Reproducing Standard Model fermion masses and mixing in string theory: A heterotic line bundle study

Andrei Constantin,[1,2,*] Lucas T.-Y. Leung,[2,†] Andre Lukas,[2,‡] and Luca A. Nutricati[2,§]

[1]*School of Mathematics, University of Birmingham, Watson Building, Edgbaston, Birmingham B15 2TT, United Kingdom*

[2]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*

(Received 10 September 2025; accepted 15 January 2026; published 9 February 2026)

Deriving the Yukawa couplings and the resulting fermion masses and mixing angles of the Standard Model (SM) from a more fundamental theory remains one of the central outstanding problems in theoretical high-energy physics. It has long been recognized that string theory provides a framework within which this question can, at least in principle, be addressed. While substantial progress has been made in studying flavor physics in string compactifications over the past few decades, a concrete string construction that reproduces the full set of observed SM flavor parameters remains unknown. Here, we take a significant step in this direction by identifying two explicit $E_8 \times E_8$ heterotic string models compactified on a Calabi-Yau threefold with Abelian, holomorphic, and polystable vector bundles with minimal supersymmetric (MS) SM spectrum. Subject to reasonable assumptions about the moduli, we show that these models reproduce the correct values of the quark and charged lepton masses, as well as the quark mixing parameters, at some point in their moduli spaces. The resulting four-dimensional theories are $\mathcal{N} = 1$ supersymmetric, contain no exotic fields, and realize a μ-term suppressed to the electroweak scale. While the issues of moduli stabilization and supersymmetry breaking are not addressed here; our main result constitutes a proof of principle: There exist choices of topology and moduli within heterotic string compactifications which allow for an MSSM spectrum with the correct flavor parameters.

DOI: 10.1103/fbdz-t73z

I. INTRODUCTION

It has been recognized for nearly four decades that heterotic string compactifications on Calabi-Yau (CY) threefolds provide a natural framework for embedding realistic particle physics into string theory [1]. Beyond unification, this approach offers a significant advantage over conventional effective field theory: In string theory, all low-energy quantities, including the free parameters of the Standard Model (SM), are, in principle, calculable from the geometry and topology of the compactification. This opens the door to deriving the SM from first principles. Before such calculations can be performed, however, one must first construct compactifications that exactly reproduce the SM gauge group and particle spectrum, with no additional exotic fields or unwanted low-energy interactions. Achieving this has proven to be highly nontrivial. Generic string compactifications produce large gauge groups and complex matter spectra. The problem is exacerbated by the vast landscape of consistent four-dimensional string theory solutions and the technical challenges of identifying explicit compactifications that satisfy key requirements such as anomaly cancellation, supersymmetry, and spectrum constraints.

While decades of progress have brought us closer, an explicit construction that reproduces the exact SM spectrum with realistic flavor data has remained elusive. Recently, however, advances in mathematical methods, computational tools, and search algorithms have enabled the systematic exploration of large classes of compactifications leading to many models with the correct minimal supersymmetric (MS) SM spectrum. Among the most promising and tractable avenues are compactifications of the $E_8 \times E_8$ heterotic string on CY threefolds with vector bundles built as direct sums of line bundles [2–5]. This framework gives rise to a large and well-controlled class of models amenable to large-scale scanning [6–13] and explicit computational analysis. Many such constructions have been shown to yield the exact SM gauge group, three

*Contact author: andrei.constantin@physics.ox.ac.uk
†Contact author: lucas.leung@physics.ox.ac.uk
‡Contact author: andre.lukas@physics.ox.ac.uk
§Contact author: luca.nutricati@physics.ox.ac.uk

chiral families, and no exotic matter. Crucially, progress in this area has been driven by the development of mathematical and algorithmic tools that allow for the efficient computation of particle spectra via sheaf cohomology [14–25] and the computation of physical Yukawa couplings in terms of the compactification data [26–42].

A key structural feature of heterotic line bundle models is the presence of additional $U(1)$ gauge symmetries originating from the internal gauge bundle. While the corresponding gauge bosons typically acquire large Stückelberg masses and decouple from the low-energy spectrum, the symmetries themselves persist as effective global selection rules in the four-dimensional theory. Both SM fields and SM gauge-singlet moduli carry nontrivial charges under these $U(1)$ symmetries, which strongly constrain the form of allowed operators in the effective theory. When the moduli acquire vacuum expectation values (VEVs), they induce effective SM couplings. This mechanism, reminiscent of Froggatt-Nielsen models [43], can generate hierarchical structures in the couplings in a way that arises naturally from the geometry of the compactification. A systematic study of how such $U(1)$ symmetries constrain the operator structure of effective field theories with features characteristic of heterotic line bundle models was recently discussed in [44]. The present work builds on that foundation but is taking a top-down rather than a bottom-up point of view. Our goal here is to identify explicit $E_8 \times E_8$ heterotic line bundle models that successfully reproduce the observed flavor structure of the SM.

The effective couplings in heterotic line bundle models, such as the Yukawa matrices, are shaped by two key ingredients. First, the $U(1)$ selection rules inherited from the compactification restrict the structure of allowed operators, often requiring the insertion of additional moduli fields to ensure symmetry invariance. When these moduli acquire VEVs, they generate suppression factors in the effective couplings, as the relevant VEVs, measured in units of the compactification scale [typically close to the grand unified theory (GUT) scale], are subunitary. Second, each operator comes with an order-one coefficient that depends on the detailed geometry of the compactification, for example, through overlap integrals of internal wave functions. While these coefficients depend nontrivially on the (complex structure and Kähler) moduli, numerical studies indicate they typically lie within an order of magnitude of unity across the moduli space [41].

Assuming the unfixed coefficients vary within a few orders of magnitude of 1, and that the moduli acquire generic, nonhierarchical VEVs, we identify two concrete line bundle models that reproduce the observed quark and charged lepton masses as well as the Cabibbo–Kobayashi–Maskawa (CKM) matrix. MSSM. While we do not address how the moduli are stabilized at the required values, or the mechanism of supersymmetry breaking, this approach is nonetheless well motivated: In many string scenarios, the leading-order flavor structure is fixed in the supersymmetric limit, while supersymmetry breaking primarily affects soft terms and the scalar potential. As long as the moduli are stabilized near the VEVs identified in our analysis, the fermion masses and mixing angles remain essentially unchanged. Supersymmetry also imposes strong structural constraints on the effective theory: Holomorphic quantities like superpotential Yukawa couplings are determined at tree level and protected from large quantum corrections, while the Kähler potential, although more intricate, is still computable in terms of the underlying geometry [41]. These features make the computation of physical Yukawa couplings in terms of moduli-dependent suppression factors and order-one geometric coefficients, feasible.

In addition to reproducing the observed flavor structure, the models presented below allow for a μ-term of the correct electroweak scale. The μ problem (that is, the question of why a supersymmetric Higgs mass parameter should lie far below the Planck or compactification scale) is resolved here by the same mechanism that controls the Yukawa hierarchies. The bare μ-term is forbidden by the compactification topology, but an effective term is generated through higher-dimensional operators involving moduli fields. Once these moduli acquire appropriately suppressed VEVs, an effective μ-term of the right magnitude is induced in our models.

II. HETEROTIC LINE BUNDLE MODELS

A. Geometry

The systematic study of heterotic line bundle models was initiated in Refs. [6–9]. These models are based on compactifying the $E_8 \times E_8$ heterotic string on smooth CY threefold quotients X/Γ, where X is a simply connected CY threefold with Picard number $h = h^{1,1}(X)$ and a freely acting finite symmetry group Γ. For the models presented here, the covering space X is a complete intersection of hypersurfaces in a product of projective spaces [a Complete Intersection Calabi-Yau (CICY)]. Gauge symmetry breaking from the 10D $E_8 \times E_8$ theory to the 4D SM is achieved by specifying a nontrivial gauge background on the internal CY manifold in one of the E_8 sectors referred to as the observable sector. This internal gauge field background corresponds to a holomorphic vector bundle V on X. We take V to be a direct sum of five line bundles over X,

$$V = \bigoplus_{a=1}^{5} \mathcal{L}_a = \bigoplus_{a=1}^{5} \mathcal{O}_X(\mathbf{k}_a), \tag{2.1}$$

where the h-dimensional integer vectors $\mathbf{k}_a = (k_a^i)$ specify the first Chern classes $c_1(\mathcal{L}_a)$ and label the line bundles $\mathcal{L}_a$. The line bundle sum V can often be chosen so that it descends to a line bundle sum $\hat{V} = \oplus_{a=1}^{5} \hat{\mathcal{L}}_a$ on X/Γ, and this is going to be the case for our examples. The bundle $\hat{V}$ is also required to be polystable somewhere inside the Kähler cone of X/Γ and must satisfy the topological conditions $c_2(\hat{V}) \leq c_2(X/\Gamma)$. The latter condition ensures anomaly cancellation can be satisfied via the inclusion of five-branes wrapping holomorphic curves or a bundle in the other, hidden E_8 sector. To embed the structure group of V into (the observable) E_8, we impose $c_1(V) = \sum_a c_1(\mathcal{L}_a) = 0$ or, equivalently, $\sum_a \mathbf{k}_a = 0$. Provided this condition is satisfied (and the line bundles are otherwise sufficiently generic), the structure group of V is $\mathcal{G} = S(U(1)^5) \cong U(1)^4$. The resulting low-energy gauge group is $SU(5) \times \mathcal{G}$, the commutant of $\mathcal{G}$ within E_8. The $SU(5)$ factor can be interpreted as a GUT gauge group but it is further broken (and indeed never realized at the 4D field theory level) to the SM gauge group by a discrete Wilson line associated with the nontrivial fundamental group of the quotient manifold X/Γ. Hence, the final low-energy symmetry group (in the observable sector) is $G_{\mathrm{SM}} \times \mathcal{G}$, where $G_{\mathrm{SM}} = SU(3) \times SU(2) \times U(1)$ is the SM gauge group, and the additional group $\mathcal{G}$ emerges as a flavor symmetry.

B. Spectrum

Although the physical 4D gauge group is precisely the SM gauge group G_{SM}, the low-energy theory has features which mimic those of a conventional field theory GUT. This happened because $\mathcal{G}$ commutes with the underlying $SU(5)$ GUT symmetry and, as a result, all SM multiplets which originate from the same $SU(5)$ GUT multiplet must have the same $\mathcal{G}$ charge. This gives rise to a crude form of Yukawa unification: The allowed singlet insertions needed to form effective couplings are identical for, say, the d-quark and the charged lepton Yukawa couplings. However, the coefficients multiplying these operators, which arise from overlap integrals of internal wave functions, are typically different, and hence, Yukawa unification is generically broken at the numerical level [45], in contrast to conventional GUTs.

It is important to note, however, that this is not the only possible situation in heterotic line bundle models. In certain constructions, such as the one discussed in Ref. [46] [which realizes a KSVZ (Kim, Shifman, Vainshtein, and Zakharov) QCD axion with a decay constant in the phenomenologically allowed range], the MSSM spectrum at the Abelian locus includes exotic vectorlike states, which can be lifted by giving suitable VEVs to $U(1)$-charged singlet fields. In such scenarios, once the moduli are displaced away from the Abelian locus, the SM multiplets are no longer required to assemble into $SU(5)$ representations with a common $\mathcal{G}$ charge. This may lead to greater flexibility in the resulting effective low-energy Yukawa textures. A systematic exploration of such models lies beyond the scope of this work and is left for future investigation.

While the group $\mathcal{G}$ is of rank four, charges are most naturally represented as five-component integer vectors $\mathbf{q} \in \mathbb{Z}^5$ subject to the equivalence relation

$$\mathbf{q} \sim \mathbf{q}' \quad \Leftrightarrow \quad \mathbf{q} - \mathbf{q}' \in \mathbb{Z}(1,1,1,1,1),$$

TABLE I. The field content and multiplet charges of rank-five heterotic line bundle models, where $\mathbf{e}_1, \ldots, \mathbf{e}_5$ are the standard five-dimensional unit vectors. The precise charge assignments depend on the topology of the underlying compactification.

Field	SM rep	Name	SU(5)	$\mathcal{G}$ charge pattern
Q	$(\mathbf{3},\mathbf{2})_1$	LH quark	10	$\mathbf{e}_a$
u	$(\bar{\mathbf{3}},\mathbf{1})_{-4}$	RH u-quark		
e	$(\mathbf{1},\mathbf{1})_6$	RH electron		
d	$(\bar{\mathbf{3}},\mathbf{1})_2$	RH d-quark	$\bar{\mathbf{5}}$	$\mathbf{e}_a + \mathbf{e}_b$
L	$(\mathbf{1},\mathbf{2})_{-3}$	LH lepton		
H^d	$(\mathbf{1},\mathbf{2})_{-3}$	Down-Higgs boson	$\bar{\mathbf{5}}^{H^d}$	$\mathbf{e}_a + \mathbf{e}_b$
H^u	$(\mathbf{1},\mathbf{2})_3$	Up-Higgs boson	$\mathbf{5}^{H^u}$	$-\mathbf{e}_a - \mathbf{e}_b$
ϕ	$(\mathbf{1},\mathbf{1})_0$	Perturbative moduli	1	$\mathbf{e}_a - \mathbf{e}_b$
Φ^i	$(\mathbf{1},\mathbf{1})_0$	Nonperturbative moduli	1	$(k_1^i, \ldots, k_5^i)$

which reflects the trace-free condition for $\mathcal{G} = S(U(1)^5)$. The charge patterns of all relevant multiplets are summarized in Table I. Let us briefly explain their origin.

Suppose a quark doublet Q arises as a zero mode of the Dirac operator on the quotient manifold X/Γ valued in one of the five line bundles, say, $\hat{\mathcal{L}}_a$, where $a \in 1, \ldots, 5$. Then Q carries $\mathcal{G}$ charge $\mathbf{e}_a$, with $(\mathbf{e}_a)_{a=1}^5$ denoting the standard five-dimensional unit vectors. In any given model, the three quark doublets may all descend from the same line bundle—leading to identical charges—or they may originate from different line bundles, resulting in distinct charges across generations. Similar considerations apply to right-handed up-quarks u and electrons e. In contrast, the right-handed down-quarks d and lepton doublets L arise as zero modes associated with tensor products $\hat{\mathcal{L}}_a \otimes \hat{\mathcal{L}}_b$, and hence carry charges $\mathbf{e}_a + \mathbf{e}_b$. Again, the specific charge pattern across families is determined by where each field originates in the line bundle sum. As a general rule, in order to explain mass hierarchies, we require $\mathcal{G}$ to be a flavor symmetry; that is, not all families should have the same $\mathcal{G}$ charge. From the above discussion, this implies a topological constraint on the underlying string models: (Q, u, e) families should be associated with at least two different line bundles; that is, we need $h^1(X/\Gamma, \hat{\mathcal{L}}_a) > 0$ for more than one line bundle $\hat{\mathcal{L}}_a$.

The final two rows of Table I refer to moduli fields. The perturbative moduli ϕ arise as zero modes of $\hat{\mathcal{L}}_a \otimes \hat{\mathcal{L}}_b^*$ and carry charges $\mathbf{e}_a - \mathbf{e}_b$. A given compactification typically yields a spectrum of such singlets, which parametrize deformations of the vector bundle. When all ϕ VEVs vanish, the model sits precisely at the line bundle locus, while nonzero ϕ VEVs correspond to moving away from this locus toward non-Abelian bundle configurations (see, e.g., [47,48] for a detailed case study).

Finally, consider the Kähler moduli T^i, $i = 1, \ldots, h$, which transform under $\mathcal{G}$ as $T^i \mapsto T^i + i\epsilon^a k^i_a$, where ϵ^a are the $\mathcal{G}$ transformation parameters, and k^i_a are the integers specifying the line bundles in Eq. (2.1). The nonperturbative moduli $\Phi^i = \exp(-T^i)$ then carry $\mathcal{G}$ charges $(k^i_1, \ldots, k^i_5)$, as indicated in Table I.

C. Moduli VEVs

SM operators are dressed with singlet insertions involving ϕ and Φ fields—representing perturbative and nonperturbative effects, respectively—in a manner controlled by $\mathcal{G}$ invariance. This mechanism provides a string-theoretic realization of Froggatt-Nielsen models. Once these singlets acquire VEVs, effective couplings, including Yukawa couplings, are generated. The detailed structure of these couplings depends on the $\mathcal{G}$ charges of the matter fields, Higgs fields and moduli, which, as we have seen, are determined by the topology and geometry of the compactification. Relying on this mechanism, we identify two explicit line bundle models that successfully reproduce the observed pattern of fermion masses and mixing angles of the SM. In doing so, we assume that the singlet VEVs and the order-one coefficients multiplying the operators can be chosen within plausible ranges, as detailed in Appendix B.

To be more precise, we reproduce the measured (infrared) values for the SM fermion masses and mixing angles in the following sense. While the VEVs for the perturbative and nonperturbative fields ϕ and Φ can be taken at face value, the order-one coefficients depend in a complex way on the underlying complex structure and Kähler moduli. In our analysis, these coefficients correspond to their infrared values. To connect these to a specific point in moduli space, one would need to (1) run the order-one coefficients up to the compactification scale, taking into account the details of supersymmetry breaking, and (2) determine the map between the geometrical moduli and the order-one coefficients, as can be done using the tools developed in Ref. [41]. The key point is that while this matching is technically involved, it is, in principle, feasible. What we take as order one in the IR will differ from the corresponding UV values by factors of order unity arising from renormalization group evolution and the mapping to the underlying geometry.

D. Couplings

Having outlined how the observed fermion mass and mixing parameters can be reproduced in our framework, we now turn to the structure of the low-energy effective action that underlies these results. The dynamics of heterotic line bundle models at low energies is governed by a four-dimensional $\mathcal{N} = 1$ supergravity theory with superpotential

$$
\begin{aligned}
&W = W_Y + W_R + \ldots,\\
&W_Y = \mu H^u H^d + Y^u_{IJ} H^u Q^I u^J + Y^d_{IJ} H^d Q^I d^J + Y^e_{IJ} H^d e^I L^J,\\
&W_R \supset \{LLe, QLd, udd, LH^u\},
\end{aligned}
$$

where $I, J = 1, 2, 3$ are family indices, Y^u_{IJ}, Y^d_{IJ}, and Y^e_{IJ} denote the effective Yukawa couplings, and μ is an effective supersymmetric mass term. All these couplings are holomorphic functions of the moduli fields. The μ-term and typically many Yukawa couplings are absent at the renormalizable level and arise effectively through insertions of singlet fields ϕ and Φ, whose VEVs spontaneously break $\mathcal{G}$.

In addition to the Abelian symmetry $\mathcal{G}$, we assume an R-parity symmetry that forbids all renormalizable R-parity violating terms in the superpotential W_R. Crucially, we also assume that R-parity remains unbroken throughout moduli space, even when $\mathcal{G}$ is spontaneously broken, ensuring the absence of dangerous effective R-parity violating operators. Geometrically, R-parity can originate from a discrete symmetry of the quotient CY threefold X/Γ. The ellipsis in W represents higher-dimensional operators, including those contributing to proton decay, which are naturally suppressed by the compactification scale.

The Kähler potential takes the standard form in $\mathcal{N} = 1$ supergravity, depending on the Kähler moduli, complex structure moduli, and matter fields. The moduli dependence of the matter field Kähler metric leads to nontrivial wave function normalizations and kinetic mixing, modifying the effective Yukawa couplings upon canonical normalization. However, following the perspective of Ref. [41], we regard these effects as order one and absorb them into the order-one coefficients multiplying the superpotential operators. Similarly, any indirect impact of the gauge kinetic function f, for example, via renormalization group evolution, can be absorbed into these coefficients.

III. RESULTS

A. Geometry

Out of the 202 line bundle models constructed in Ref. [6], we identified two that reproduce the observed quark and charged lepton masses and CKM matrix. Our search strategy is described in Appendix B, and the source code of the search algorithm is available at [49]. The models are based on the CICY threefold:

$$
X_{6770} \in \begin{array}{c} \mathbb{P}^1 \\ \mathbb{P}^1 \\ \mathbb{P}^1 \\ \mathbb{P}^1 \\ \mathbb{P}^1 \end{array} \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 2 & 0 \\ 0 & 2 \end{array} \right]^{5,37},
$$

where the superscripts on the configuration matrix indicate the nontrivial Hodge numbers. The manifold is realized as a complete intersection of two hypersurfaces $\{p_1 = 0\} \cap \{p_2 = 0\} \subseteq (\mathbb{P}^1)^{\times 5}$, where p_1, p_2 are homogeneous polynomials with multidegrees given by the columns of the above matrix. To describe the relevant discrete symmetry [50], we introduce the notation $[x_{i,0} : x_{i,1}]$ for the homogeneous coordinates on the ith $\mathbb{P}^1$ factor. The manifold X_{6770} admits a free action of $\Gamma = \mathbb{Z}_2$ defined by $x_{i,0} \to -x_{i,0}$, $x_{i,1} \to x_{i,1}$, $p_1 \to p_1$, $p_2 \to -p_2$. The resulting quotient $X_{6770}/\mathbb{Z}_2$ has Hodge numbers (5,21).

The line bundle sums for the two models are given by

$$V_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ -2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & -2 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 & 0 \end{pmatrix}, \qquad V_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ -2 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix},$$

where the columns define the first Chern class vectors $\mathbf{k}_a$. In both cases, all line bundles are individually equivariant with respect to the relevant freely acting symmetry. The choice of equivariant structure amounts to a choice of one Γ representation for each line bundle. For our examples, all line bundles are associated with the trivial representation, except for the second line bundle in V_1 and the fifth line bundle in V_2. These choices are correlated with the Wilson line, which is introduced to project out the Higgs triplets while breaking $SU(5)$ to $G_{\rm SM}$. The Wilson line is specified by two one-dimensional Γ representations, W_2 and W_3, with $W_2 \neq W_3$. Specifically, we choose $(W_2, W_3) = (1, 0)$ for both models. Crucially, the four gauge bosons associated with the symmetry $\mathcal{G} = S(U(1)^5)$ acquire masses at the compactification scale via the Green-Schwarz mechanism and do not give rise to low-energy fifth forces.

We now present the low-energy physics associated with the two models. In both cases, the massless spectra consist of three chiral families of quarks and leptons, along with a single vectorlike pair of Higgs doublets and several singlet fields. The $\mathcal{G}$ charges and the number of singlet fields are model dependent. The perturbative parts of the spectra are determined through computations of the cohomologies $H^1(X, V)$, $H^1(X, \wedge^2 V)$, $H^1(X, \wedge^2 V^*)$, and $H^1(X, V \otimes V^*)$ which split into Γ representations. The parts that survive the $SU(5)$ breaking are determined by projecting along the directions indicated by the Wilson line [7]. For the nonperturbative fields, we have checked that the Gopakumar–Vafa (GV) invariants associated with the generators of the Mori cone of X_{6770} are nonvanishing. However, we have not attempted to compute the corresponding Pfaffians, and our implicit assumption is that these are nonvanishing.

1. Model 1 on $X_{6770}/\mathbb{Z}_2$

The low-energy spectrum includes the following fields:

Matter	$(Q,u,e)_{\mathbf{e}_1}, 2(Q,u,e)_{\mathbf{e}_2}$, $(d,L)_{\mathbf{e}_1+\mathbf{e}_2}, (d,L)_{\mathbf{e}_1+\mathbf{e}_5}, (d,L)_{\mathbf{e}_2+\mathbf{e}_5}$
Higgs	$H^u_{-\mathbf{e}_1-\mathbf{e}_2}, H^d_{\mathbf{e}_1+\mathbf{e}_2}$
Moduli	$\phi_{1,2} \coloneqq \phi_{\mathbf{e}_1-\mathbf{e}_2}, 4\phi_{2,1} \coloneqq 4\phi_{\mathbf{e}_2-\mathbf{e}_1}$, $4\phi_{1,3} \coloneqq 4\phi_{\mathbf{e}_1-\mathbf{e}_3}, 3\phi_{5,1} \coloneqq 3\phi_{\mathbf{e}_5-\mathbf{e}_1}$, $2\phi_{2,3} \coloneqq 2\phi_{\mathbf{e}_2-\mathbf{e}_3}, 7\phi_{2,5} \coloneqq 7\phi_{\mathbf{e}_2-\mathbf{e}_5}$, $\Phi_1 \coloneqq \Phi_{(1,0,0,0,-1)}, \Phi_2 \coloneqq \Phi_{(-2,1,1,0,0)}$, $\Phi_3 \coloneqq \Phi_{(0,1,0,-1,0)}, \Phi_4 \coloneqq \Phi_{(0,-2,0,1,1)}$, $\Phi_5 \coloneqq \Phi_{(0,0,-1,1,0)}$.

The first step in assigning VEVs to the moduli fields is to ensure the correct scale for the μ-term. We achieve this by setting

$$\langle\Phi_3\rangle, \langle\phi_{2,1}\rangle, \langle\phi_{1,3}\rangle, \langle\phi_{5,1}\rangle, \langle\phi_{2,3}\rangle = \mathcal{O}(M_{\rm EW}/M_{\rm GUT}),$$

which guarantees that the μ-term is of the electroweak scale. With these choices, the singlet insertions for the Yukawa matrices are given by the rank-two matrices

$$\Lambda^u = \begin{pmatrix} 0 & 1 & 1 \\ 1 & \phi_{1,2} & \phi_{1,2} \\ 1 & \phi_{1,2} & \phi_{1,2} \end{pmatrix},$$

$$\Lambda^d_{11} = \Phi_2\Phi_4,$$

$$\Lambda^d_{12} = \Phi_2^2\Phi_5\phi_{1,2}^2 + \Phi_2\Phi_4\phi_{2,5},$$

$$\Lambda^d_{13} = \Phi_2^2\Phi_5\phi_{1,2}^3 + \Phi_2\Phi_4\phi_{2,5}\phi_{1,2} + \Phi_1\Phi_2\Phi_4,$$

$$\Lambda^d_{21} = \Lambda^d_{31} = \Phi_2\Phi_4\phi_{1,2},$$

$$\Lambda^d_{22} = \Lambda^d_{32} = \Phi_2^2\Phi_5\phi_{1,2}^3 + \Phi_2\Phi_4\phi_{2,5}\phi_{1,2} + \Phi_1\Phi_2\Phi_4,$$

$$\Lambda^d_{23} = \Lambda^d_{33} = \Phi_2^2\Phi_5\phi_{1,2}^4 + \Phi_2\Phi_4\phi_{2,5}\phi_{1,2}^2 + \Phi_1\Phi_2\Phi_4\phi_{1,2},$$

omitting all terms of order $\mathcal{O}(M_{\rm EW}/M_{\rm GUT})$. Each entry in these matrices—more precisely, each individual term contributing to an entry when it is a sum of multiple terms—is multiplied by an order-one coefficient to produce the effective Yukawa matrices $Y^u_{IJ} = c^u_{IJ}\cdot\Lambda^u_{IJ}$, $Y^d_{IJ} = c^d_{IJ}\cdot\Lambda^d_{IJ}$, and $Y^e_{IJ} = c^e_{IJ}\cdot\Lambda^d_{IJ}$ (no summation). As discussed earlier, Y^d_{IJ} and Y^e_{IJ} are treated as independent even at the GUT scale. There is a large degeneracy among the possible choices of order-one coefficients and remaining VEVs that reproduce the observed masses and mixings in the IR. For illustration, we set

$$\langle\phi_{1,2}\rangle = \langle\phi_{2,5}\rangle = 0.7,$$
$$\langle\Phi_1\rangle = \langle\Phi_2\rangle = \langle\Phi_4\rangle = \langle\Phi_5\rangle = 0.07.$$

For the order-one coefficients in the up Yukawa matrix, one possible assignment is

$$c^u_{12} = -0.2117,\qquad c^u_{13} = -0.1764,$$
$$c^u_{21} = 0.2599,\qquad c^u_{22} = 3.0096,\qquad c^u_{23} = 2.4959,$$
$$c^u_{31} = -0.2061,\qquad c^u_{32} = -2.3937,\qquad c^u_{33} = -1.9841.$$

For the down and lepton Yukawa matrices, there are multiple insertions contributing to each entry. To illustrate a simple choice, we assume that all such insertions contributing to a given entry are multiplied by a common order-one coefficient:

$$c^d_{11} = -1.5947,\qquad c^d_{12} = 0.1109,\qquad c^d_{13} = 0.4794,$$
$$c^d_{21} = 5.3584,\qquad c^d_{22} = 0.1450,\qquad c^d_{23} = -1.5017,$$
$$c^d_{31} = -4.2375,\qquad c^d_{32} = -0.1191,\qquad c^d_{33} = 1.1682,$$
$$c^e_{11} = 0.8238,\qquad c^e_{12} = 1.2625,\qquad c^e_{13} = 0.8958,$$
$$c^e_{21} = -0.8783,\qquad c^e_{22} = -1.2210,\qquad c^e_{23} = -1.1892,$$
$$c^e_{31} = 1.0613,\qquad c^e_{32} = 1.5147,\qquad c^e_{33} = 1.6658.$$

It is straightforward to verify that these choices, together with the choice

$$(\langle H^u\rangle, \langle H^d\rangle) = (48.930, 166.979)\ \text{GeV}$$

for the Higgs VEVs reproduce the observed quark and charged lepton masses, as well as the CKM matrix. While many alternative choices are possible, the above values serve as a concrete proof of principle.

This model also admits a second, qualitatively distinct mechanism for generating an electroweak-scale μ-term. The key observation is that by choosing

$$\langle\Phi_1\rangle, \langle\phi_{5,1}\rangle, \langle\phi_{2,5}\rangle = \mathcal{O}\left(\sqrt{M_{\rm EW}/M_{\rm GUT}}\right),$$
$$\langle\Phi_3\rangle, \langle\phi_{2,1}\rangle, \langle\phi_{1,3}\rangle, \langle\phi_{2,3}\rangle = \mathcal{O}(M_{\rm EW}/M_{\rm GUT}),$$

all singlet insertions allowed in the μ-term are suppressed by at least a factor of $M_{\rm EW}/M_{\rm GUT}$. This effectively introduces an intermediate scale reminiscent of the Kim-Nilles mechanism [51] for solving the μ problem. In this scenario, the observed fermion masses and mixing parameters can still be reproduced with small adjustments to the order-one coefficients presented above, as can be verified by direct calculation.

2. Model 2 on $X_{6770}/\mathbb{Z}_2$

The low-energy spectrum includes the following fields:

Matter	$2(Q,u,e)_{\mathbf{e}_1}, (Q,u,e)_{\mathbf{e}_2}$, $2(d,L)_{\mathbf{e}_1+\mathbf{e}_5}, (d,L)_{\mathbf{e}_2+\mathbf{e}_5}$
Higgs	$H^u_{-\mathbf{e}_2-\mathbf{e}_5}, H^d_{\mathbf{e}_2+\mathbf{e}_5}$
Moduli	$8\phi_{1,2} \coloneqq 8\phi_{\mathbf{e}_1-\mathbf{e}_2}, 6\phi_{1,3} \coloneqq 6\phi_{\mathbf{e}_1-\mathbf{e}_3}$, $2\phi_{1,4} \coloneqq 2\phi_{\mathbf{e}_1-\mathbf{e}_4}, 4\phi_{5,1} \coloneqq 4\phi_{\mathbf{e}_5-\mathbf{e}_1}$, $4\phi_{2,4} \coloneqq 4\phi_{\mathbf{e}_2-\mathbf{e}_4}, \phi_{2,5} \coloneqq \phi_{\mathbf{e}_2-\mathbf{e}_5}$, $6\phi_{3,5} \coloneqq 6\phi_{\mathbf{e}_3-\mathbf{e}_5}$, $\Phi_1 \coloneqq \Phi_{(1,0,0,0,-1)}, \Phi_2 \coloneqq \Phi_{(-2,1,1,0,0)}$, $\Phi_3 \coloneqq \Phi_{(0,0,-1,0,1)}, \Phi_4 \coloneqq \Phi_{(1,-2,0,1,0)}$, $\Phi_5 \coloneqq \Phi_{(0,0,1,-1,0)}$.

To obtain the correct scale for the μ-term, we assign

$$\langle\Phi_1\rangle, \langle\Phi_3\rangle, \langle\phi_{1,2}\rangle, \langle\phi_{1,3}\rangle, \langle\phi_{1,4}\rangle, \langle\phi_{2,4}\rangle = \mathcal{O}(M_{\rm EW}/M_{\rm GUT}).$$

With these choices, the singlet insertions for the Yukawa matrices are given by the rank-two matrices

$$\Lambda^u = \begin{pmatrix} \phi_{2,5}\phi_{5,1}^2 & \phi_{2,5}\phi_{5,1}^2 & \phi_{5,1} \\ \phi_{2,5}\phi_{5,1}^2 & \phi_{2,5}\phi_{5,1}^2 & \phi_{5,1} \\ \phi_{5,1} & \phi_{5,1} & 0 \end{pmatrix},$$

$$\Lambda^d_{11} = \Lambda^d_{12} = \Lambda^d_{21} = \Lambda^d_{22} = \Phi_2\Phi_4\phi_{2,5} + \Phi_4\phi_{2,5}^2\phi_{3,5}\phi_{5,1}^2,$$
$$\Lambda^d_{13} = \Lambda^d_{23} = \Lambda^d_{31} = \Lambda^d_{32} = \Phi_4\phi_{2,5}\phi_{3,5}\phi_{5,1},$$
$$\Lambda^d_{33} = \Phi_4\phi_{3,5} + \Phi_4^2\Phi_5\phi_{2,5}^2\phi_{5,1},$$

where we have omitted all terms of order $\mathcal{O}(M_{\rm EW}/M_{\rm GUT})$. Analogous to the model discussed above, there is significant degeneracy among the possible choices of order-one coefficients and the remaining VEVs that reproduce the observed masses and mixings in the IR. For illustration, we display the following VEV choices:

$$\langle\phi_{5,1}\rangle = \langle\phi_{2,5}\rangle = 0.7,\qquad \langle\phi_{3,5}\rangle = 0.5,$$
$$\langle\Phi_2\rangle = 0.07, \langle\Phi_4\rangle = 0.01,\qquad \langle\Phi_5\rangle = 0.06.$$

For the order-one coefficients in the up Yukawa matrix, one possible assignment is

$$c^u_{11} = -1.8576, \quad c^u_{12} = 2.6696, \quad c^u_{13} = 0.1381,$$
$$c^u_{21} = -2.8519, \quad c^u_{22} = 4.1054, \quad c^u_{23} = 0.2093,$$
$$c^u_{31} = 0.1084, \quad c^u_{32} = -0.1776.$$

For the down-type quark and charged lepton Yukawa matrices, multiple insertions contribute to each entry. As before, for simplicity, we assume that all such insertions contributing to a given entry are multiplied by a single, common order-one coefficient:

$$c^d_{11} = 1.2406, \quad c^d_{12} = 2.5324, \quad c^d_{13} = -5.3256,$$
$$c^d_{21} = 1.9296, \quad c^d_{22} = 4.0020, \quad c^d_{23} = -8.4645,$$
$$c^d_{31} = -0.4503, \quad c^d_{32} = -0.8237, \quad c^d_{33} = 1.6341,$$
$$c^e_{11} = -2.8422, \quad c^e_{12} = -1.9381, \quad c^e_{13} = -2.3298,$$
$$c^e_{21} = 2.9123, \quad c^e_{22} = 2.0472, \quad c^e_{23} = 1.8753,$$
$$c^e_{31} = -0.9057, \quad c^e_{32} = -0.6317, \quad c^e_{33} = -0.4543.$$

For the Higgs VEVs, we set

$$(\langle H^u \rangle, \langle H^d \rangle) = (83.764, 152.511)\ \text{GeV}.$$

It is straightforward to verify that these choices reproduce the observed quark and charged lepton masses, as well as the CKM matrix. As in the previous model, here it is also possible to address the μ problem by introducing an intermediate scale of order $\sqrt{M_{\rm EW}/M_{\rm GUT}}$ and assigning the VEVs $\langle \phi_{3,5} \rangle$ and $\langle \Phi_3 \rangle$ to this scale.

IV. DISCUSSION AND CONCLUSIONS

In this work, we have presented explicit heterotic $E_8 \times E_8$ string compactifications on smooth Calabi-Yau threefolds with polystable, holomorphic, Abelian vector bundles that yield the exact (MS)SM particle spectrum with realistic flavor data and no exotic states. Building on the framework of line bundle models developed over the past decade, we have exploited the residual global $U(1)$ selection rules inherited from the internal gauge bundle to generate hierarchical Yukawa couplings in the effective theory, in analogy to Froggatt-Nielsen models. By analyzing these couplings under the constraints imposed by the $U(1)$ selection rules, we have identified explicit models capable of reproducing the observed quark and charged lepton masses and CKM mixing angles. While most singlet moduli were assumed to acquire generic, nonhierarchical VEVs, for each identified model we have determined the minimal subset that must be suppressed to generate an electroweak-scale μ-term via higher-dimensional operators.

Our main result is a proof of principle: There exist explicit choices of CY topology and vector bundle data in heterotic string theory that yield a fully realistic (MS)SM spectrum with the correct flavour structure and no exotic states. The models rely on the existence of an R-parity symmetry—geometrically realized via a discrete symmetry of the quotient CY threefold—that forbids all renormalizable R-parity violating operators and remains unbroken throughout moduli space, ensuring the absence of dangerous low-energy couplings even after the spontaneous breaking of the $U(1)$ s. The models identified here represent a significant step toward realizing string compactifications that reproduce the SM in full detail.

While the number of free parameters in each model is significant—typically ranging from 50 to 80 due to the many moduli fields and order-one coefficients—this should not be mistaken for an unconstrained fit. The $U(1)$ selection rules inherited from the compactification sharply restrict the allowed operator structure, enforcing hierarchical Yukawa textures via higher-dimensional couplings involving $U(1)$-charged singlet moduli fields. These selection rules, along with the charges and multiplicities of the singlet fields themselves, are entirely dictated by the compactification data and not introduced *ad hoc*. The (infrared) values of the order-one coefficients are themselves constrained: They are expected to remain within a reasonable order-one range consistent with what can typically be computed from the compactification and subsequently evolved down to the infrared via renormalization group running. In practice, we find that the vast majority of three-generation models in our scan fail to reproduce the observed flavor structure within these constraints, with only a small fraction successfully matching the measured masses and mixings. This indicates that the flavor structure arises nontrivially from the underlying compactification, rather than from arbitrary parameter tuning.

Of course, several open questions remain. Most notably, our analysis does not address the problem of moduli stabilization. While we identify regions of moduli space consistent with the observed flavor data, we do not demonstrate that these regions can be dynamically realized through a concrete stabilization mechanism. Additionally, the precise values of the order-one coefficients in the effective operators depend nontrivially on the CY geometry and can vary across moduli space. A complete UV realization would require computing these coefficients explicitly at a stabilized point using, for example, the neural network approach developed in Ref. [41].

Although we have not specified a stabilization mechanism, our analysis implicitly relies on the assumption of low-energy supersymmetry, which ensures that holomorphic quantities such as the superpotential Yukawa couplings are protected from large quantum corrections, making their tree-level determination meaningful. Even though supersymmetry breaking and moduli stabilization will generically shift the vacuum, one expects the leading-order flavor textures to remain robust provided the stabilized VEVs lie near the identified regions.

Future work should aim to embed these constructions into fully stabilized compactifications, incorporating fluxes or nonperturbative effects to fix the moduli dynamically while preserving the desired low-energy physics. It would also be valuable to extend the analysis to include neutrino masses and mixings, which could arise from similar higher-dimensional operators in the effective superpotential. Another promising direction is the systematic study of models in which exotic states at the Abelian locus can be lifted via singlet VEVs, realizing, for example, viable QCD axion scenarios with greater flexibility in the Yukawa structures. Further line bundle models capable of realizing the full range of SM flavor parameters can be identified by systematically searching the larger sets of three-generation models obtained in Refs. [8,9,11,13]. The main lesson of this work is that the systematic exploration of large classes of heterotic compactifications, supported by appropriate computational geometry tools, offers a promising route to uncovering fully realistic string-derived models of particle physics.

ACKNOWLEDGMENTS

We would like to thank Jonathan Patterson for assistance with the Oxford Theoretical Physics computing cluster Hydra and Russell Jones for general technical support. A. C. and L. A. N. are supported by a Royal Society Dorothy Hodgkin Fellowship, Grant No. DHF/R1/231142. L. T. Y. L. is supported by the Oxford-Croucher Scholarship. A. L. acknowledges support by the STFC consolidated Grant No. ST/X000761/1.

DATA AVAILABILITY

The data that support the findings of this article are openly available [49], embargo periods may apply.

APPENDIX A: SM YUKAWA COUPLINGS

The quark and charged lepton mass matrices are related to the Yukawa matrices through the relations

$$M^u = \langle H^u \rangle Y^u, \quad M^d = \langle H^d \rangle Y^d, \quad M^e = \langle H^d \rangle Y^e, \tag{A1}$$

where $\langle H^u \rangle$ and $\langle H^d \rangle$ satisfy

$$\sqrt{\langle H^u \rangle^2 + \langle H^d \rangle^2} \simeq 174 \text{ GeV}. \tag{A2}$$

The quark and charged lepton masses are obtained from the singular value decomposition of the mass matrices:

$$\begin{aligned} M^u &= U^u \hat{M}^u (V^u)^\dagger \quad \hat{M}^u = \mathrm{diag}(m_u, m_c, m_t), \\ M^d &= U^d \hat{M}^d (V^d)^\dagger \quad \hat{M}^d = \mathrm{diag}(m_d, m_s, m_b), \\ M^e &= U^e \hat{M}^e (V^e)^\dagger \quad \hat{M}^e = \mathrm{diag}(m_e, m_\mu, m_\tau), \end{aligned} \tag{A3}$$

TABLE II. Masses of the quarks and charged leptons [52].

Quark	m_u (GeV)	m_c (GeV)	m_t (GeV)
Mass	$0.00216^{+0.00049}_{-0.00026}$	1.27 ± 0.02	172.4 ± 0.07
Quark	m_d (GeV)	m_s (GeV)	m_b (GeV)
Mass	$0.00467^{+0.00048}_{-0.00017}$	$0.093^{+0.011}_{-0.005}$	$4.18^{+0.03}_{-0.02}$
Lepton	m_e (MeV)	m_μ (GeV)	m_τ (GeV)
Mass	0.511 ± 0.001	0.1057 ± 0.0001	1.77682 ± 0.00016

while the CKM quark mixing matrix is given by

$$V_{\mathrm{CKM}} = (U^u)^\dagger U^d. \tag{A4}$$

For completeness, the standard reference values for the quark and lepton masses are listed in Table II. We recall that the values for the charged lepton masses correspond to their physical (pole) masses, as these particles are directly observable. For quarks, which are confined, the masses are scheme- and scale-dependent parameters defined in a particular renormalization prescription. The light-quark (u, d, s) masses are computed using dimensional regularization in the minimal subtraction scheme at a scale of 2 GeV, while the charm and bottom quark masses, $m_c(m_c)$ and $m_b(m_b)$, are given at their own mass scales in the same scheme. The top quark mass is conventionally quoted as its pole mass.

APPENDIX B: SEARCH STRATEGY

The present search builds on the 202 line bundle models identified in Ref. [6] that yield three chiral families of fermions. After GUT breaking and computation of the graded cohomologies using the methods of Ref. [53], these models give rise to 2113 Standard-Model-like spectra without exotic fields, in particular, without Higgs triplets. Among these, 1568 models contain exactly one pair of Higgs doublets. However, many of these models are physically equivalent in the sense that they share identical low-energy spectra and couplings, leaving 226 inequivalent models in total. This subset forms the starting point of our analysis.

For each of these models, we systematically compute the $\mathcal{G}$-invariant operators in the superpotential that contribute to the effective Yukawa couplings (including all terms with up to seven singlet insertions) and to the effective μ-term (including terms with up to 15 singlet insertions). The μ-term must be suppressed to the electroweak scale—approximately 14 orders of magnitude below the GUT scale—and we identify the minimal subsets of singlet VEVs that need to be suppressed by this factor in order to achieve the required scale while allowing the remaining VEVs to take generic values of order 0.01–0.1. We then proceed as follows:

(1) Given the required suppression of certain singlet VEVs (by $\sim 10^{-14}$), these VEVs are set to zero when computing the flavor parameters. This can reduce the rank of the Yukawa matrices. If both up- and down-type Yukawa matrices become rank one under this restriction, the model is discarded.
(2) The CKM matrix in Eq. (A4) is enforced by randomly generating orthogonal matrices U^u and U^d until the product $(U^u)^\dagger U^d$ matches the target CKM matrix within a 10% average deviation (in the absolute values of its entries). Note that fitting the CP-violating phase in the CKM matrix can be easily accomplished, since every singlet VEV and order-one coefficient comes with a phase. We, therefore, choose all VEVs and order-one coefficients to be real in order to keep the dimension of the parameter space manageable.
(3) For each realization of the CKM matrix and each choice of suppressed singlet VEVs, there remain approximately 50–80 additional parameters that must be optimized to reproduce realistic quark and charged lepton masses. These include the Yukawa coefficients, the up-type Higgs VEV, the unsuppressed singlet VEVs, and the entries of V^u, V^d, U^e, and V^e. The number of moduli fields and the Yukawa coefficients varies from model to model, with the latter depending on the number of leading operators appearing in the Yukawa matrices. We perform the optimization of parameters using the trust-region reflective algorithm, a numerical method that enforces the parameter bounds specified below while minimizing the loss function

$$L = \sum_{\lambda=u,d,e} \|M^\lambda - U^\lambda \hat{M}^\lambda (V^\lambda)^\dagger\|_F^2, \qquad \text{(B1)}$$

where $\|\cdot\|_F$ is the Frobenius norm. Models for which $L < 10^{-4}$ are classified as "viable." For each configuration of suppressed singlet VEVs consistent with solving the μ problem, we run the optimization over several hundred random initialization points and CKM matrix realizations.

The bounds imposed on the parameters in the optimization are chosen to reflect physically reasonable values. The bundle moduli VEVs $\langle \phi_a \rangle$ are allowed to vary between 0.1 and 0.7, while the nonperturbative moduli VEVs $\langle \Phi_i \rangle$ range from 0.01 to 0.07. The order-one Yukawa coefficients c^λ_{IJ} (for $\lambda = u, d, e$) are constrained as $0.1 < |c^\lambda_{IJ}| < 9$. The up-type Higgs VEV v_u is varied between 40 and 170 GeV, consistent with electroweak symmetry breaking constraints. Finally, the 12 Euler angles used to parametrize the residual orthogonal matrices (after fixing U^u and U^d) are left unconstrained.

[1] P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Vacuum configurations for superstrings, Nucl. Phys. **B258**, 46 (1985).
[2] J. Distler and B. R. Greene, Aspects of (2,0) string compactifications, Nucl. Phys. **B304**, 1 (1988).
[3] R. Blumenhagen, G. Honecker, and T. Weigand, Loop-corrected compactifications of the heterotic string with line bundles, J. High Energy Phys. 06 (2005) 020.
[4] R. Blumenhagen, S. Moster, and T. Weigand, Heterotic GUT and standard model vacua from simply connected Calabi-Yau manifolds, Nucl. Phys. **B751**, 186 (2006).
[5] R. Blumenhagen, S. Moster, R. Reinbacher, and T. Weigand, Massless spectra of three generation U(N) heterotic string vacua, J. High Energy Phys. 05 (2007) 041.
[6] L. B. Anderson, J. Gray, A. Lukas, and E. Palti, Two hundred heterotic standard models on smooth Calabi-Yau threefolds, Phys. Rev. D **84**, 106005 (2011).
[7] L. B. Anderson, J. Gray, A. Lukas, and E. Palti, Heterotic line bundle standard models, J. High Energy Phys. 06 (2012) 113.
[8] L. B. Anderson, A. Constantin, J. Gray, A. Lukas, and E. Palti, A comprehensive scan for heterotic SU(5) GUT models, J. High Energy Phys. 01 (2014) 047.
[9] Y.-H. He, S.-J. Lee, A. Lukas, and C. Sun, Heterotic model building: 16 Special manifolds, J. High Energy Phys. 06 (2014) 077.
[10] A. P. Braun, C. R. Brodie, and A. Lukas, Heterotic line bundle models on elliptically fibered Calabi-Yau threefolds, J. High Energy Phys. 04 (2018) 087.
[11] A. Constantin, Y.-H. He, and A. Lukas, Counting string theory standard models, Phys. Lett. B **792**, 258 (2019).
[12] M. Larfors and R. Schneider, Explore and exploit with heterotic line bundle models, Fortschr. Phys. **68**, 2000034 (2020).
[13] S. A. Abel, A. Constantin, T. R. Harvey, A. Lukas, and L. A. Nutricati, Decoding nature with nature's tools: Heterotic line bundle models of particle physics with genetic algorithms and quantum annealing, Fortschr. Phys. **72**, 2300260 (2024).
[14] T. Hubsch, *Calabi-Yau Manifolds: A Bestiary for Physicists* (World Scientific, Singapore, 1994).
[15] L. B. Anderson, J. Gray, Y.-H. He, S.-J. Lee, and A. Lukas, CICY package, based on methods described in arXiv:0911.1569, arXiv:0911.0865, arXiv:0805.2875, arXiv:hep-th/0703249, arXiv:hep-th/0702210.
[16] cohomCalg package, 2010, High-performance line bundle cohomology computation based on methods described in

arXiv:1003.5217, arXiv:1006.2392, arXiv:1006.0780. Download link: http://wwwth.mppmu.mpg.de/members/blumenha/cohomcalg/.

[17] M. Larfors and R. Schneider, pyCICY—A Python CICY toolkit, 10.5281/zenodo.3243914, [github-link] (2019).

[18] A. Constantin and A. Lukas, Formulae for line bundle cohomology on Calabi-Yau threefolds, Fortschr. Phys. **67**, 1900084 (2019).

[19] D. Klaewer and L. Schlechter, Machine learning line bundle cohomologies of hypersurfaces in toric varieties, Phys. Lett. B **789**, 438 (2019).

[20] M. Larfors and R. Schneider, Line bundle cohomologies on CICYs with Picard number two, Fortschr. Phys. **67**, 1900083 (2019).

[21] C. R. Brodie, A. Constantin, R. Deen, and A. Lukas, Machine learning line bundle cohomology, Fortschr. Phys. **68**, 1900087 (2020).

[22] C. R. Brodie and A. Constantin, Cohomology chambers on complex surfaces and elliptically fibered Calabi–Yau threefolds, Commun. Math. Phys. **405**, 151 (2024).

[23] C. R. Brodie, A. Constantin, and A. Lukas, Flops, Gromov-Witten invariants and symmetries of line bundle cohomology on Calabi-Yau three-folds, J. Geom. Phys. **171**, 104398 (2022).

[24] C. R. Brodie, A. Constantin, A. Lukas, and F. Ruehle, Geodesics in the extended Kähler cone of Calabi-Yau threefolds, J. High Energy Phys. 03 (2022) 024.

[25] A. Constantin, Generating functions for line bundle cohomology dimensions on complex projective varieties, Exp. Math. **34**, 459 (2024).

[26] V. Braun, T. Brelidze, M. R. Douglas, and B. A. Ovrut, Calabi-Yau metrics for quotients and complete intersections, J. High Energy Phys. 05 (2008) 080.

[27] M. R. Douglas, R. L. Karp, S. Lukic, and R. Reinbacher, Numerical Calabi-Yau metrics, J. Math. Phys. (N.Y.) **49**, 032302 (2008).

[28] M. R. Douglas, R. L. Karp, S. Lukic, and R. Reinbacher, Numerical solution to the Hermitian Yang-Mills equation on the Fermat quintic, J. High Energy Phys. 12 (2007) 083.

[29] L. B. Anderson, V. Braun, and B. A. Ovrut, Numerical Hermitian Yang-Mills connections and Kahler cone substructure, J. High Energy Phys. 01 (2012) 014.

[30] L. B. Anderson, V. Braun, R. L. Karp, and B. A. Ovrut, Numerical Hermitian Yang-Mills connections and vector bundle stability in heterotic theories, J. High Energy Phys. 06 (2010) 107.

[31] S. Blesneag, E. I. Buchbinder, P. Candelas, and A. Lukas, Holomorphic Yukawa couplings in heterotic string theory, J. High Energy Phys. 01 (2016) 152.

[32] S. Blesneag, E. I. Buchbinder, and A. Lukas, Holomorphic Yukawa couplings for complete intersection Calabi-Yau manifolds, J. High Energy Phys. 01 (2017) 119.

[33] A. Ashmore, Y.-H. He, and B. A. Ovrut, Machine learning Calabi–Yau metrics, Fortschr. Phys. **68**, 2000068 (2020).

[34] L. B. Anderson, M. Gerdes, J. Gray, S. Krippendorf, N. Raghuram, and F. Ruehle, Moduli-dependent Calabi-Yau and SU(3)-structure metrics from machine learning, J. High Energy Phys. 05 (2021) 013.

[35] V. Jejjala, D. K. Mayorga Pena, and C. Mishra, Neural network approximations for Calabi-Yau metrics, J. High Energy Phys. 08 (2022) 105.

[36] M. R. Douglas, S. Lakshminarasimhan, and Y. Qi, Numerical Calabi-Yau metrics from holomorphic networks, arXiv:2012.04797.

[37] M. Larfors, A. Lukas, F. Ruehle, and R. Schneider, Learning size and shape of Calabi-Yau spaces, arXiv:2111.01436.

[38] M. Larfors, A. Lukas, F. Ruehle, and R. Schneider, Numerical metrics for complete intersection and Kreuzer–Skarke Calabi–Yau manifolds, Mach. Learn. Sci. Tech. **3**, 035014 (2022).

[39] A. Ashmore, R. Deen, Y.-H. He, and B. A. Ovrut, Machine learning line bundle connections, Phys. Lett. B **827**, 136972 (2022).

[40] A. Ashmore, Y.-H. He, E. Heyes, and B. A. Ovrut, Numerical spectra of the Laplacian for line bundles on Calabi-Yau hypersurfaces, J. High Energy Phys. 07 (2023) 164.

[41] A. Constantin, C. S. Fraser-Taliente, T. R. Harvey, A. Lukas, and B. Ovrut, Computation of quark masses from string theory, Nucl. Phys. **B1010**, 116778 (2025).

[42] G. Butbaia, D. Mayorga Peña, J. Tan, P. Berglund, T. Hübsch, V. Jejjala, and C. Mishra, Physical Yukawa couplings in heterotic string compactifications, Adv. Theor. Math. Phys. **28**, 2783 (2024).

[43] C. Froggatt and H. Nielsen, Hierarchy of quark masses, cabibbo angles and *CP* violation, Nucl. Phys. **B147**, 277 (1979).

[44] A. Constantin, C. S. Fraser-Taliente, T. R. Harvey, L. T. Y. Leung, and A. Lukas, Fermion masses and mixing in string-inspired models, J. High Energy Phys. 06 (2025) 175.

[45] E. I. Buchbinder, A. Constantin, J. Gray, and A. Lukas, Yukawa unification in heterotic string theory, Phys. Rev. D **94**, 046005 (2016).

[46] E. I. Buchbinder, A. Constantin, and A. Lukas, Heterotic QCD axion, Phys. Rev. D **91**, 046010 (2015).

[47] E. I. Buchbinder, A. Constantin, and A. Lukas, The moduli space of heterotic line bundle models: A case study for the tetra-quadric, J. High Energy Phys. 03 (2014) 025.

[48] E. I. Buchbinder, A. Constantin, and A. Lukas, Non-generic couplings in supersymmetric standard models, Phys. Lett. B **748**, 251 (2015).

[49] A. Constantin, L. T. Y. Leung, A. Lukas, and L. A. Nutricati, Heterotic Line Bundle Models—Search Algorithm, https://github.com/dfgoe55/heterotic_line_bundle_models (2026). GitHub repository.

[50] V. Braun, On free quotients of complete intersection Calabi-Yau manifolds, J. High Energy Phys. 04 (2011) 005.

[51] J. E. Kim and H. P. Nilles, The mu problem and the strong *CP* problem, Phys. Lett. **138B**, 150 (1984).

[52] S. Navas *et al.* (Particle Data Group), Review of particle physics, Phys. Rev. D **110**, 030001 (2024).

[53] A. Constantin, J. Gray, and A. Lukas, Hodge numbers for all CICY quotients, J. High Energy Phys. 01 (2017) 001.