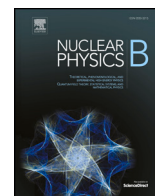




ELSEVIER





Contents lists available at ScienceDirect

Nuclear Physics, Section B

journal homepage: www.elsevier.com/locate/nuclphysb

Frontiers Article

Computation of quark masses from string theory

Andrei Constantin^{a, }, Cristoforo S. Fraser-Taliente^{a, }*, Thomas R. Harvey^{a, },
Andre Lukas^{a, }, Burt Ovrut^{b, }^a Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road Oxford, OX1 3PU, UK^b Department of Physics, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Editor: Stephan Stieberger

Dedicated to the memory of Graham G. Ross

ABSTRACT

We present a numerical computation, based on neural network techniques, of the physical Yukawa couplings in a heterotic string theory compactification on a smooth Calabi-Yau threefold with non-standard embedding. The model belongs to a large class of heterotic line bundle models that have previously been identified and whose low-energy spectrum precisely matches that of the MSSM plus fields uncharged under the Standard Model group. The relevant quantities for the calculation, that is, the Ricci-flat Calabi-Yau metric, the Hermitian Yang-Mills bundle metrics and the harmonic bundle-valued forms, are all computed by training suitable neural networks. For illustration, we consider a one-parameter family in complex structure moduli space. The computation at each point along this locus takes about half a day on a single twelve-core CPU. Our results for the Yukawa couplings are estimated to be within 10% of the expected analytic result. We find that the effect of the matter field normalisation can be significant and can contribute towards generating hierarchical couplings. We also demonstrate that a zeroth order, semi-analytic calculation, based on the Fubini-Study metric and its counterparts for the bundle metric and the bundle-valued forms, leads to roughly correct results, about 25% away from the numerical ones. The method can be applied to other heterotic line bundle models and generalised to other constructions, including to F-theory models.

1. Introduction

Computing the values of the quark and lepton masses and understanding their hierarchical structure from first principles is a long-standing and fundamental open problem in theoretical particle physics. Arguably, string theory currently provides the only framework which allows for such a computation. However, the low-energy particle content obtained from string compactification was, for a long time, not sufficiently realistic to warrant detailed computations of couplings. Now that many models with the Standard Model spectrum are available, particularly in the context of $E_8 \times E_8$ heterotic string compactifications on Calabi-Yau (CY) threefolds with holomorphic, poly-stable vector bundles [1–22], computing Yukawa couplings and the resulting fermion masses and mixing angles is the obvious next step.

The physical Yukawa couplings are completely specified by two moduli-dependent quantities in the low-energy four-dimensional $N = 1$ supersymmetric Lagrangian: the holomorphic Yukawa couplings, which arise in the superpotential, and the matter field Kähler

* Corresponding author.

E-mail addresses: andrei.constantin@physics.ox.ac.uk (A. Constantin), cristoforo.fraser-taliente@physics.ox.ac.uk (C.S. Fraser-Taliente), thomas.harvey@physics.ox.ac.uk (T.R. Harvey), lukas@physics.ox.ac.uk (A. Lukas), ovrut@physics.upenn.edu (B. Ovrut).<https://doi.org/10.1016/j.nuclphysb.2024.116778>

Received 6 August 2024; Accepted 12 December 2024

Available online 17 December 2024

0550-3213/© 2024 The Author(s). Published by Elsevier B.V. Funded by SCOAP³. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

metric which determines the normalisations of the $N = 1$ chiral superfields. The holomorphic Yukawa couplings, defined in Eq. (3.1), are quasi-topological and do not depend on the Ricci-flat metric of the CY threefold. Consequently, they can often be calculated analytically using algebraic or differential geometric tools. This has been carried out for a number of models [23–27]. On the other hand, the computation of the matter field Kähler metric, defined in Eq. (3.2), requires full knowledge of the compactification geometry, including the Ricci-flat CY metric, the Hermitian Yang-Mills (HYM) connection on the holomorphic vector bundle, as well as various harmonic bundle-valued forms on the CY threefold. Obtaining these quantities has been the single major hurdle (other than finding models with a realistic particle spectrum) in the computation of Yukawa couplings from string theory for nearly four decades, as no closed-form expressions are known.

A salient exception to this rule is represented by standard embedding CY compactifications of the $E_8 \times E_8$ heterotic string (and singular limits thereof, such as toroidal orbifold compactifications) [28–32]. In this setting, the matter field Kähler metric and the holomorphic Yukawa couplings can be explicitly calculated from period integrals via special geometry or by using conformal field theory techniques. In fact, the pioneering work on fermion masses from string theory [33–36] was carried out in this context. These analytic results for models with standard embedding have recently been successfully reproduced, and extended, using neural networks methods [37]. Unfortunately, the class of standard embedding models is very limited in terms of what can be achieved, both at the level of the spectrum and couplings.

The methods presented in this paper apply to a wide class of phenomenologically attractive models relying on line bundle sums over smooth CY threefolds embedded in products of projective spaces with freely acting discrete symmetries [7–9], and may be further extended to compactifications with non-Abelian bundles. The core idea is to use neural networks to approximate the above-mentioned geometrical quantities, which are obtained as solutions of certain partial differential equations. As zeroth order ‘reference’ quantities (and for benchmarking) we use analytic expressions for the Fubini-Study Kähler form, the Chern connection and various bundle-valued differential forms, defined on the ambient space and subsequently restricted to the CY manifold, which represent the correct cohomology classes. To these reference quantities we add terms exact in cohomology which are represented by neural networks and optimised in order to obtain the Ricci-flat CY metric, the HYM bundle metrics and the harmonic forms.¹ The optimisation process involves sampling points on the CY threefold and minimising a loss function that takes into account how well the associated partial differential equations and various patching conditions are satisfied. Our methods build directly on the tools developed in the cymetric package [38,39] for the construction of numerical CY metrics, as well as earlier numerical work on: (a) numerical methods for Ricci-flat CY metrics based on Donaldson’s algorithm [40–42], functional minimisation [43], and machine learning techniques [38,39,44–49]; (b) numerical methods for the computation of Hermitian Yang-Mills connections [50–55]; (c) numerical harmonic functions and harmonic bundle-valued forms [55–58].

The main goal of this paper is to develop the neural-network approach up to the point where explicit values for the quark and lepton masses can be computed for arbitrary values of the moduli fields. In practice, we carry out the computation of perturbative up-quark Yukawa couplings in a specific heterotic line bundle model, originally introduced in Refs. [10,11], thus proving that such calculations are now feasible. A systematic exploration of the moduli space for this and other models is beyond the scope of the present letter, but likely within reach. The aim of this future work is to identify concrete line bundle models and specific loci in their moduli space that give rise to the observed flavour structure of the Standard Model. Combining such an analysis with moduli stabilisation may lead to a comprehensive explanation of the parameters in the Standard Model.

It is worth pointing out that the present work does not rely on any simplifying assumptions, such as localisation, the only limiting factor being numerical accuracy. Localisation techniques have been heavily used in other contexts, including calculations of zero-mode wavefunctions on toroidal backgrounds [59,60], for F-theory models [61–69] and for heterotic models [70]. For compactifications on non-flat spaces, localisation relies on the observation that sufficiently large fluxes lead to localised matter field wave functions, so that approximate calculations can be carried out with a (locally) flat metric. This appears to circumvent the need to know the Ricci-flat CY metric explicitly. However, this method comes with a number of problems: it is difficult to assess its accuracy and to express the results in terms of standard CY moduli and there is a tension between the large fluxes required for localisation and the requirements of three families and anomaly cancellation.

The structure of the paper is as follows. We begin in Section 2 by reviewing the details of the heterotic string model under consideration, before moving on to describing the mathematical background and the computational setup in Section 3. In Section 4, we present our results for the physical up-quark Yukawa couplings and resulting fermion masses. We conclude in Section 5. Further details of the calculation will be presented in a forthcoming longer paper [71].

2. The heterotic string model

The model underlying our computations was originally introduced in Ref. [10,11]. It is based on compactifying the $E_8 \times E_8$ heterotic string on a smooth $\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_2$ quotient of a CY hypersurfaces X of multi-degree $(2, 2, 2, 2)$ in a product of four \mathbb{P}^1 -spaces. These hypersurfaces are, from now on, referred to as tetra-quadric CY threefolds or TQ threefolds, for short. The quotient X/Γ has four Kähler parameters and 20 complex structure parameters. The $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry descends from the ambient space [72], where it is generated by the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (2.1)$$

¹ This is not to say we are computing the leading order correction in some expansion. The final results are precise, up to numerical error.

acting simultaneously on the homogeneous coordinates of each \mathbb{P}^1 .

The standard Kähler forms on the four \mathbb{P}^1 factors, restricted to X , provide a basis, (J_i) , where $i = 1, 2, 3, 4$, for the second cohomology of X . The first Chern class of a line bundle $\mathcal{L} \rightarrow X$ can be written as $c_1(\mathcal{L}) = k^i J_i$, where $k^i \in \mathbb{Z}$, and such a line bundle is also denoted by $\mathcal{L} = \mathcal{O}_X(\mathbf{k})$, with $\mathbf{k} = (k^1, k^2, k^3, k^4)^T \in \mathbb{Z}^4$.

The vector bundle $V \rightarrow X$ is chosen to be a sum of five line bundles, that is $V = \bigoplus_{a=1}^5 \mathcal{L}_a$, with Chern classes $c_1(\mathcal{L}_a) = k_a^i J_i$. This bundle breaks one of the two E_8 gauge factors to $SU(5) \times S(U(1)^5)$. For the model under considerations, the five line bundles are specified by the column vectors of the matrix

$$(k_a^i) = \begin{bmatrix} \mathcal{L}_1 & \mathcal{L}_2 & \mathcal{L}_3 & \mathcal{L}_4 & \mathcal{L}_5 \\ -1 & -1 & 0 & 1 & 1 \\ 0 & -3 & 1 & 1 & 1 \\ 0 & 2 & -1 & -1 & 0 \\ 1 & 2 & 0 & -1 & -2 \end{bmatrix}. \quad (2.2)$$

The additional $U(1)$ -symmetries are Green-Schwarz anomalous and their associated gauge bosons are super-heavy. At low energy they appear as global symmetries which constrain the allowed couplings. Finally, the $SU(5)$ -gauge symmetry is broken to the Standard Model gauge group by specifying a discrete Wilson line with structure group $\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_2$. The model is free of gauge and gravitational anomalies (by a suitable choice of a five-brane or hidden bundle) and whatever remains from the second E_8 gauge group at low energy is ‘hidden’, in the sense that all observable fields are uncharged under it. It is also supersymmetric along the locus where all Kähler moduli take the same value, that is,

$$t := t_1 = t_2 = t_3 = t_4. \quad (2.3)$$

The particle content is that of the MSSM, plus a number of singlet fields uncharged under the SM gauge group. These fields are decorated with $U(1)$ charges and given by

$$\begin{array}{ll} 2Q_2, 2U_2, 2E_2 \leftrightarrow \mathcal{L}_2 & Q_5, U_5, E_5 \leftrightarrow \mathcal{L}_5 \\ 2D_{2,5}, 2L_{2,5} \leftrightarrow \mathcal{L}_4 \otimes \mathcal{L}_5 & D_{2,4}, L_{2,4} \leftrightarrow \mathcal{L}_2 \otimes \mathcal{L}_4 \\ H_{2,5}^d \leftrightarrow \mathcal{L}_2 \otimes \mathcal{L}_5 & H_{2,5}^u \leftrightarrow \mathcal{L}_2^* \otimes \mathcal{L}_5^* \\ 3S_{2,4} \leftrightarrow \mathcal{L}_2 \otimes \mathcal{L}_4^* & 12 \text{ other singlets} \end{array} \quad (2.4)$$

The subscripts label the $U(1)$ symmetries under which the particles carry charge 1, while being uncharged under all other $U(1)$ symmetries. The exceptions are $H_{2,5}^u$, whose only non-zero charges are -1 under the second and fifth $U(1)$ symmetry, and $S_{2,4}$ with charge $+1$ under the second symmetry and charge -1 under the fourth.² The fields are in one-to-one correspondence with certain harmonic bundle-valued one-form on specific line bundles, which have been indicated in Eq. (2.4).

The $U(1)$ symmetries enforce the vanishing of down-quark and lepton Yukawa matrices at the perturbative level; for a realistic model, they would have to be generated non-perturbatively. Writing the left-handed quarks as $(Q^i) = (Q_2^1, Q_2^2, Q_5)$, the right-handed up-quarks as $(U^i) = (U_2^1, U_2^2, U_5)$ and the up-Higgs as $H^u = H_{2,5}^u$, the holomorphic up-quark Yukawa couplings are of the form

$$W_u = Y_{ij}^u H^u Q^i U^j, \quad (2.5)$$

where $i, j = 1, 2, 3$ label the three quark families. The $U(1)$ symmetries enforce a specific structure of the up-Yukawa matrix given by

$$Y^u = \begin{pmatrix} 0 & 0 & \lambda_1 \\ 0 & 0 & \lambda_2 \\ \lambda_3 & \lambda_4 & 0 \end{pmatrix}. \quad (2.6)$$

Its entries $\lambda_1, \dots, \lambda_4$ are quasi-topological and can be computed using differential geometric techniques, as detailed in Refs. [25–27].³ For completeness, we note that the full perturbative superpotential is

$$W = W_u + \rho_{\alpha i} S_{2,4}^\alpha L_{4,5}^i H_{2,5}^u, \quad (2.7)$$

where the index $\alpha = 1, 2, 3$ labels the three singlets $S_{2,4}$ present in the spectrum. These are interpreted as right-handed neutrinos.

The part of the Kähler potential relevant for the calculation of the physical up-Yukawa couplings has the form

$$K = K_{ij}^Q Q^i \bar{Q}^j + K_{ij}^U U^i \bar{U}^j + k H^u \bar{H}^u,$$

and $U(1)$ -invariance dictates the following structure for the Kähler metrics:

² Note that the up and down Higgs triplets have been projected out by the $\mathbb{Z}_2 \times \mathbb{Z}_2$ quotient and the inclusion of the Wilson line.

³ Unfortunately, there appears to be a mistake in the calculation carried out in Ref. [25] of the holomorphic Yukawa couplings for this model, due to a missed boundary term, an issue which we correct in the present paper.

$$K^Q = \mathcal{V}^{-\frac{1}{3}} \begin{pmatrix} \mathcal{K}^Q & 0 \\ 0 & 0 \\ 0 & 0 & k^Q \end{pmatrix}, K^u = \mathcal{V}^{-\frac{1}{3}} \begin{pmatrix} \mathcal{K}^u & 0 \\ 0 & 0 \\ 0 & 0 & k^u \end{pmatrix}. \quad (2.8)$$

The factor $\mathcal{V}^{-1/3}$ captures the full Kähler moduli dependence in this model due to Eq. (2.3). This means the complex 2×2 matrices \mathcal{K}^Q , \mathcal{K}^u and the real numbers k , k^Q , k^u in Eq. (2.8) are Kähler moduli independent, but they still depend on complex structure. After bringing the resulting kinetic terms into canonical form, one finds the physical up-Yukawa matrix

$$Y_{\text{phys}}^u = \begin{pmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ b_1 & b_2 & 0 \end{pmatrix} \quad (2.9)$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{e^{-\phi}}{\sqrt{kk^u}} P_Q \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \quad \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \frac{e^{-\phi}}{\sqrt{kk^Q}} P_u \begin{pmatrix} \lambda_3 \\ \lambda_4 \end{pmatrix},$$

where ϕ is the dilaton, and P_Q and P_u are 2×2 diagonalising matrices satisfying

$$P_Q \mathcal{K}^Q P_Q^\dagger = \mathbb{1}_2, \quad P_u \mathcal{K}^u P_u^\dagger = \mathbb{1}_2. \quad (2.10)$$

Note that the volume dependence of the physical Yukawa couplings drops out due to the additional factor of $e^{-\phi}/\sqrt{\mathcal{V}}$ which arises from the $\exp(K/2)$ prefactor to the Yukawa couplings in the component supergravity Lagrangian.⁴ Finally, the up-quark masses are

$$(m_1, m_2, m_3) = |\langle H^u \rangle| e^{-\phi} \left(0, \frac{\left| P_Q \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \right|}{\sqrt{kk^u}}, \frac{\left| P_u \begin{pmatrix} \lambda_3 \\ \lambda_4 \end{pmatrix} \right|}{\sqrt{kk^Q}} \right). \quad (2.11)$$

A number of comments are in order at this point. Firstly, in this model, the holomorphic up-quark Yukawa matrix (2.6) on its own cannot lead to a non-zero mass for the first generation due to its reduced rank. A non-zero up-quark mass would have to be generated non-perturbatively. Secondly, a potential split between the second and third generation up-quark masses can be induced by the structure of the holomorphic up-Yukawa matrix as well as by the non-canonical matter field Kähler metric, both of which depend on complex structure moduli. The dependence of the physical Yukawa couplings on the Kähler moduli completely drops out in this particular model, as a consequence of the volume-independence mentioned earlier, and of working at the special locus Eq. (2.3) in Kähler moduli space. Thirdly, in order to make contact with the measured values of the quark masses, one should include the RG running from the compactification scale down to the electroweak scale in the presence of supersymmetry breaking. This is amenable to standard methods and will not be discussed further. Finally, loop, α' , and non-perturbative corrections can affect the physical Yukawa couplings. These are suppressed in the large volume and weak coupling regime and, in these limits, will only lead to small corrections to non-zero perturbative masses (of course, they may be the leading effect if a perturbative mass vanishes, as in our present example). For precise predictions from string theory, all these corrections must ultimately be considered.

3. Metrics and harmonic forms

In geometric heterotic compactifications on smooth CY threefolds X/Γ , the holomorphic Yukawa couplings and the matter field Kähler metric are given by the following expressions:

$$\lambda_{IJK} = \gamma_{IJK} \frac{c 2\sqrt{2}}{|\Gamma|} \int_X \nu_I \wedge \nu_J \wedge \nu_K \wedge \Omega, \quad (3.1)$$

$$K_{IJ} = \frac{\zeta_{IJ}}{2\mathcal{V}|\Gamma|} \int_X \nu_I \wedge \star_{\mathcal{V}} \nu_J = \frac{\zeta_{IJ}}{2\mathcal{V}|\Gamma|} \int_X \nu_I \wedge \star(H_J \bar{\nu}_J). \quad (3.2)$$

Here ν_I are harmonic $(0,1)$ -forms which represent the matter fields and take values in line bundles \mathcal{L}_I , whilst H_I are HYM bundle metrics on \mathcal{L}_I . Concretely, for the model outlined in the previous section, the line bundles \mathcal{L}_I are the ones given in Eq. (2.4). Furthermore, \mathcal{V} is the CY volume and the Hodge star is taken with respect to the Ricci-flat CY metric. The holomorphic $(3,0)$ -form Ω on X is normalised such that $\int_X \Omega \wedge \bar{\Omega} = 1$. The integrals are performed on the ‘upstairs’ manifold X and the result is transferred to the smooth ‘downstairs’ quotient X/Γ by dividing by the group order, $|\Gamma|$. Concretely, for our specific model, we have $|\Gamma| = |\mathbb{Z}_2 \times \mathbb{Z}_2| = 4$. In line with the constraints from the low-energy $U(1)$ symmetries, holomorphic Yukawa couplings λ_{IJK} can be non-zero only if $\mathcal{L}_I \otimes \mathcal{L}_J \otimes \mathcal{L}_K = \mathcal{O}_X$ and entries K_{IJ} for $\mathcal{L}_I \neq \mathcal{L}_J$ must vanish. The numerical pre-factors arise from the dimensional reduction, whilst the factor $c = \sqrt{H_I H_J H_K}$ originates from transforming between conventions used in the physics and mathematics literature [51]. Note that $H_I H_J H_K$ is constant for any Yukawa coupling allowed by the $U(1)$ symmetries. The factors ζ_{IJ} and γ_{IJK} are group theoretic factors, coming from the branching of the 10D E_8 gauge group. Their derivation will be given in the upcoming paper, and for the case of up-quark Yukawa couplings they are given by

⁴ Note that this is a general feature. The physical Yukawa couplings in heterotic theories are independent of the overall CY volume modulus.

$$\zeta_{IJ} = \frac{\delta_{IJ}}{2}, \quad \gamma_{IJK} = \frac{1}{8\sqrt{30}}. \quad (3.3)$$

We write γ_{IJK} as a constant as it takes the same value, up to a sign, for all terms allowed by the symmetries. The signs are such to cancel the anti-symmetry of the forms inside (3.1).

As mentioned in the previous section, to compute the Ricci-flat CY metric g , the HYM bundle metrics H_I and the harmonic forms v_I , we start with certain reference quantities which represent the correct cohomology classes. To these, we add exact terms which are determined by training suitable neural networks. We now describe how this is done for each of the three types of quantities in turn.

3.1. The Ricci-flat CY metric

3.1.1. Mathematical background

Yau's theorem, applied to a CY manifold X , asserts that in any given Kähler class, associated to a reference metric $g^{(\text{ref})}$, there exists a unique Ricci-flat metric

$$g_{a\bar{b}} = g_{a\bar{b}}^{(\text{ref})} + \partial_a \bar{\partial}_{\bar{b}} \phi, \quad (3.4)$$

where ϕ is a real function on X determined by solving the relevant Monge-Ampère equation. In practice, this can be done by training a neural network which represents ϕ . This approach has been realised in the cymetric package [38,39], where it is referred to as the ‘ ϕ -model’.

Our specific simply-connected CY threefold is defined as a tetra-quadric (TQ) hypersurface given as the zero locus of a defining polynomial p of multi-degree $(2, 2, 2, 2)$, in the ambient space $\mathcal{A} = \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. Homogeneous coordinates on the four \mathbb{P}^1 s are denoted by $x_\alpha, y_\alpha, u_\alpha, v_\alpha$, where $\alpha = 0, 1$. The standard patches on \mathcal{A} with $x_\alpha, y_\beta, u_\gamma, v_\delta \neq 0$ are denoted by $U_{\alpha\beta\gamma\delta}$ and affine coordinates on the patch U_{0000} are defined by $z_1 = x_1/x_0$, $z_2 = y_1/y_0$, $z_3 = u_1/u_0$ and $z_4 = v_1/v_0$. We also introduce the convenient shorthand $\kappa_i = 1 + |z_i|^2$.

Computing the holomorphic Yukawa couplings (3.1) requires the holomorphic $(3, 0)$ -form Ω . On the standard patch with coordinates z_i defined above, it can be written explicitly as

$$\Omega \propto \hat{\Omega} = \left. \frac{dz_1 \wedge dz_2 \wedge dz_3}{\frac{\partial p}{\partial z_4}} \right|_X, \quad (3.5)$$

with the proportionality constant fixed by $\int_X \Omega \wedge \bar{\Omega} = 1$, up to an arbitrary phase that drops out of physical quantities.

For the reference metric in Eq. (3.4) we choose the Fubini-Study metric restricted to X :

$$g_{a\bar{b}}^{(\text{ref})} = \left. \sum_{i=1}^4 \frac{t^i}{2\pi} \partial_a \bar{\partial}_{\bar{b}} \ln(\kappa_i) \right|_X, \quad (3.6)$$

where $t^i \in \mathbb{R}^{>0}$ are the four Kähler parameters. In terms of these parameters, the CY volume \mathcal{V} reads:

$$\mathcal{V} = 2(t_1 t_2 t_3 + t_1 t_2 t_4 + t_1 t_3 t_4 + t_2 t_3 t_4). \quad (3.7)$$

At the supersymmetric locus (2.3), this expression simplifies to $\mathcal{V} = 8t^3$, with the overall Kähler parameter t .

For most of the calculations below, we will be working with the two-parameter family of TQs defined by the vanishing of the polynomial

$$p = \sum_{\substack{\text{even} \\ \alpha+\beta+\delta+\gamma}} x_\alpha^2 y_\beta^2 u_\gamma^2 v_\delta^2 + \psi_0 \sum_{\substack{\text{odd} \\ \alpha+\beta+\delta+\gamma}} x_\alpha^2 y_\beta^2 u_\gamma^2 v_\delta^2 + \psi \prod_{\alpha,\beta,\delta,\gamma} x_\alpha y_\beta u_\gamma v_\delta, \quad (3.8)$$

constructed in analogy with the Dwork pencil of quintic threefolds. For $\psi_0 \neq 1$, this leads to a smooth hypersurface for generic values of ψ . This polynomial is of course invariant under the action (2.1) of $\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_2$. However, it is also invariant under an additional symmetry which enforces equality between the two non-zero up-quark masses, a degeneration reflected in the numerical calculation below (see, for instance, Fig. 4). To illustrate that this degeneracy can be lifted, we also consider a more general $\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_2$ invariant polynomial which breaks the additional symmetry present in Eq. (3.8), namely,

$$\begin{aligned}
\tilde{p} = & 6x_1^2 y_0 y_1 u_0 u_1 v_0^2 + x_0^2 y_0 y_1 u_0 u_1 v_0^2 - 18x_0^2 u_0 u_1 v_0 v_1 y_0^2 \\
& - 2x_0 x_1 y_1^2 v_0 v_1 u_0^2 + 7x_1^2 y_0 y_1 v_0 v_1 u_0^2 - 11x_0^2 y_0^2 u_0^2 v_0^2 \\
& - 2x_0 x_1 y_0^2 v_0 v_1 u_0^2 + 7x_1^2 y_1^2 u_1^2 v_0^2 - 12x_0^2 y_1^2 u_1^2 v_0^2 \\
& - 6x_0 x_1 y_0 y_1 u_1^2 v_0^2 + 20x_1^2 y_0^2 u_1^2 v_0^2 + 4x_0^2 y_0^2 u_1^2 v_0^2 \\
& - 13x_0 x_1 y_1^2 u_0 u_1 v_0^2 - 9x_0 x_1 y_0^2 u_0 u_1 v_0^2 + x_1^2 y_0^2 u_0^2 v_0^2 \\
& - 15x_1^2 y_1^2 u_0^2 v_0^2 - 4x_0^2 y_1^2 u_0^2 v_0^2 - 4x_0 x_1 y_0 y_1 u_0^2 v_0^2 \\
& - 13x_0^2 y_0 y_1 v_0 v_1 u_0^2 + (1 \leftrightarrow 0).
\end{aligned} \tag{3.9}$$

Here, $(1 \leftrightarrow 0)$ is the instruction to copy over all the previous terms with coordinate indices swapped as indicated.

Throughout the entire calculation below, we will set the overall Kähler parameter to $t = 1$. Recall that in the present model the physical Yukawa couplings are independent of the Kähler parameters, so this choice does not limit the scope of our calculation.

3.1.2. Computational realisation

An approximate Ricci-flat CY metric is constructed via Eq. (3.4), where the function ϕ is represented by a neural network and trained on a loss function that includes the Monge-Ampère loss, the transition loss and the Kähler class loss:

$$\begin{aligned}
\mathcal{L} &= \alpha_1 \mathcal{L}_{\text{MA}} + \alpha_2 \mathcal{L}_{\text{tr.}} + \alpha_3 \mathcal{L}_{\text{Kähler}} \\
\mathcal{L}_{\text{MA}}[\phi] &= \left\| \left| 1 - \frac{1}{\kappa} \frac{J(\phi) \wedge J(\phi) \wedge J(\phi)}{\hat{\Omega} \wedge \bar{\hat{\Omega}}} \right| \right\|_1 \\
\mathcal{L}_{\text{tr.}}[\phi] &= \sum_{s \neq t} \left\| \left| \phi_s - \phi_t \right| \right\|_1.
\end{aligned} \tag{3.10}$$

Here $\alpha_1, \alpha_2, \alpha_3$ are weights for the three contributions and $\|\cdot\|_1$ is the L_1 -norm computed by performing Monte-Carlo integration on X as detailed in Ref. [39]. Furthermore, $J(\phi)$ is the Kähler form associated to the metric (3.4), $\hat{\Omega}$ is defined in Eq. (3.5), ϕ_s is the version of ϕ computed on the patch U_s , while the normalisation factor κ and the Kähler class loss $\mathcal{L}_{\text{Kähler}}$ are defined in Ref. [39].

We carry out the computation of the CY metric at the points along the pencil of tetra-quadratics (3.8) defined by $\psi_0 = 2$ and for $\psi \in \{0, 0.5, 1, 2, 4, 6\}$. For each value of ψ , the cymetric package [38,39] is used to create a sample of 300,000 points on X , distributed according to the measure defined in Ref. [39]. These points are used to train the neural network for ϕ as well as the neural networks discussed below and to perform Monte-Carlo integration over X . We split the point sample into training and validation sets at a ratio of 9:1.

The neural network is fully connected with GeLU activation [73], four layers and a width of 128. Training is carried out for 100 epochs, with batch size 64 and learning rate 0.001. The change in loss over the course of a typical training round with $\psi_0 = 2$ and $\psi = 1$ is illustrated in Fig. 1a and Fig. 1b. We define the following measures for the training performance,

$$M_{\text{MA}}[\phi] = \frac{\mathcal{L}_{\text{MA}}[\phi]}{\mathcal{L}_{\text{MA}}[\phi = 0]}, \tag{3.11}$$

$$M_{\text{tr.}}[\phi] = \frac{1}{\mathcal{V}} \frac{\mathcal{L}_{\text{tr.}}[\phi]}{\text{std. dev. of } \phi}, \tag{3.12}$$

where the integration is performed over the validation set. For a typical run with $\psi_0 = 2$ and $\psi = 1$, we find $M_{\text{MA}}(\phi) = 0.04$ and $M_{\text{tr.}}(\phi) = 0.05$. These values indicate that the neural network ϕ performs well and leads to a good approximation of the Ricci-flat CY metric.

3.2. The HYM bundle metric

3.2.1. Mathematical background

Let $\mathcal{L} = \mathcal{O}_X(\mathbf{k})$ be a line bundle on X associated with one of the matter fields in (2.4). To determine the correct field normalisations, knowledge of the HYM metric on $\mathcal{O}_X(\mathbf{k})$ is required. As reference metric $H^{(\text{ref})}$, we choose the standard bundle metric associated with the Chern connection on $\mathcal{O}_{\mathcal{A}}(\mathbf{k})$ restricted to X . This can be written down explicitly as

$$H^{(\text{ref})} = \prod_{i=1}^4 \kappa_i^{-k^i} \Big|_X. \tag{3.13}$$

The HYM bundle metric H on $\mathcal{O}_X(\mathbf{k})$ is related to $H^{(\text{ref})}$ by

$$H = e^\beta H^{(\text{ref})}, \tag{3.14}$$

where β is a real function on X . The HYM equation implies that β must satisfy the following Poisson equation

$$\Delta\beta = \rho_\beta = -g^{a\bar{b}} \partial_a \bar{\partial}_b \ln(\bar{H}^{(\text{ref})}). \tag{3.15}$$

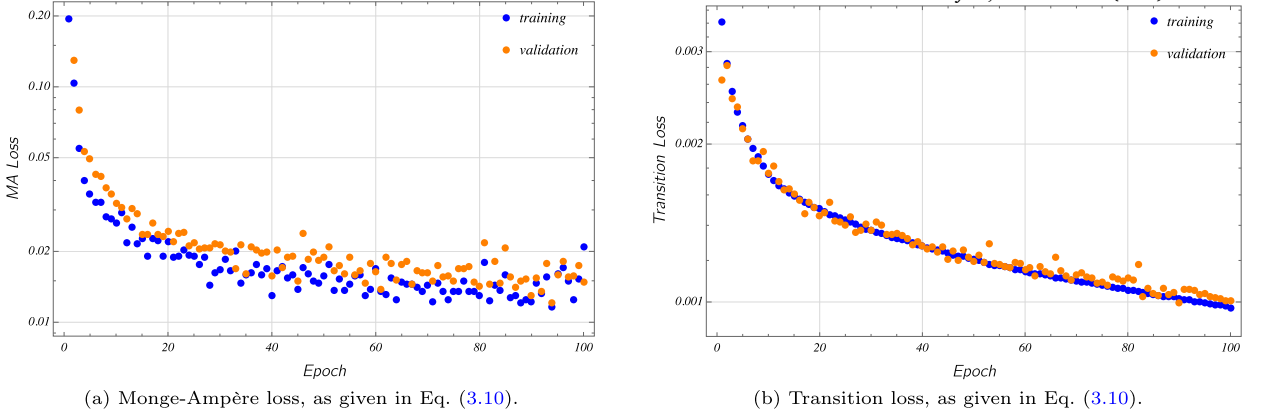


Fig. 1. Monge-Ampère and transition losses for the neural network approximating the function ϕ defined in Eq. (3.4). For this example, we have chosen $\psi_0 = 2$ and $\psi = 1$.

The integrability condition for this equation amounts to the requirement that the slope of the line bundle vanishes, which is guaranteed by working at the supersymmetric Kähler locus in Eq. (2.3). From the spectrum (2.4), in order to determine the up-quark Yukawa couplings, we need to compute the three Hermitian bundle metrics on \mathcal{L}_2 , \mathcal{L}_5 and $\mathcal{L}_2^* \otimes \mathcal{L}_5^*$.

3.2.2. Computational realisation

A possible method to solve Eq. (3.15), which has, for instance, been used in Ref. [39], is to construct a function basis, by starting with the sections of a given line bundle $\mathcal{L}' \rightarrow X$. Then, Eq. (3.15) can be converted into a linear system by expanding β in terms of this basis, and by computing the matrix elements of the Laplacian as well as the components of the source ρ_β . In our case, the simplest line bundle for this purpose is $\mathcal{L}' = \mathcal{O}_X(1, 1, 1, 1)$ with 16 sections, leading to $16^2 = 256$ functions. Working with this basis requires computing 256² Laplacian matrix elements. On the other hand, this is equivalent to working with the $l = 0, 1$ spherical harmonics in each of the $\mathbb{P}^1 \cong S^2$ directions only, which indicates a rather poor approximation. Trying to improve this by starting with $\mathcal{L}' = \mathcal{O}_X(2, 2, 2, 2)$ instead requires the calculation of an unfeasible number of matrix elements. In conclusion, it appears difficult to achieve sufficient accuracy with this method.

For this reason, we have opted to determine the HYM bundle metric by training a neural network, in analogy with what has been done for the CY metric. In practice, the neural network approximating the function β in Eq. (3.14) has two contributions for the loss function: the HYM loss based on the failure to satisfy Eq. (3.15), and the transition loss which ensures that β transforms as a function. Explicitly, the loss function is given by

$$\begin{aligned} \mathcal{L} &= \alpha_1 \mathcal{L}_{\text{HYM}} + \alpha_2 \mathcal{L}_{\text{tr}}. \\ \mathcal{L}_{\text{HYM}}[\beta] &= \left\| \Delta\beta - \rho_\beta \right\|_1 \\ \mathcal{L}_{\text{tr}}[\beta] &= \sum_{s \neq t} \left\| \beta_s - \beta_t \right\|_1, \end{aligned} \tag{3.16}$$

where α_1, α_2 are weights for the two contributions, and β_s, β_t are the versions of β computed on the patches U_s, U_t , respectively.

The neural network is fully connected with GeLU activation, three layers and a width of 128. Training is carried out for 100 epochs, with batch size 64 and learning rate 0.001. Each of the three required bundle metrics is computed for $\psi_0 = 2$ and $\psi \in \{0, 0.5, 1, 2, 4, 6\}$. To discuss training and accuracy, we focus on the case $\psi = 1$ which turns out to be typical. A typical change in loss over the course of training is shown in Fig. 2a and Fig. 2b. Appropriate measures for the success or failure of training the neural networks can be defined by

$$M_{\text{HYM}}[\beta] = \frac{\mathcal{L}_{\text{HYM}}[\beta]}{\int_X |\rho|}, \tag{3.17}$$

$$M_{\text{tr}}[\beta] = \frac{1}{\mathcal{V}} \frac{\mathcal{L}_{\text{tr}}[\beta]}{\text{std. dev. of } \beta}, \tag{3.18}$$

where the integration goes over the validation set.

Typical numerical values for these measures, computed for the three line bundles, range between 0.05 – 0.06 for M_{HYM} and between 0.02 – 0.04 for M_{tr} . We expect the product of the three bundle metrics to give the trivial bundle metric. It follows that this product must be constant over the CY manifold, and this is indeed true within an error of about 4%.

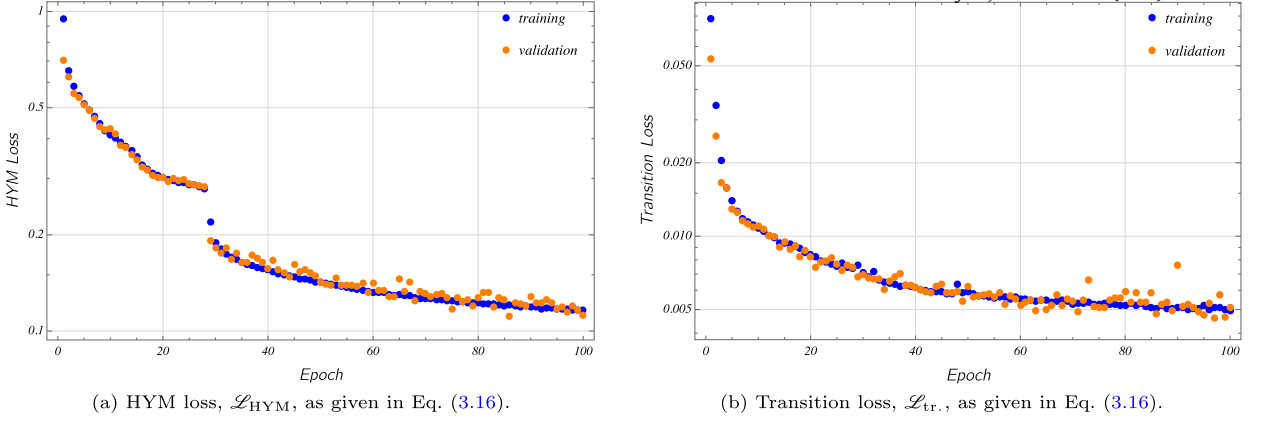


Fig. 2. HYM and transition losses for the neural network approximating the function β defined in Eq. (3.15). This example, typical of the cases we consider, solves for the bundle metric on the line bundle \mathcal{L}_2 . We have selected the values $\psi_0 = 2$ and $\psi = 1$.

3.3. The harmonic forms

3.3.1. Mathematical background

The matter fields in the spectrum (2.4) are associated with harmonic bundle-valued $(0, 1)$ -forms. Let \mathcal{L} be one of the relevant line bundles on X with HYM bundle metric $H = e^\beta H^{(\text{ref})}$. Each matter field is represented by a specific cohomology class in $H^1(X, \mathcal{L})$ and we choose a reference form $v^{(\text{ref})}$ to represent this class. The unique harmonic form v in the same cohomology class is related to $v^{(\text{ref})}$ by

$$v = v^{(\text{ref})} + \bar{\partial}_{\mathcal{L}} \sigma. \quad (3.19)$$

Here, σ is a global section of \mathcal{L} determined by the Poisson equation

$$\Delta_{\mathcal{L}} \sigma = \rho_{\sigma} = -g^{a\bar{b}} \partial_a \left(H v_b^{(\text{ref})} \right), \quad (3.20)$$

where $\Delta_{\mathcal{L}}$ is the Laplacian on \mathcal{L} relative to the Ricci-flat CY metric g and the HYM bundle metric H .

For our model, we need to calculate seven harmonic forms, one for each of the matter fields in Table 1. These are precisely the forms which survive the $\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_2$ projection and the inclusion of the Wilson line. Explicit expressions for these reference forms have been given in Refs. [25–27], which, being somewhat lengthy, will not be reproduced here.

3.3.2. Computational realisation

Following the same principles as before, we find the harmonic bundle-valued form v from Eq. (3.19), by designing a neural network which approximates σ , trained on a loss function with two contributions: the Laplacian loss, which measures the failure to satisfy Eq. (3.20), and the transition loss, which ensures that σ transforms as a section of \mathcal{L} . In analogy with Eq. (3.16), this loss function has the form

$$\begin{aligned} \mathcal{L} &= \tilde{\alpha}_1 \mathcal{L}_{\Delta} + \tilde{\alpha}_2 \mathcal{L}_{\text{tr.}} \\ \mathcal{L}_{\Delta}[\sigma] &= \left\| \Delta_{\mathcal{L}} \sigma - \rho_{\sigma} \right\|_1 \\ \mathcal{L}_{\text{tr.}}[\sigma] &= \sum_{s \neq t} \left\| \sigma_s - T_{(s,t)} \sigma_t \right\|_1, \end{aligned} \quad (3.21)$$

where $\tilde{\alpha}_1, \tilde{\alpha}_2$ are weights, σ_s, σ_t are the versions of σ computed on the patches U_s, U_t , respectively and $T_{(s,t)}$ denotes the transition function between U_t and U_s .

Enforcing the right transformation for σ is more complicated than in the previous cases for ϕ and β due to the presence of the transition functions. These functions can become large or small away from the natural overlap region between two patches. As a result, for any $i \in \{1, \dots, 4\}$ with a line bundle integer $k^i \neq 0$, the transition loss is evaluated at points on a ‘belt’ region \mathcal{B} , defined by $1/2 < |z_i| < 2$, where z_i is the affine coordinate on the i^{th} \mathbb{P}^1 space introduced previously.

The neural network is fully connected with GeLU activation, three layers and a width of 128. Training is carried out for 75 epochs, with batch size 64 and learning rate 0.001. Each of the required seven harmonic forms is computed for $\psi_0 = 2$ and $\psi \in \{0, 0.5, 1, 2, 4, 6\}$. To discuss training and accuracy, we focus on the case $\psi = 1$ which turns out to be typical. The loss over the course of training for this case is shown in Fig. 3a and Fig. 3b. Appropriate measures for the trained network’s performance are defined as

$$M_{\Delta}[\sigma] = \frac{\mathcal{L}_{\Delta}[\sigma]}{\int_X |\rho|}, \quad (3.22)$$

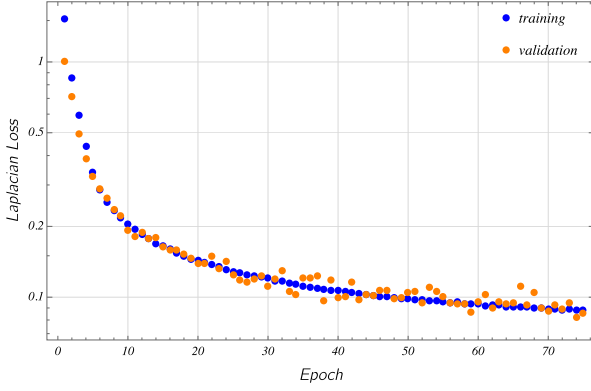
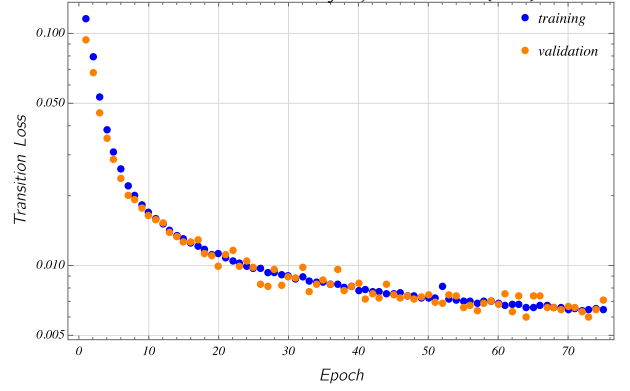

(a) Laplacian loss, \mathcal{L}_Δ , as given in Eq. (3.21).

(b) Transition loss, $\mathcal{L}_{\text{tr.}}$, as given in Eq. (3.21).

Fig. 3. Laplacian and transition losses for the neural network approximating the section σ defined in Eq. (3.20). This example, typical of the cases we consider, solves for the correction to the up-Higgs one-form. We have selected the values $\psi_0 = 2$ and $\psi = 1$.

Table 1

Typical values of the errors, evaluated on the trained neural networks representing σ , for $\psi_0 = 2$ and $\psi = 1$.

| | $H_{2,5}^u$ | Q_5 | U_5 | Q_2^1 | Q_2^2 | U_2^1 | U_2^2 |
|--------------------------|-------------|-------|-------|---------|---------|---------|---------|
| $M_\Delta(\sigma)$ | 0.06 | 0.08 | 0.12 | 0.15 | 0.20 | 0.14 | 0.12 |
| $M_{\text{tr.}}(\sigma)$ | 0.04 | 0.09 | 0.13 | 0.15 | 0.17 | 0.21 | 0.14 |

$$M_{\text{tr.}}[\sigma] = \frac{1}{\mathcal{V}_B} \frac{\mathcal{L}_{\text{tr.}}[\sigma]}{\text{std. dev. of } \sigma}, \quad (3.23)$$

where the integration goes over the validation set. In the evaluation of the transition measure for σ , we only consider the ‘belt’ region with volume \mathcal{V}_B . Both measures are given in Table 1, for each of the seven matter fields.

We emphasise that, whilst this setup is very general, in the present case we are also able to construct neural networks which automatically obey the transition relations. This is done by choosing a set of (non-holomorphic) globally generating sections $\{\sigma_i, i = 1, \dots, N\}$ of the bundle, that is, sections which at each point on the base space span the entire fibre over that point. For a rank one bundle, it is sufficient to choose a set of sections such that at least one is nonzero at each point on the manifold. A general section σ is then written as

$$\sigma(x) = \sum_{i=1}^N f_i(x) \sigma_i, \quad (3.24)$$

where f_i are functions which are represented by the output of a neural network. Crucially these functions, along with the neural networks used for ϕ and β , can be made invariant under projective transformations by making use of architectures analogous to those introduced in Ref. [47]. As a result, $\sigma(x)$ automatically transforms as required between the different patches. Further details of these architectures will be explained in Ref. [71].

4. Yukawa couplings and masses

We now outline the main steps in the numerical calculation. For concreteness, we refer to the details of the model introduced in Section 2, but we note that the procedure is general for heterotic line bundle models.

- (1) At the reference Kähler locus $(t_i) = (1, 1, 1, 1)$, and for the defining polynomial (3.8) with $\psi_0 = 2$ and for each $\psi \in \{0, 0.5, 1, 2, 4, 6\}$, a sample of 300,000 points on the TQ manifold is generated. Then, the Ricci-flat CY metric is computed by machine-learning the function ϕ in Eq. (3.4), with this point sample as a training/validation set.
- (2) For each relevant line bundle \mathcal{L} , the corresponding HYM bundle metric H is computed by machine-learning the function β in Eq. (3.14), with a newly generated training/validation point set and the loss function (3.16). This needs to be done for three line bundles, namely, for \mathcal{L}_2 associated to $Q_2^1, Q_2^2, U_2^1, U_2^2$, for \mathcal{L}_5 , associated to Q_5, U_5 and for $\mathcal{L}_2^* \otimes \mathcal{L}_5^*$, associated to $H_{2,5}^u$.
- (3) For each matter field involved, the corresponding bundle-valued harmonic form ν is computed by starting with the corresponding reference bundle-valued form taken from Refs. [25–27], machine-learning σ in Eq. (3.19), with an additional point sample as training/validation set and loss function (3.21). This needs to be done for seven cases: the three left-handed quarks Q^i , the three right-handed up-quarks U^i and the up-Higgs H^u .
- (4) Using the aforementioned quantities, Eqs. (3.1) and (3.2) are used to compute the holomorphic Yukawa couplings and matter field metrics. This is done by Monte-Carlo integration using the given point sample. These results must be consistent with the

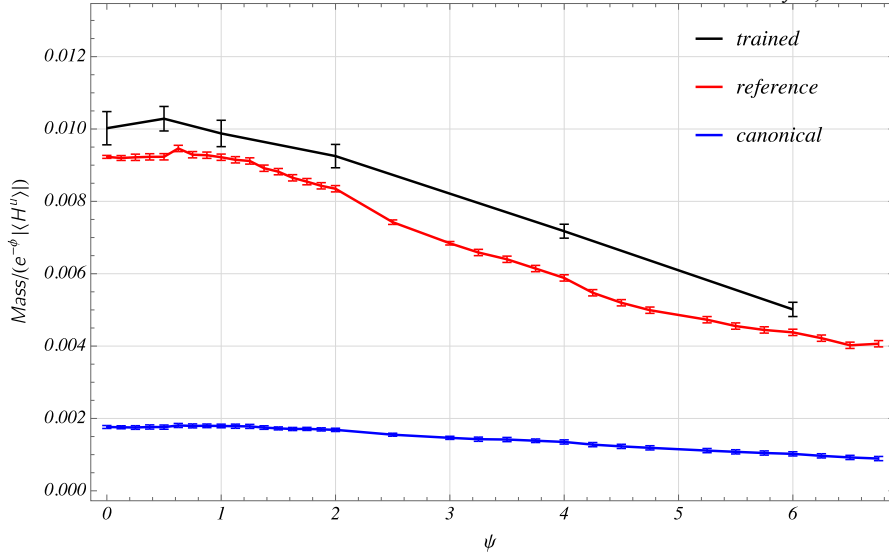


Fig. 4. Plot of the (degenerate) top/charm mass in units of $e^{-\phi} |\langle H^u \rangle|$ for the defining polynomial (3.8) with $\psi_0 = 2$, as a function of the complex structure modulus ψ . The black curve corresponds to the full neural network calculation, whilst the red curve gives the masses computed with the reference metrics and differential forms. The blue curve shows the mass for canonical kinetic terms, obtained by setting $\mathcal{K}^Q = \mathcal{K}^u = \mathbb{1}_2$ and $k^u = k^Q = k = 1$ in Eq. (2.8). Comparison of the blue and black curve demonstrates the importance of including the field normalisations. Error bars are statistical, and average five independent calculations.

structure of Yukawa couplings and matter field metrics as given in Eqs. (2.6) and (2.8), which provides an important check of our calculation. Furthermore, this determines the entries λ_i of the holomorphic Yukawa couplings in Eq. (2.6) and the quantities $\mathcal{K}^Q, \mathcal{K}^u, k^Q, k^u, k$ in Eq. (2.8).

- (5) Finally, inserting these quantities into Eqs. (2.9), (2.10) and (2.11), we determine the physical up Yukawa couplings and up-quark masses.

The above calculation is performed in three modes. Initially, we carry out a quick calculation with the analytic reference quantities, that is, we set ϕ and all β and σ to zero. In this case, no neural networks need to be trained—the above first three steps are trivial—but the integrals still need to be carried out numerically. For comparison, we also calculate the masses which arise if one assumes canonical kinetic terms,⁵ that is, by setting $\mathcal{K}^Q = \mathcal{K}^u = \mathbb{1}_2$ and $k^u = k^Q = k = 1$ in Eq. (2.8). Finally, the results are then compared with the full calculation carried out by following all of the above five steps. In this case, a total of 11 neural networks are trained to obtain the correct quantities ϕ , β and σ .

The final result for the up-quark mass is shown in Fig. 4, with the black curve showing the full calculation, the red curve the calculation based on the reference metrics and forms, and the blue curve the calculation with canonical kinetic terms. Due to the enhanced symmetry of the one-parameter family (3.8), the two non-zero masses are forced to be identical and, within numerical errors, this is confirmed by our results. This is the reason for why Fig. 4 shows results for only a single mass. Some further remarks are in order. First, the physical mass has a significant dependence on complex structure (black curve). Secondly, comparison of the blue and black curves shows that taking into account the field normalisation has a significant effect. Finally, the calculation using the reference quantities (red curve) leads to a reasonable approximation, significantly better than the one based on assuming a canonical Kähler metric (blue curve).

Can the degeneracy between the two non-zero masses be lifted if we move away from the one-parameter family (3.8) of complex structures? To show that this is indeed possible, we consider the new defining polynomial in Eq. (3.9). This point in moduli space was selected by randomly generating 5 polynomials with the $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry and integer coefficients between 0 and 100 and then taking the example with the largest mass ratio. As before, we carry out the calculation in three modes, that is, we perform the full calculation leading to up-quark masses (m_i), the calculation with reference quantities leading to masses ($m_i^{(\text{ref})}$) and the calculation with canonical kinetic terms with masses ($m_i^{(\text{can})}$). The results are

$$\begin{aligned}
(m_1, m_2, m_3) &\approx e^{-\phi} |\langle H^u \rangle| (0, 0.009, 0.016) , \\
(m_1^{(\text{ref})}, m_2^{(\text{ref})}, m_3^{(\text{ref})}) &\approx e^{-\phi} |\langle H^u \rangle| (0, 0.007, 0.013) , \\
(m_1^{(\text{can})}, m_2^{(\text{can})}, m_3^{(\text{can})}) &\approx e^{-\phi} |\langle H^u \rangle| (0, 0.004, 0.008) .
\end{aligned} \tag{4.1}$$

⁵ Note that with this definition ‘canonical kinetic terms’, the physical Yukawa couplings remain independent of the overall Kähler modulus.

This, as expected, lifts the degeneracy between the two non-zero masses. The results lend further evidence to the claim that reference quantities serve as a better approximation than canonical kinetic terms. However, the larger of the two masses is still too small to account for the top mass and the split is not sufficient to explain the top to charm mass ratio. It is reasonable to expect that a more systematic exploration of the 20-parameter complex structure moduli space leads to masses which are more phenomenologically acceptable. This will be investigated in future work [71].

5. Conclusion

In this paper, we have presented the first calculation of physical Yukawa couplings in a heterotic string model compactified on a CY manifold with non-standard embedding. In particular, we have calculated the physical up-quark Yukawa couplings and resulting masses. There are two main parts to our methodology. First, we have introduced analytic ‘reference’ expressions for all relevant quantities which are contained in the right cohomology classes. Specifically, we have used the restricted ambient Fubini-Study metric as the reference CY metric, the restriction of the standard line bundle metrics on projective spaces as bundle reference metrics and certain restricted ambient bundle-valued forms as reference forms to represent the matters fields. In a second step, we have added exact terms to these reference quantities. For these we have determined numerical solutions using neural network techniques, in order to obtain the Ricci-flat CY metric, the Hermitian Yang-Mills bundle metrics and the harmonic bundle-valued forms.

It turns out, the presence of additional $U(1)$ symmetries in our model leads to a rank two up-quark Yukawa matrix. The calculation has been carried out for the two-parameter family (3.8) of tetra-quadric CY threefolds, whose additional symmetry forces the two non-zero masses to be equal. Our main result is shown in Fig. 4 which displays the value of the (degenerate) non-zero quark mass as a function of the complex structure parameter ψ in Eq. (3.8). This is computed in three ways: (1) the full numerical calculation (black curve), (2) the semi-analytic calculation using the reference quantities (red curve), and (3) a semi-analytic calculation where the kinetic terms have been set to canonical form ‘by hand’ (blue curve). The full numerical calculation for each point in Fig. 4 involves training 11 neural networks and takes about half a day on a twelve-core CPU. We estimate that the full numerical result is within 10% of the actual values.

While the value of the mass and the degeneracy are clearly phenomenologically unacceptable, a number of interesting features can be observed in Fig. 4. First, the mass has significant complex structure dependence, even for the limited two-parameter family under consideration (black curve). Secondly, the reference calculation (red curve) provides a reasonable approximation, within 25% of the full numerical one (black curve), which is certainly better than the one based on canonical kinetic terms (blue curve). This observation might have important practical consequences for investigating the phenomenology of fermion masses from string theory. For a ‘first-pass’ analysis, it may well be sufficient to calculate with the reference quantities.

As emphasised earlier, the purpose of this paper is to provide proof of concept that a calculation of fermion masses can be carried out, by using machine learning techniques. Obtaining a realistic (up-quark) mass spectrum was not expected (and is indeed not achieved). Nevertheless, it is instructive to discuss the physical implications of our results for the model at hand. The fact that the mass matrix has, at the perturbative level considered here, one vanishing eigenvalue, is not necessarily a problem. Non-perturbative effects might well fill in some of the zero entries. Further problems are the degeneracy of the two non-zero masses and the fact, evident from Fig. 4, that the value is too small to account for the top quark mass.

To explore this further, we have moved away from the one-parameter family (3.8). Instead, we have carried out the calculation for a different point (3.9) in moduli space without any additional symmetry. The results in Eq. (4.1) show that the degeneracy between the two non-zero masses is indeed lifted, as expected. While this is encouraging, the actual values in Eq. (4.1) are still not phenomenologically viable. Whether realistic top and charm quark masses can be obtained somewhere in the full 20-dimensional moduli space remains to be investigated. These and other issues will be further studied in a forthcoming paper [71], which will also present a systematic error analysis and more sophisticated network architectures which obviate the need for transition losses.

The present work opens up many directions for future investigation. The methods presented here apply to a wide range of CY manifolds, including complete intersections in products of projective spaces and hyper-surfaces in toric four-folds [74]. With suitable modifications, they should also be suitable for calculating masses in F-theory [75]. In this paper, we have concentrated on vector bundles given as sum of line bundles. This leads to considerable technical simplifications, but also means that our techniques are directly applicable to the $\mathcal{O}(10^4)$ known line bundle models with the right particle content. Nevertheless, a generalisation to vector bundles with a non-Abelian structure group is feasible and desirable. Developing these techniques will open up the possibility for calculating fermion masses in large classes of string models and search for phenomenologically viable cases. Perhaps most crucially, this effort must ultimately be combined with moduli stabilisation [76–82].

Dedication

This paper is dedicated to our colleague Graham G. Ross who passed away in October 2021. Graham’s many important contributions to high energy physics include his pioneering work on string phenomenology and fermion masses from string theory [33–36]. Some of the authors have greatly benefited from discussing these and related issues with Graham over the years, and these discussions have, in part, motivated and guided our work.

Declaration of competing interest

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

Acknowledgements

AC was supported by a Stephen Hawking Fellowship (EPSRC grant EP/T016280/1), and by a Royal Society Dorothy Hodgkin Fellowship grant DHF/R1/231142. CSFT is supported by the Gould-Watson Scholarship. TRH is supported by an STFC studentship. AL acknowledges support by the STFC consolidated grant ST/X000761/1. BO is supported in part by both the research grant Department of Energy (DOE) No. DESC0007901 and School of Arts and Sciences (SAS) Account 020-0188-2-010202-6603-0338.

The authors would like to thank Fabian Ruehle for support with cymetric, Jonathan Patterson for assistance with the Oxford Theoretical Physics computing cluster Hydra, and Russell Jones for general technical support. BO would like to acknowledge the hospitality of the CCPP at New York University, where his contribution to this work was carried out.

Data availability

Code will be made available with the publication of subsequent papers.

References

- [1] P. Candelas, G.T. Horowitz, A. Strominger, E. Witten, Vacuum configurations for superstrings, *Nucl. Phys. B* 258 (1985) 46–74.
- [2] V. Braun, Y.-H. He, B.A. Ovrut, T. Pantev, The exact MSSM spectrum from string theory, *J. High Energy Phys.* 05 (2006) 043, arXiv:hep-th/0512177.
- [3] R. Blumenhagen, S. Moster, T. Weigand, Heterotic GUT and standard model vacua from simply connected Calabi-Yau manifolds, *Nucl. Phys. B* 751 (2006) 186–221, arXiv:hep-th/0603015.
- [4] R. Blumenhagen, S. Moster, R. Reinbacher, T. Weigand, Massless spectra of three generation U(N) heterotic string vacua, *J. High Energy Phys.* 05 (2007) 041, arXiv:hep-th/0612039.
- [5] V. Braun, P. Candelas, R. Davies, A three-generation Calabi-Yau manifold with small Hodge numbers, *Fortschr. Phys.* 58 (2010) 467–502, arXiv:0910.5464 [hep-th].
- [6] V. Braun, P. Candelas, R. Davies, R. Donagi, The MSSM spectrum from (0, 2)-deformations of the heterotic standard embedding, *J. High Energy Phys.* 05 (2012) 127, arXiv:1112.1097 [hep-th].
- [7] L.B. Anderson, J. Gray, A. Lukas, E. Palti, Two hundred heterotic standard models on smooth Calabi-Yau threefolds, *Phys. Rev. D* 84 (2011) 106005, arXiv:1106.4804 [hep-th].
- [8] L.B. Anderson, J. Gray, A. Lukas, E. Palti, Heterotic line bundle standard models, *J. High Energy Phys.* 06 (2012) 113, arXiv:1202.1757 [hep-th].
- [9] L.B. Anderson, A. Constantin, J. Gray, A. Lukas, E. Palti, A comprehensive scan for heterotic SU(5) GUT models, *J. High Energy Phys.* 01 (2014) 047, arXiv:1307.4787 [hep-th].
- [10] E.I. Buchbinder, A. Constantin, A. Lukas, The moduli space of heterotic line bundle models: a case study for the tetra-quadric, *J. High Energy Phys.* 03 (2014) 025, arXiv:1311.1941 [hep-th].
- [11] E.I. Buchbinder, A. Constantin, A. Lukas, A heterotic standard model with $B - L$ symmetry and a stable proton, *J. High Energy Phys.* 06 (2014) 100, arXiv:1404.2767 [hep-th].
- [12] E.I. Buchbinder, A. Constantin, A. Lukas, Non-generic couplings in supersymmetric standard models, *Phys. Lett. B* 748 (2015) 251–254, arXiv:1409.2412 [hep-th].
- [13] A. Constantin, Y.-H. He, A. Lukas, Counting string theory standard models, *Phys. Lett. B* 792 (2019) 258–262, arXiv:1810.00444 [hep-th].
- [14] A. Constantin, T.R. Harvey, A. Lukas, Heterotic string model building with monad bundles and reinforcement learning, *Fortsch. Phys.* 70 (o), 2100186, 2-3, arXiv:2108.07316 [hep-th], 2022.
- [15] S. Abel, A. Constantin, T.R. Harvey, A. Lukas, Evolving heterotic gauge backgrounds: genetic algorithms versus reinforcement learning, *Fortschr. Phys.* 70 (5) (2022) 2200034, arXiv:2110.14029 [hep-th].
- [16] S.A. Abel, A. Constantin, T.R. Harvey, A. Lukas, L.A. Nutricati, Decoding nature with nature's tools: heterotic line bundle models of particle physics with genetic algorithms and quantum annealing, arXiv:2306.03147 [hep-th].
- [17] M. Ambroso, B.A. Ovrut, The mass spectra, hierarchy and cosmology of $B - L$ MSSM heterotic compactifications, *Int. J. Mod. Phys. A* 26 (2011) 1569–1627, arXiv:1005.5392 [hep-th].
- [18] B.A. Ovrut, A. Purves, S. Spinner, The minimal SUSY $B - L$ model: from the unification scale to the LHC, *J. High Energy Phys.* 06 (2015) 182, arXiv:1503.01473 [hep-ph].
- [19] L.B. Anderson, J. Gray, M. Larfors, M. Magill, R. Schneider, Generalized vanishing theorems for Yukawa couplings in heterotic compactifications, *J. High Energy Phys.* 2021 (2021) 85, DOI:10.1007/JHEP05(2021)085.
- [20] M. Larfors, R. Schneider, Explore and Exploit with Heterotic Line Bundle Models, *Fortschritte der Physik* 68 (2020) 2000034, DOI:10.1002/prop.202000034.
- [21] H. Otsuka, K. Takemoto, Deep learning and k-means clustering in heterotic string vacua with line bundles, *J. High Energy Phys.* 2020 (2020) 47, DOI:10.1007/JHEP05(2020)047.
- [22] J. Gray, J. Wang, Jumping spectra and vanishing couplings in heterotic Line Bundle Standard Models, *J. High Energy Phys.* 2019 (2019) 73, DOI:10.1007/JHEP11(2019)073.
- [23] V. Braun, Y.-H. He, B.A. Ovrut, Yukawa couplings in heterotic standard models, *J. High Energy Phys.* 04 (2006) 019, arXiv:hep-th/0601204.
- [24] L.B. Anderson, J. Gray, B. Ovrut, Yukawa textures from heterotic stability walls, *J. High Energy Phys.* 05 (2010) 086, arXiv:1001.2317 [hep-th].
- [25] S. Blesneag, E.I. Buchbinder, P. Candelas, A. Lukas, Holomorphic Yukawa couplings in heterotic string theory, *J. High Energy Phys.* 01 (2016) 152, arXiv:1512.05322 [hep-th].
- [26] S. Blesneag, E.I. Buchbinder, A. Lukas, Holomorphic Yukawa couplings for complete intersection Calabi-Yau manifolds, *J. High Energy Phys.* 01 (2017) 119, arXiv:1607.03461 [hep-th].
- [27] S. Blesneag, Holomorphic Yukawa Couplings in Heterotic String Theory, PhD thesis, Oxford U, 2021, arXiv:2204.01165 [hep-th].
- [28] A. Strominger, Yukawa couplings in superstring compactification, *Phys. Rev. Lett.* 55 (1985) 2547.
- [29] P. Candelas, Yukawa couplings between (2, 1) forms, *Nucl. Phys. B* 298 (1988) 458.
- [30] L.J. Dixon, V. Kaplunovsky, J. Louis, On effective field theories describing (2, 2) vacua of the heterotic string, *Nucl. Phys. B* 329 (1990) 27–82.
- [31] P. Candelas, X. de la Ossa, Moduli space of Calabi-Yau manifolds, *Nucl. Phys. B* 355 (1991) 455–481.
- [32] K. Ishiguro, T. Kobayashi, H. Otsuka, Hierarchical structure of physical Yukawa couplings from matter field Kähler metric, *J. High Energy Phys.* 07 (2021) 064, arXiv:2103.10240 [hep-th].
- [33] B.R. Greene, K.H. Kirklin, P.J. Miron, G.G. Ross, A superstring inspired standard model, *Phys. Lett. B* 180 (1986) 69.
- [34] B.R. Greene, K.H. Kirklin, P.J. Miron, G.G. Ross, A three generation superstring model. 1. Compactification and discrete symmetries, *Nucl. Phys. B* 278 (1986) 667–693.

- [35] B.R. Greene, K.H. Kirklin, P.J. Miron, G.G. Ross, A three generation superstring model. 2. Symmetry breaking and the low-energy theory, *Nucl. Phys. B* 292 (1987) 606–652.
- [36] B.R. Greene, K.H. Kirklin, P.J. Miron, G.G. Ross, $27^{*}3$ Yukawa couplings for a three generation superstring model, *Phys. Lett. B* 192 (1987) 111–118.
- [37] G. Butbaia, D. Mayorga Peña, J. Tan, P. Berglund, T. Hübsch, V. Jejjala, C. Mishra, Physical Yukawa couplings in heterotic string compactifications, arXiv:2401.15078 [hep-th].
- [38] M. Larfors, A. Lukas, F. Ruehle, R. Schneider, Learning size and shape of Calabi-Yau spaces, arXiv:2111.01436 [hep-th].
- [39] M. Larfors, A. Lukas, F. Ruehle, R. Schneider, Numerical metrics for complete intersection and Kreuzer–Skarke Calabi–Yau manifolds, *Mach. Learn.: Sci. Technol.* 3 (3) (2022) 035014, arXiv:2205.13408 [hep-th].
- [40] S.K. Donaldson, Some numerical results in complex differential geometry, arXiv:math.DG/0512625.
- [41] V. Braun, T. Brelidze, M.R. Douglas, B.A. Ovrut, Calabi-Yau metrics for quotients and complete intersections, *J. High Energy Phys.* 05 (2008) 080, arXiv:0712.3563 [hep-th].
- [42] M.R. Douglas, R.L. Karp, S. Lukic, R. Reinbacher, Numerical Calabi-Yau metrics, *J. Math. Phys.* 49 (2008) 032302, arXiv:hep-th/0612075.
- [43] M. Headrick, A. Nassar, Energy functionals for Calabi-Yau metrics, *Adv. Theor. Math. Phys.* 17 (5) (2013) 867–902, arXiv:0908.2635 [hep-th].
- [44] A. Ashmore, Y.-H. He, B.A. Ovrut, Machine learning Calabi–Yau metrics, *Fortschr. Phys.* 68 (9) (2020) 2000068, arXiv:1910.08605 [hep-th].
- [45] L.B. Anderson, M. Gerdes, J. Gray, S. Krippendorff, N. Raghuram, F. Ruehle, Moduli-dependent Calabi-Yau and SU(3)-structure metrics from machine learning, *J. High Energy Phys.* 05 (2021) 013, arXiv:2012.04656 [hep-th].
- [46] V. Jejjala, D.K. Mayorga Pena, C. Mishra, Neural network approximations for Calabi-Yau metrics, *J. High Energy Phys.* 08 (2022) 105, arXiv:2012.15821 [hep-th].
- [47] M.R. Douglas, S. Lakshminarasimhan, Y. Qi, Numerical Calabi-Yau metrics from holomorphic networks, arXiv:2012.04797 [hep-th].
- [48] A. Ashmore, L. Calmon, Y.-H. He, B.A. Ovrut, Calabi-Yau metrics, energy functionals and machine-learning, *Int. J. Data Sci. Math. Sci.* 1 (1) (2023) 49–61, arXiv:2112.10872 [hep-th].
- [49] M. Gerdes, S. Krippendorff, CYJAX: a package for Calabi-Yau metrics with JAX, *Mach. Learn.: Sci. Technol.* 4 (2) (2023) 025031, arXiv:2211.12520 [hep-th].
- [50] X. Wang, Canonical metrics on stable vector bundles, *Commun. Anal. Geom.* 13 (2) (2005) 253–385.
- [51] M.R. Douglas, R.L. Karp, S. Lukic, R. Reinbacher, Numerical solution to the hermitian Yang-Mills equation on the Fermat quintic, *J. High Energy Phys.* 12 (2007) 083, arXiv:hep-th/0606261.
- [52] L.B. Anderson, V. Braun, B.A. Ovrut, Numerical Hermitian Yang-Mills connections and Kahler cone substructure, *J. High Energy Phys.* 01 (2012) 014, arXiv:1103.3041 [hep-th].
- [53] L.B. Anderson, V. Braun, R.L. Karp, B.A. Ovrut, Numerical Hermitian Yang-Mills connections and vector bundle stability in heterotic theories, *J. High Energy Phys.* 06 (2010) 107, arXiv:1004.4399 [hep-th].
- [54] A. Ashmore, R. Deen, Y.-H. He, B.A. Ovrut, Machine learning line bundle connections, *Phys. Lett. B* 827 (2022) 136972, arXiv:2110.12483 [hep-th].
- [55] A. Ashmore, Y.-H. He, E. Heyes, B.A. Ovrut, Numerical spectra of the Laplacian for line bundles on Calabi-Yau hypersurfaces, *J. High Energy Phys.* 07 (2023) 164, arXiv:2305.08901 [hep-th].
- [56] V. Braun, T. Brelidze, M.R. Douglas, B.A. Ovrut, Eigenvalues and eigenfunctions of the scalar Laplace operator on Calabi-Yau manifolds, *J. High Energy Phys.* 07 (2008) 120, arXiv:0805.3689 [hep-th].
- [57] A. Ashmore, Eigenvalues and eigenforms on Calabi–Yau threefolds, *J. Geom. Phys.* 195 (2024) 105028, arXiv:2011.13929 [hep-th].
- [58] A. Ashmore, F. Ruehle, Moduli-dependent KK towers and the swampland distance conjecture on the quintic Calabi-Yau manifold, *Phys. Rev. D* 103 (10) (2021) 106028, arXiv:2103.07472 [hep-th].
- [59] D. Cremades, L.E. Ibanez, F. Marchesano, Computing Yukawa couplings from magnetized extra dimensions, *J. High Energy Phys.* 05 (2004) 079, arXiv:hep-th/0404229.
- [60] S. Krippendorff, M.J. Dolan, A. Maharana, F. Quevedo, D-branes at toric singularities: model building, Yukawa couplings and flavour physics, *J. High Energy Phys.* 06 (2010) 092, arXiv:1002.1790 [hep-th].
- [61] A. Font, L.E. Ibanez, Yukawa structure from U(1) fluxes in F-theory grand unification, *J. High Energy Phys.* 02 (2009) 016, arXiv:0811.2157 [hep-th].
- [62] J.J. Heckman, C. Vafa, Flavor hierarchy from F-theory, *Nucl. Phys. B* 837 (2010) 137–151, arXiv:0811.2417 [hep-th].
- [63] H. Hayashi, T. Kawano, R. Tatar, T. Watari, Codimension-3 singularities and Yukawa couplings in F-theory, *Nucl. Phys. B* 823 (2009) 47–115, arXiv:0901.4941 [hep-th].
- [64] S. Cecotti, M.C.N. Cheng, J.J. Heckman, C. Vafa, Yukawa couplings in F-theory and non-commutative geometry, arXiv:0910.0477 [hep-th].
- [65] A. Font, L.E. Ibanez, Matter wave functions and Yukawa couplings in F-theory grand unification, *J. High Energy Phys.* 09 (2009) 036, arXiv:0907.4895 [hep-th].
- [66] J.P. Conlon, E. Palti, Aspects of flavour and supersymmetry in F-theory GUTs, *J. High Energy Phys.* 01 (2010) 029, arXiv:0910.2413 [hep-th].
- [67] H. Hayashi, T. Kawano, Y. Tsuchiya, T. Watari, Flavor structure in F-theory compactifications, *J. High Energy Phys.* 08 (2010) 036, arXiv:0910.2762 [hep-th].
- [68] L. Aparicio, A. Font, L.E. Ibanez, F. Marchesano, Flux and instanton effects in local F-theory models and hierarchical fermion masses, *J. High Energy Phys.* 08 (2011) 152, arXiv:1104.2609 [hep-th].
- [69] E. Palti, Wavefunctions and the point of E_8 in F-theory, *J. High Energy Phys.* 07 (2012) 065, arXiv:1203.4490 [hep-th].
- [70] c. Blesneag, E.I. Buchbinder, A. Constantin, A. Lukas, E. Palti, Matter field Kähler metric in heterotic string theory from localisation, *J. High Energy Phys.* 04 (2018) 139, arXiv:1801.09645 [hep-th].
- [71] A. Constantin, C.S. Fraser-Taliente, T.R. Harvey, A. Lukas, B. Ovrut, in preparation.
- [72] V. Braun, On free quotients of complete intersection Calabi-Yau manifolds, *J. High Energy Phys.* 04 (2011) 005, arXiv:1003.3235 [hep-th].
- [73] D. Hendrycks, K. Gimpel, Gaussian error linear units (GELUs), arXiv:1606.08415 [cs.LG], <https://doi.org/10.48550/arXiv.1606.08415>, 2016.
- [74] Y.-H. He, S.-J. Lee, A. Lukas, C. Sun, Heterotic model building: 16 special manifolds, *J. High Energy Phys.* 06 (2014) 077, arXiv:1309.0223 [hep-th].
- [75] A.P. Braun, C.R. Brodie, A. Lukas, Heterotic line bundle models on elliptically fibered Calabi-Yau three-folds, *J. High Energy Phys.* 04 (2018) 087, arXiv:1706.07688 [hep-th].
- [76] L.B. Anderson, J. Gray, A. Lukas, B. Ovrut, Stability walls in heterotic theories, *J. High Energy Phys.* 09 (2009) 026, arXiv:0905.1748 [hep-th].
- [77] L.B. Anderson, J. Gray, A. Lukas, B. Ovrut, Stabilizing the complex structure in heterotic Calabi-Yau vacua, *J. High Energy Phys.* 02 (2011) 088, arXiv:1010.0255 [hep-th].
- [78] L.B. Anderson, J. Gray, A. Lukas, B. Ovrut, Stabilizing all geometric moduli in heterotic Calabi-Yau vacua, *Phys. Rev. D* 83 (2011) 106011, arXiv:1102.0011 [hep-th].
- [79] L.B. Anderson, J. Gray, A. Lukas, B. Ovrut, The Atiyah class and complex structure stabilization in heterotic Calabi-Yau compactifications, *J. High Energy Phys.* 10 (2011) 032, arXiv:1107.5076 [hep-th].
- [80] C. Deffayet, B.A. Ovrut, P.J. Steinhardt, Moduli axions, stabilizing moduli and the large field swampland conjecture in heterotic M-theory, arXiv:2312.04656 [hep-th].
- [81] X. de la Ossa, E.E. Svanes, Holomorphic bundles and the moduli space of $N=1$ supersymmetric heterotic compactifications, *J. High Energy Phys.* 10 (2014) 123, [https://doi.org/10.1007/JHEP10\(2014\)123](https://doi.org/10.1007/JHEP10(2014)123).
- [82] L.B. Anderson, J. Gray, E. Sharpe, Algebroids, heterotic moduli spaces and the Strominger system, *J. High Energy Phys.* 07 (2014) 037, [https://doi.org/10.1007/JHEP07\(2014\)037](https://doi.org/10.1007/JHEP07(2014)037).