent estimate was made in [21]). The comprehensive scan of [11] was performed on CICY manifolds with a number of Kähler parameters less than  $h^{1,1}(X) = 6$  and vector bundles constructed from sum of line bundles. A total of about 35,000  $SU(5)$  heterotic line bundle models were found, all with the right field content to induce low-energy standard-like models.

It is obviously important to have a count of MSSM models expected within string theory. While ours is still a relatively unrefined notion of the standard model, even counting at this level is not easy since it requires information on all vector bundles and their cohomology on CYs which is not available in any systematic form. An exception to this rule is the class of vector bundles that split into a sum of line bundles. Holomorphic line bundles are classified by their first Chern class, which can be expressed in terms of  $h^{1,1}(X)$  integers. As such, line bundles can be enumerated.

Moreover, there is enough evidence that indicates the existence of analytic formulae for the ranks of line bundle valued cohomology groups in terms of the line bundle integers [23–26]. Finally, line bundle sums offer an accessible window into the moduli space of non-abelian bundles [26–28]: if a line bundle sum corresponds to a standard-like model, then usually it can be deformed into non-abelian bundles that also lead to standard-like models.

Our model building experience for such (rank five) line bundle models on CICYs suggests that a significant number of consistent models with the correct chiral asymmetry will descend to standard models after dividing by the freely acting discrete symmetry group  $\Gamma$ . Moreover, by far the most frequent symmetry is  $\mathbb{Z}_2$ . This suggests that an indication of the number of standard models should be provided by counting the consistent upstairs line bundle models with chiral asymmetry 6, relevant for  $\mathbb{Z}_2$  symmetries.

We start our analysis by outlining the constraints on the compactification data that guarantee an exact MSSM spectrum. For a fixed manifold, most of these constraints take the form of Diophantine equations and inequalities, where the unknown variables are the line bundle integers. For the class of CICY manifolds with less than six Kähler parameters this system was solved in Ref. [11] by explicitly checking every possible line bundle sum. We augment this dataset of line bundle models with results for 7 new manifolds with  $h^{1,1}(X) = 6, 7$ .

These scans suggest a simple rule: the number of line bundle models increases roughly by an order of magnitude with every increment of  $h^{1,1}(X)$  by one. However, it is difficult to test this relation for larger values of  $h^{1,1}(X)$  due to computer limitations. Instead, we estimate the number of solutions to the Diophantine system of constraints using a result from the mathematical literature [29]. For this to hold, we define a bound on the line bundle integers in terms of topological data of the manifold.

Finally, we come back to the empirical dataset of line bundles and correlate the number of solutions not only with  $h^{1,1}(X)$ , but also with a number of topological invariants, constructed from the intersection form and the second Chern class of the CY manifold, that display little variation with increasing  $h^{1,1}(X)$ . Extrapolating the multi-linear regression to the maximal value of  $h^{1,1}(X)$  found in the CICY dataset, we estimate a total of  $N_{\text{CICY}} \simeq 10^{23}$  line bundle models (we owe the expression “a mole of models” to Tristan Hübsch) while for the manifolds in the Kreuzer-Skarke list we expect  $N_{\text{KS}} \simeq 10^{723}$  line bundle models.

**Acknowledgments.** We are grateful to K. Dienes and T. Hübsch for valuable comments on the draft. AC and AL would like to thank the Mainz Institute for Theoretical Physics for hospitality during part of the completion of this project. AL is partially supported by the EPSRC network grant EP/N007158/1. YHH thanks STFC for grant ST/J00037X/1.

## 2. Counting line bundle MSSMs

The models of interest for our count have an exact MSSM particle content and are constructed from heterotic compactifications on a smooth, compact Calabi-Yau threefolds  $X$  endowed with slope-zero, poly-stable direct sums of line bundles. Let  $h$  denote the Picard number of  $X$ ,  $h := h^{1,1}(X)$ , and choose an integral basis of  $H^2(X)$  denoted by  $\{J_i\}$ , where  $i = 1, \dots, h$ . In this basis, let the second Chern class of  $X$  be  $c_{2,i}$  and the triple intersection numbers be  $d_{ijk} = \int_X J_i \wedge J_j \wedge J_k$ . Line bundles  $L \rightarrow X$  with first Chern class  $c_1(L) = k^i J_i$  are denoted by  $L = \mathcal{O}_X(\mathbf{k})$ . The reader is referred to Sec. 4 of Ref. [11] for further details on the constraints.

First, we focus on  $SU(5)$  bundles  $V$  for the following reason. From a group theoretic point of view [22], there are many ways to break the GUT group to the MSSM group using an appropriate discrete Wilson line, for example, the exact MSSM spectrum of [4] was achieved with a  $\mathbb{Z}_3 \times \mathbb{Z}_3$  Wilson line from an  $SO(10)$  GUT group. However, CY manifolds  $\tilde{X}$  with a large freely acting discrete symmetry group  $\Gamma$  are quite rare. This can be seen, for instance, from the complete classification of freely-acting [33] and residual [34,35] symmetries on all CICYs, or from the KS dataset of hyper-surfaces in toric Fano fourfolds [30,32].

Therefore, generically, it is expected that Calabi-Yau manifolds with a small fundamental group  $\pi_1(X)$ , should far exceed in number those with a large  $\pi_1(X)$  (this should be contrasted with the relative paucity of Calabi-Yau manifolds of small Hodge numbers [36,37,39]). The smallest possible  $\pi_1(X)$  that breaks the  $SU(5)$  GUT group to the Standard Model gauge group is  $\Gamma = \mathbb{Z}_2$  and this setup is expected to dominate. Now, in  $SU(5)$  (commutant of the  $SU(5)$  of the bundle in  $E_8$ ) GUTs, the  $\mathbf{10}$  representation corresponds entirely to anti-families which we desire to be absent. Under the branching of  $E_8$  to  $SU(5)$ , this corresponds to the condition that  $h^2(X, V) = 0$ , so that stability (implying that  $h^0(X, V) = h^3(X, V) = 0$ ) in conjunction with the index theorem gives

$$-3|\Gamma| = \text{ind}(V) = \sum_{i=0}^3 h^i(X, V) = -h^1(X, V). \tag{1}$$

Thus, we impose  $\text{ind}(V) = -h^1(X, V) = -6$ .

In summary, the counting problem can then be formulated as follows:

**PROBLEM:** What is the number  $N = N(h, c_{2,i}, d_{ijk})$  of rank five line bundle sums  $V = \bigoplus_{a=1}^5 L_a$ , where  $L_a = \mathcal{O}_X(\mathbf{k}_a)$  satisfying the following constraints:

**$E_8$  embedding:**  $c_1(V) = \sum_{a=1}^5 k_a^i \stackrel{!}{=} 0$  for all  $i = 1, \dots, h$ ;

**Anomaly cancellation:**

$$c_{2,i}(V) = -\frac{1}{2}d_{ijk} \sum_a k_a^j k_a^k \stackrel{!}{\leq} c_{2,i} \text{ for all } i = \dots, h;$$

**Supersymmetry/Zero Slope:** there is a common solution  $t^i$  to the vanishing slopes

$$\mu(L_a) = d_{ijk} k_a^j t_a^k \stackrel{!}{=} 0 \text{ for } a = 1, \dots, 5$$

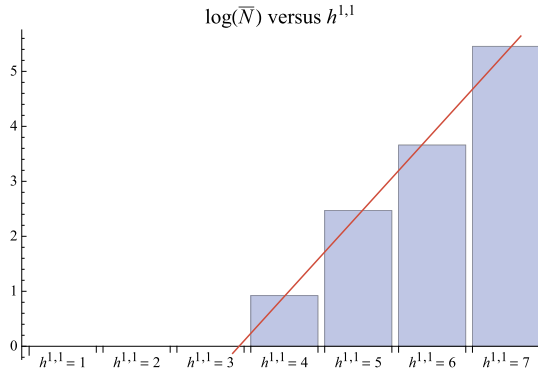
such that  $J = t^i J_i \in$  interior of the Kahler cone;

**Particle generations:** the chiral asymmetry is six, i.e.

$$\text{ind}(V) = \frac{1}{6}d_{ijk} \sum_a k_a^i k_a^j k_a^k \stackrel{!}{=} -6.$$

**Table 1**  
Number of models found in a computer scan for favourable CICYs with  $4 \leq h \leq 7$  for each  $h$ .

$h = h^{1,1}(X)$	4	5	6	7
number of CYs	9	10	8	5
number of models	75	2949	36692	1428856
number, $\bar{N}$ , of models per CY	8.3	294.9	4586.5	285771.0



**Fig. 1.** The logarithm of  $\bar{N}$ , the average number of models per CY, as a function of  $h = h^{1,1}(X)$ , taken from the data in Table 1. The red line is a linear fit to the data.

We emphasize that  $N$  is a function of the prescribed Hodge number  $h$ , the second Chern class  $c_{2,i}$ , as well as the triple intersection numbers  $d_{ijk}$  of  $X$ .

### 2.1. Preliminary count

For the subset of favourable CICYs with  $h \leq 6$ , the number  $N$  was determined by the computer scan [11] and we have extended this scan to include manifolds for  $h^{1,1}(X) = 7$  as well as some non-favourable manifold for  $h^{1,1} \leq 7$  which were previously discarded. The cumulative and average number of models found for each Picard number  $h$  is summarised in Table 1. Note there are no viable models for  $h = 1, 2, 3$  since the supersymmetry (slope zero) conditions are too constraining for those cases.

A simple approach is to assume that the main dependence of  $N$  is on  $h$ , and to neglect the possible effects of  $c_{2,i}$  and  $d_{ijk}$ . The average number,  $\bar{N} = \bar{N}(h)$  of models per CY as a function of  $h$ , taken from the last row of Table 1, has been plotted (logarithmically) in Fig. 1.

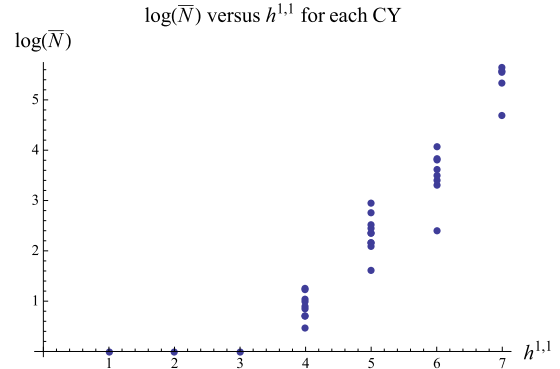
A linear fit to this data (which corresponds to the red line in Fig. 1) leads to

$$\log(\bar{N}(h)) \simeq -5.0 + 1.5h. \quad (2)$$

The largest known Picard number of any CY threefold is  $h_{\max} = 491$ , which appears within the KS data set, and the largest value within the CICY list is  $h_{\text{CICY}} = 19$ . Using (2) to boldly extrapolate to those values we find

$$\bar{N}(h_{\text{CICY}}) \simeq 10^{23}, \quad \bar{N}(h_{\max}) \simeq 10^{721}. \quad (3)$$

Clearly, these numbers are quite dramatic, even if we restrict ourselves to the CICYs. The predicted number of standard models even within this set is significantly larger than can be currently stored, let alone found by a scan. However, the method so far is quite crude and the extrapolation to large  $h$  adventurous. To see some of the problems, consider Fig. 2 which shows the number of models as a function of  $h$  for each CICY, rather than the average over all CICYs with the same  $h$ , as in Fig. 1. The variation within given  $h$  can be seen to be considerable - clear indication that there is a strong dependence of  $N$  on  $c_{2,i}$  and  $d_{ijk}$ , in addition to  $h$ .



**Fig. 2.** The logarithm of  $N$ , the number of models for each of the CICY manifolds, as a function of  $h = h^{1,1}(X)$ .

### 2.2. Some theoretical considerations

While the computer scan gives a finite number of models in each case, it is actually not easy to prove that  $N$  is finite. A succinct argument was presented in Ref. [15] based on the moduli space metric

$$G_{ij} = -3 \left( \frac{\kappa_{ij}}{\kappa} - \frac{3}{2} \frac{\kappa_i \kappa_j}{\kappa^2} \right), \quad (4)$$

where  $\kappa = d_{ijk} t^i t^j t^k$ ,  $\kappa_i = d_{ijk} t^j t^k$  and  $\kappa_{ij} = d_{ijk} t^k$ , with  $t^i$  being the Kähler moduli. We note that the slope zero conditions can be expressed as  $\kappa_i k_a^i = 0$  so that

$$\begin{aligned} 0 &\leq \sum_a \mathbf{k}_a^T G \mathbf{k}_a = -\frac{3}{\kappa} d_{ijk} \sum_a k_a^i k_a^j t^k = \frac{6}{\kappa} t^i c_{2,i}(V) \\ &\leq \frac{6}{\kappa} t^i c_{2,i}(TX) \leq \frac{6}{\kappa} |\mathbf{t}| |c_{2,i}(TX)|. \end{aligned} \quad (5)$$

Then, introducing the scale-invariant modified metric  $\tilde{G} = \frac{\kappa}{6|\mathbf{t}|} G$ , we get the bound

$$\sum_a \mathbf{k}_a^T \tilde{G} \mathbf{k}_a \leq |c_{2,i}(TX)|. \quad (6)$$

This by itself does unfortunately not bound the vectors  $\mathbf{k}_a$  since the metric  $\tilde{G}$  might become singular at the boundary of the Kähler cone. However, if we require that we stay in a “physical” region of the Kähler cone where all curve volumes to be greater than 1 (so that the supergravity approximation is valid) and the volume  $\kappa$  is bounded from above (so that we are not de-compactifying) then the eigenvalues of  $\tilde{G}$  are bounded from below by a strictly positive number.

Eq. (6) then implies that the length  $\sum_a |\mathbf{k}_a|^2$  is bounded from above and, hence, that there is only a finite number of possible integer vectors  $\mathbf{k}_a$ . More quantitative statements depend very much on the specific example but what can be said is that the length of the  $\mathbf{k}$  vectors is bounded by a radius  $R$  which roughly scales as

$$R \sim \left( \frac{c_2}{d} \right)^{1/2}, \quad (7)$$

where  $c_2$  and  $d$  are typical values of  $c_{2,i}$  and  $d_{ijk}$ .

Apart from the slope zero conditions the constraints on the line bundles can be written as a system of Diophantine equations

$$\begin{aligned} \sum_{a=1}^5 k_a^i &\stackrel{!}{=} 0, & -\frac{1}{2} d_{ijk} \sum_a k_a^j k_a^k &\stackrel{!}{=} c_{2,i}, \\ \frac{1}{6} d_{ijk} \sum_a k_a^i k_a^j k_a^k &\stackrel{!}{=} -6, \end{aligned} \quad (8)$$

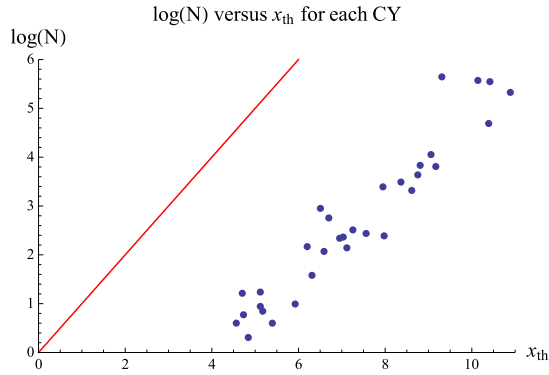


Fig. 3. The logarithm of  $N$  from the computer scan of CICYs versus the “theoretical” upper bound  $x_{th}$ , as defined in Eq. (9).

for  $i = 1, \dots, h$ . Here we have replaced the inequality in the anomaly condition with an equality, assuming that the bulk of the contribution comes from line bundles with the largest allowed integers. We can homogenise these equations (by introducing one additional coordinate) and think of them as a set of equations in  $\mathbb{P}^n$ , where  $n = 5h$ . Assuming they provide a complete intersection,  $Z$ , its dimension is given by  $m = n - 2h - 1 = 3h - 1$ .

Ref. [29] provides an upper bound for the number of rational points  $N_Z(B)$  within a box of size  $B$  (where the size is measured by the maximum norm) on  $Z$  which is given by  $N_Z(B) \ll B^m$ . Using the above radius  $R$  as an upper bound for  $B$  we find

$$N \ll R^{3h-1} \sim \left(\frac{c_2}{d}\right)^{(3h-1)/2} =: x_{th}. \tag{9}$$

The comparison between this upper bound and the results from the computer scan on CICY manifolds is shown in Fig. 3. With all the data points well below the red line it is clear that Eq. (9) indeed provides an upper bound but it is equally obvious that this upper bound is rather weak. There are a number of possible reasons for this. First, the result of Ref. [29] is only an upper bound (which, in addition, counts rational rather than integer points). Second, the radius  $R$  from Eq. (7) is a crude estimate and is, by itself, only an upper bound on the size  $B$  of the box considered in Ref. [29].

### 2.3. A more sophisticated count

Our count in §2.1, based on considering only the dependence of  $N$  on the Picard number  $h$  is clearly somewhat unrefined, while the theoretical upper bound  $x_{th}$  from §2.2 is clearly too weak to allow for a meaningful extrapolation to larger values of  $h$ . In this subsection, we seek a more sophisticated equation for  $N$  as a function of  $h$ ,  $c_{2,i}$  and  $d_{ijk}$ , drawing inspiration from the above discussions.

There is an obvious difficulty of writing down even an ansatz for  $N$  as a function of  $c_{2,i}$  and  $d_{ijk}$ : both of these quantities are basis-dependent (on a choice of integral basis  $J_i$  for  $H^2(X)$ ), but  $N$  clearly cannot depend on such a choice of basis. This means we should think about basis-independent quantities which can be constructed from  $c_{2,i}$  and  $d_{ijk}$ .

Unfortunately, both quantities have “all indices down” so there is no invariant which can be obtained by a simple contraction of indices, given that the only available metric,  $G$  from Eq. (4), is moduli-dependent. This problem has been encountered before, in the context of practical applications of Wall’s theorem, and a solution has been proposed in Ref. [17], p. 174.

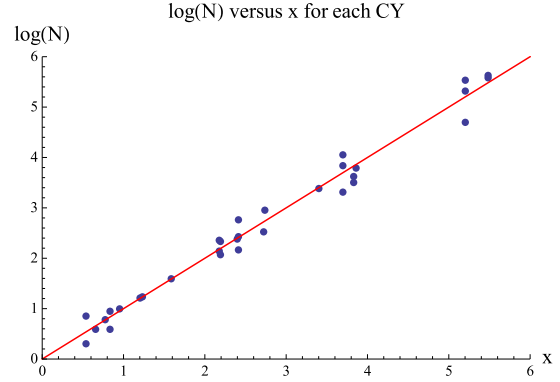


Fig. 4. The logarithm of  $N$  from the computer scan of CICYs versus the quantity  $x$  in Eq. (12), where the constants  $A_i$  and  $B_i$  have been determined to provide the best fit.

From the intersection form  $\lambda(\alpha, \beta, \gamma) := \int_X \alpha \wedge \beta \wedge \gamma \in \mathbb{Z}_{\geq 0}$ , completely symmetric in  $\alpha, \beta, \gamma$ , the following invariants can be constructed:

$$\begin{aligned} \ell_1 &= \gcd\{\lambda(\alpha, \beta, \gamma) \mid \alpha, \beta, \gamma \in H^2(X, \mathbb{Z})\} \\ \ell_2 &= \gcd\{\lambda(\alpha, \beta, \beta) \mid \alpha, \beta \in H^2(X, \mathbb{Z})\} \\ \ell_3 &= \gcd\{\lambda(\alpha, \alpha, \alpha) \mid \alpha \in H^2(X, \mathbb{Z})\}. \end{aligned} \tag{10}$$

Furthermore, combining the intersection form and  $c_2 = c_2(TX)$  we can define the form  $\Lambda(\alpha, \beta, \gamma, \delta) = (\lambda(\alpha, \beta, \gamma)c_2(\delta) + 3 \text{ permutations})$  which gives rise to the invariants

$$\begin{aligned} \ell_4 &= \gcd\{\Lambda(\alpha, \beta, \gamma, \delta) \mid \alpha, \beta, \gamma, \delta \in H^2(X, \mathbb{Z})\} \\ \ell_5 &= \gcd\{\Lambda(\alpha, \beta, \gamma, \gamma) \mid \alpha, \beta, \gamma \in H^2(X, \mathbb{Z})\} \\ \ell_6 &= \gcd\{\Lambda(\alpha, \beta, \beta, \beta) \mid \alpha, \beta \in H^2(X, \mathbb{Z})\} \\ \ell_7 &= \gcd\{\Lambda(\alpha, \alpha, \alpha, \alpha) \mid \alpha \in H^2(X, \mathbb{Z})\}. \end{aligned} \tag{11}$$

Ref. [17] also provides a practical way of computing these invariants which involves a scan over only a finite subset of  $H^2(X, \mathbb{Z})$ , so that they can be worked out from  $d_{ijk}$  and  $c_{2,i}$ .

Combining the approaches of §2.1 and §2.2 and using invariants  $\ell_{i=1, \dots, 7}$  a plausible ansatz for  $x := \log N$  is

$$x = \sum_{i=1}^7 (A_i + B_i h) \log \ell_i, \tag{12}$$

where  $A_i$  and  $B_i$  are constants to be determined by regression. Using the data from the CICY scan, together with the Hodge numbers  $h$  and the invariants  $\ell_i$  for each CICY involved, the best fit values of  $A_i$  and  $B_i$  are

$$\begin{aligned} (A_i) &\simeq (12.2, 0, 11.6, -3.5, -11.7, 1.6, 1.8) \\ (B_i) &\simeq (-2.3, 0, -1.9, 0.9, 2.0, -0.7, 0.2). \end{aligned} \tag{13}$$

The comparison of this fit with the data is provided in Fig. 4. Each point corresponds to a CICY with  $x$  is the value computed from the RHS of Eq. (12), using the values in (13) for  $A_i$ ,  $B_i$  and  $\log(N)$  is the value of standard models on this CICY found by the computer scan. The red line is the diagonal,  $\log(N) = x$ , which represents a perfect fit. It seems from Eq. (13) that  $x$  depends most heavily on  $\ell_{1,3,5}$  as the corresponding coefficients dominate in magnitude. Examining these, we see that they essentially come from the basic intersection form (including the self-intersection), which is after all our most fundamental topological quantity.

It turns out that the invariants  $\ell_i$  show relatively little variation with increasing Picard number  $h$ . If we insert typical values for these quantities, together with  $h_{\text{CICY}} = 19$  and  $h_{\text{max}} = 491$  into the above fit result the extrapolated numbers become

$$N(h_{\text{CICY}}) \simeq 10^{23}, \quad N(h_{\text{max}}) \simeq 10^{723}, \quad (14)$$

which is not too far away from the earlier result (3).

### 3. Conclusions

The fit illustrated in Fig. 4 looks rather convincing and we believe that an extrapolation to large  $h = h^{1.1}$  is trustworthy (since the invariants  $\ell_i$  show little variation relative to  $h$ ), as long as the underlying model-building assumptions continue to be satisfied for large  $h$ . We believe that this is the case for the CICY dataset and, hence, the number of  $\simeq 10^{23}$  standard models within this set should be taken seriously.

The extrapolation to  $h_{\text{max}} = 491$ , the maximal known Picard number of any CY, is more questionable. Some of the model building assumptions made here have not yet been checked for the KS set, and there are even indications that they may not be satisfied. First, it is not clear that the KS set contains manifolds that admit freely-acting symmetries with the same frequency as CICYs. The only systematic checks carried out for low  $h^{1.1}$  where the frequency of symmetries is lower than the CICYs [30]. No information is available for large  $h^{1.1}$  yet.

Another generic feature of CICY models is the frequent absence of large numbers of vector-like pairs, so that checking the index was sufficient to guarantee the correct spectrum for a significant fraction of the models. It is not clear that this feature persists for constructions based on other CY manifolds. In fact, the results of Ref. [31] suggest that the presence of phenomenologically problematic numbers of vector-like pairs might be a generic feature of some other CY constructions. Again, no definite statement on this is available for the KS set.

In summary, the number of  $\simeq 10^{723}$  standard models for the extrapolation to  $h^{1.1} = 491$  should be viewed with considerable caution. However, the extrapolation to  $h^{1.1} = 19$ , the maximal Picard number in the CICY set has to be taken seriously and leads to  $\simeq 10^{23}$  standard models. Even this number, almost certainly a conservative lower bound, is frighteningly large and beyond current computer storage and systematic search.

### References

- [1] P. Candelas, G. Horowitz, A. Strominger, E. Witten, Nucl. Phys. B 258 (1985) 46.
- [2] B. Greene, K. Kirklín, P. Miron, G. Ross, Phys. Lett. B 180 (1986) 69.
- [3] V. Braun, Y.H. He, B.A. Ovrut, T. Pantev, Phys. Lett. B 618 (2005) 252, arXiv:hep-th/0501070.
- [4] V. Braun, Y.H. He, B.A. Ovrut, T. Pantev, J. High Energy Phys. 0605 (2006) 043, arXiv:hep-th/0512177.
- [5] V. Bouchard, R. Donagi, Phys. Lett. B 633 (2006) 783, arXiv:hep-th/0512149.
- [6] V. Braun, Y.H. He, B.A. Ovrut, J. High Energy Phys. 0606 (2006) 032, arXiv:hep-th/0602073.
- [7] L.B. Anderson, J. Gray, Y.H. He, A. Lukas, J. High Energy Phys. 1002 (2010) 054, arXiv:0911.1569 [hep-th].
- [8] L.B. Anderson, Y.H. He, A. Lukas, J. High Energy Phys. 0707 (2007) 049, arXiv:hep-th/0702210.
- [9] L.B. Anderson, J. Gray, A. Lukas, E. Palti, Phys. Rev. D 84 (2011) 106005, arXiv:1106.4804 [hep-th].
- [10] L.B. Anderson, J. Gray, A. Lukas, E. Palti, J. High Energy Phys. 1206 (2012) 113, arXiv:1202.1757 [hep-th]; L.B. Anderson, J. Gray, A. Lukas, E. Palti, PoS CORFU 2011 (2011) 096.
- [11] L.B. Anderson, A. Constantin, J. Gray, A. Lukas, E. Palti, J. High Energy Phys. 1401 (2014) 047, arXiv:1307.4787 [hep-th].
- [12] K.R. Dienes, Phys. Rev. D 73 (2006) 106010, arXiv:hep-th/0602286; K.R. Dienes, M. Lennek, D. Senechal, V. Wasnik, Phys. Rev. D 75 (2007) 126005, arXiv:0704.1320 [hep-th]; K.R. Dienes, M. Lennek, Phys. Rev. D 75 (2007) 026008, arXiv:hep-th/0610319.
- [13] F. Gmeiner, R. Blumenhagen, G. Honecker, D. Lust, T. Weigand, J. High Energy Phys. 0601 (2006) 004, arXiv:hep-th/0510170.
- [14] M.R. Douglas, J. High Energy Phys. 0305 (2003) 046, arXiv:hep-th/0303194.
- [15] A. Constantin, A. Lukas, C. Mishra, J. High Energy Phys. 1603 (2016) 173, arXiv:1509.02729 [hep-th].
- [16] P. Candelas, A.M. Dale, C.A. Lutken, R. Schimmrigk, Nucl. Phys. B 298 (1988) 493; P. Candelas, C.A. Lutken, R. Schimmrigk, Manifolds, Nucl. Phys. B 306 (1988) 113; M. Gagnon, Q. Ho-Kim, Mod. Phys. Lett. A 9 (1994) 2235.
- [17] T. Hübsch, Calabi-Yau Manifolds, World Scientific, 1992.
- [18] A.C. Avram, M. Kreuzer, M. Mandelberg, H. Skarke, Nucl. Phys. B 505 (1997) 625, arXiv:hep-th/9703003; M. Kreuzer, H. Skarke, Adv. Theor. Math. Phys. 4 (2002) 1209, arXiv:hep-th/0002240.
- [19] M. Kreuzer, H. Skarke, Comput. Phys. Commun. 157 (2004) 87, arXiv:math/0204356; A.P. Braun, J. Knapp, E. Scheidegger, H. Skarke, N.O. Walliser, arXiv:1205.4147.
- [20] R. Altman, J. Gray, Y.H. He, V. Jejjala, B.D. Nelson, J. High Energy Phys. 1502 (2015) 158, arXiv:1411.1418 [hep-th].
- [21] R. Altman, J. Carifio, J. Halverson, B.D. Nelson, arXiv:1811.06490 [hep-th].
- [22] R. Slansky, Phys. Rep. 79 (1981) 1.
- [23] A. Constantin, A. Lukas, arXiv:1808.09992.
- [24] D. Klaeuer, L. Schlechter, arXiv:1809.02547 [hep-th].
- [25] A. Constantin, arXiv:1808.09993 [hep-th].
- [26] E.I. Buchbinder, A. Constantin, A. Lukas, J. High Energy Phys. 1403 (2014) 025, arXiv:1311.1941 [hep-th].
- [27] E.I. Buchbinder, A. Constantin, A. Lukas, J. High Energy Phys. 1406 (2014) 100, arXiv:1404.2767 [hep-th].
- [28] E.I. Buchbinder, A. Constantin, A. Lukas, Phys. Lett. B 748 (2015) 251–254, arXiv:1409.2412 [hep-th].
- [29] T.D. Browning, D.R. Heath-Brown, P. Salberger, arXiv:math/0410117.
- [30] A. Braun, A. Lukas, C. Sun, Commun. Math. Phys. 360 (3) (2018) 935, arXiv:1704.07812.
- [31] A.P. Braun, C.R. Brodie, A. Lukas, J. High Energy Phys. 1804 (2018) 087, arXiv:1706.07688.
- [32] Y.H. He, S.J. Lee, A. Lukas, C. Sun, J. High Energy Phys. 1406 (2014) 077, arXiv:1309.0223 [hep-th].
- [33] V. Braun, J. High Energy Phys. 1104 (2011) 005, arXiv:1003.3235 [hep-th].
- [34] A. Lukas, C. Mishra, arXiv:1708.08943.
- [35] P. Candelas, C. Mishra, Fortschr. Phys. 66 (2018) 1800017, arXiv:1709.01081 [hep-th].
- [36] P. Candelas, A. Constantin, C. Mishra, Fortschr. Phys. 66 (2018) 1800029, arXiv:1602.06303 [hep-th].
- [37] P. Candelas, R. Davies, Fortschr. Phys. 58 (2010) 383, arXiv:0809.4681 [hep-th]; P. Candelas, A. Constantin, Fortschr. Phys. 60 (2012) 345–369, arXiv:1010.1878; P. Candelas, A. Constantin, C. Mishra, Fortschr. Phys. 64 (6–7) (2016) 463, arXiv:1511.01103.
- [38] P. Candelas, X. de la Ossa, Y.H. He, B. Szendroi, Adv. Theor. Math. Phys. 12 (2) (2008) 429, arXiv:0706.3134.
- [39] A. Constantin, J. Gray, A. Lukas, J. High Energy Phys. 01 (2017) 001, arXiv:1607.01830.