Commun. Math. Phys. 326, 287–327 (2014)
Digital Object Identifier (DOI) 10.1007/s00220-013-1878-8

Communications in
**Mathematical
Physics**

Higgs Bundles and UV Completion in F-Theory

Ron Donagi[1], **Martijn Wijnholt**[2]

[1] Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104-6395, USA
[2] Max Planck Institute (Albert Einstein Institute), Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.
E-mail: Martijn.Wijnholt@physik.uni-muenchen.de

Received: 3 October 2012 / Accepted: 6 June 2013
Published online: 31 January 2014 –

Abstract: F-theory admits 7-branes with exceptional gauge symmetries, which can be compactified to give phenomenological four-dimensional GUT models. Here we study general supersymmetric compactifications of eight-dimensional Yang–Mills theory. They are mathematically described by meromorphic Higgs bundles, and therefore admit a spectral cover description. This allows us to give a rigorous and intrinsic construction of local models in F-theory. We use our results to prove a no-go theorem showing that local $SU(5)$ models with three generations do not exist for generic moduli. However we show that three-generation models do exist on the Noether–Lefschetz locus. We explain how F-theory models can be mapped to non-perturbative orientifold models using a scaling limit proposed by Sen. Further we address the construction of global models that do not have heterotic duals, considering models with base $\mathbf{CP}^3$ or a blow-up thereof as examples. We show how one may obtain a contractible worldvolume with a two-cycle not inherited from the bulk, a necessary condition for implementing GUT breaking using fluxes.

Contents

1. Introduction

Recently [1–3] initiated a systematic effort to study a new class of Kaluza–Klein models for Grand Unification. The basic idea was to use an eight-dimensional gauge theory with an exceptional gauge group, coupled to ten-dimensional type IIb supergravity. Although this theory is non-renormalizable, there is a lot of evidence that it admits a UV completion which is called F-theory [4]. For practical purposes however, very little is known about this non-perturbative completion, as it is not in reach of perturbative string theory. We essentially only know the low energy gauge theory and supergravity Lagrangians, which are uniquely determined by the symmetries. To get a weakly coupled description in which these Lagrangians can be trusted, the fields must be slowly varying. Thus these models can still have a weakly coupled description if we take a large volume limit. Recent work on F-theory models includes [5–20].

A central role in the analysis is played by local models, by which we mean Calabi–Yau fourfolds consisting of an ALE fibration over a compact complex surface S, or sometimes a minimal compactification thereof. Despite the conceptual progress in [1–3], there were a number of unanswered questions about the explicit construction of local models in F-theory. In particular, the strategy in [1] (and also [3]) relied on taking a limiting form of models with a heterotic dual. This approach yields manifestly consistent models, but it was perhaps less than clear if the most general local F-theory model is recovered this way. The approach in [2] is to postulate the matter curves and the fluxes restricted to the matter curves. At first sight this looks more flexible, but in this case it is less clear if the data is mutually consistent. Given the uncertainties, it can be hard to evaluate what F-theory does or does not predict.

The first purpose of this paper is to give a rigorous and intrinsic construction of local F-theory models. The chain of logic is as follows. As mentioned above, the basic idea is that in order to get a weakly coupled description, we want to construct compactifications of a supersymmetric eight-dimensional gauge theory. We show that such compactifications are mathematically described by meromorphic Higgs bundles on S. Unfortunately such Higgs bundles are practically impossible to write down explicitly, and furthermore they look very different from the traditional description of F-theory, which uses elliptically fibered fourfolds. To address this, we are going to discuss explicit isomorphisms between the (holomorphic data of the) following three types of objects:

$$\text{Higgs bundles} \quad \longleftrightarrow \quad \text{spectral covers} \quad \longleftrightarrow \quad \text{ALE fibrations.} \tag{1.1}$$

The first arrow is a relatively well-known fact about Higgs bundles. (Its relevance to F-theory was noted independently in [10], which appeared while this project was written up.) The second arrow however has not been appreciated, and it is the one that allows us to make contact with the older descriptions of F-theory. The existence of these equivalences, explained in Sect. 2.4, is one of our key messages in the first part of this paper.

Now the first and the third object in (1.1) are not so suitable for constructions. In the case of Higgs bundles, although this description is weakly coupled in the derivative expansion, the problem is that the hermitian metric on the bundle satisfies a non-linear PDE and is practically impossible to write explicitly. For ALE fibrations, the problem is that a crucial part of the data is given by the G-flux (or more precisely a relative Deligne cohomology class). The spectral cover approach is the most transparent, thus it is this approach that allows us to make constructions, particularly of the fluxes, and compute the spectrum and interactions.

Finally, our ALE fibrations should admit an elliptic fibration as well. If we assume this elliptic fibration to be generically non-degenerate, then this singles out the ALE spaces corresponding to the exceptional groups.[1] These are also exactly the ALE fibrations that one obtains from a scaling limit of heterotic/F-theory duality. Thus, the models obtained from our original strategy are actually rather general, in particular they do not require heterotic duals. We will discuss in Sect. 2.1 how an E_8 structure emerges when we extract a local model from a global model, in agreement with the argument above. Furthermore, it is clear that our logic is not limited to F-theory. A completely parallel construction of local M-theory models based on a set of equivalences analogous to (1.1) will appear in [22].

The spectral cover approach allows us to give a precise description of the configuration space of local F-theory models, which is important for phenomenological applications. We will use this to classify the possible matter curve configurations and prove a no-go theorem, showing that the fluxes which were known to exist do not allow for a local $SU(5)$ model with three generations. (The use of 'local' in this claim is slightly different from the definition above, and explained in Sect. 2.6.) This is seen to imply that in order to find realistic models, we have to solve a Noether–Lefschetz problem, i.e. we have to tune the complex structure moduli of a local model in order to find supersymmetric solutions with three generations (which will then automatically have stabilized some of the moduli). We then write down some new classes of fluxes which are available on the Noether–Lefschetz locus, and find the first examples of three-generation models.

Such more general Noether–Lefschetz fluxes are also available in heterotic models, where they generally get mapped to *rigid* bundles after a Fourier–Mukai transform. This gives rise to a large (in fact, exponentially large) number of new heterotic models which have been overlooked in the literature. In fact we will point out that heterotic constructions to date have been very special and essentially missed the landscape seen on the type II side, most of which gets mapped to a landscape of rigid bundles under heterotic/F-theory duality.

Along the way we discuss several other interesting issues. In Sect. 2.2 we will discuss orientifold limits of F-theory models with non-trivial gauge groups. For $SU(5)_{GUT}$ models we find that one typically ends up with a IIb model with singularities, and this is related to the problems of generating perturbative up-type Yukawa couplings in such models. In Sect. 2.3 we will give an interpretation of a relation between the homology classes of the matter curves in terms of anomalies.

The second purpose of this paper is to begin the systematic construction of global UV completions of local models. At present, the only known global UV completions have heterotic duals. Here we want to construct global models which do not have a heterotic

[1] If we allow for conic bundles (which can be promoted to degenerate elliptic fibrations) then we also get classical groups [21].

dual, and in which hypercharge remains massless. This section was originally to appear as Sect. 2.5 of [5], but seemed to fit better with this paper. We will give some examples based on $\mathbf{P}^3$ and a blow-up thereof, which should make the general strategy clear.

One might have thought that requiring the existence of a global UV completion places severe constraints on the local model, but this does not appear to be true. For example we do not find any meaningful constraints on extending desired values of complex structure moduli from a local model to a global model, thereby further validating the idea of studying local models. In particular we find that it is possible to set the complex structure moduli so that no dimension four or five proton decay can be generated. The remarkable agreement between local models and global models has recently been explained more conceptually in [23].

The understanding of global models is unfortunately still somewhat incomplete. Our discussion here focuses on constructing compact models with desired 7-brane configurations. Constructing suitable G-fluxes is again an incarnation of a Noether–Lefschetz problem. While one could construct examples of such fluxes (eg. following the work of Noether and Lefschetz), there is a landscape problem, and this affects the phenomenological questions one would like to ask.

2. Higgs Bundles in F-Theory

In this section we will give a detailed description of local F-theory models. Although much of this material is described implicitly or explicitly in our previous papers, writing out the chain of logic more carefully allows us to make sharper statements about the configuration space of such models.

The reader should be aware that on occasion we use two different definitions of the notion of a local model. The more physical definition, which we adopt in Sect. 2.6, is that of a model in which M_{GUT}/M_{Pl} can be made parametrically small. The other definition, which we adopt in most of the rest of this paper, is that of a non-compact CY_4 consisting of an ALE fibration over a compact, complex surface. Hopefully it is clear from the context which notion we use.

2.1. Local model from global model. Let us start with a global model, which is defined as a compact elliptically fibered Calabi–Yau complex fourfold with a section $\sigma(B_3)$ (often simply written as B_3). The elliptic fibration can be described by a Weierstrass model

$$y^2 = x^3 + fx + g, \tag{2.1}$$

where f, g are sections of $K_{B_3}^{-4}, K_{B_3}^{-6}$ respectively. For the purpose of detecting singularities, it is more useful to write the Weierstrass equation in generalized form as

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{2.2}$$

where the a_i are sections of $K_{B_3}^{-i}$. By completing the square and the cube, this may be written as (2.1), but the generalized form is more convenient for prescribing singular elliptic fibers along loci in B_3.

Suppose that we have a surface S of singularities in B_3. This will put certain restrictions on the sections a_i above. Let us take z to be a coordinate on the normal bundle to S in B_3, so S corresponds to $z = 0$. We will often denote $c_1(NS) = -t$. Then the order

Table 1. Results from Tate's algorithm [24,25]

Type	Group	a_1	a_2	a_3	a_4	a_6	Δ
I_0	–	0	0	0	0	0	0
I_1	–	0	0	1	1	1	1
I_2	$SU(2)$	0	0	1	1	2	2
I_3^{ns}	Unconven	0	0	2	2	3	3
I_3^{s}	Unconven	0	1	1	2	3	3
I_{2k}^{ns}	$Sp(k)$	0	0	k	k	$2k$	$2k$
I_{2k}^{s}	$SU(2k)$	0	1	k	k	$2k$	$2k$
I_{2k+1}^{ns}	Unconven	0	0	$k+1$	$k+1$	$2k+1$	$2k+1$
I_{2k+1}^{s}	$SU(2k+1)$	0	1	k	$k+1$	$2k+1$	$2k+1$
II	–	1	1	1	1	1	2
III	$SU(2)$	1	1	1	1	2	3
IV^{ns}	Unconven	1	1	1	2	2	4
IV^{s}	$SU(3)$	1	1	1	2	3	4
I_0^{*ns}	G_2	1	1	2	2	3	6
I_0^{*ss}	$SO(7)$	1	1	2	2	4	6
I_0^{*s}	$SO(8)^*$	1	1	2	2	4	6
I_1^{*ns}	$SO(9)$	1	1	2	3	4	7
I_1^{*s}	$SO(10)$	1	1	2	3	5	7
I_2^{*ns}	$SO(11)$	1	1	3	3	5	8
I_2^{*s}	$SO(12)^*$	1	1	3	3	5	8
I_{2k-3}^{*ns}	$SO(4k+1)$	1	1	k	$k+1$	$2k$	$2k+3$
I_{2k-3}^{*s}	$SO(4k+2)$	1	1	k	$k+1$	$2k+1$	$2k+3$
I_{2k-2}^{*ns}	$SO(4k+3)$	1	1	$k+1$	$k+1$	$2k+1$	$2k+4$
I_{2k-2}^{*s}	$SO(4k+4)^*$	1	1	$k+1$	$k+1$	$2k+1$	$2k+4$
IV^{*ns}	F_4	1	2	2	3	4	8
IV^{*s}	E_6	1	2	2	3	5	8
III^*	E_7	1	2	3	3	5	9
II^*	E_8	1	2	3	4	5	10
Non-min	–	1	2	3	4	6	12

The superscripts s/ns stand for split/non-split, indicating the absence/presence of a monodromy action by an outer automorphism on the vanishing cycles along the singular locus

of vanishing of the a_i may increase at $z = 0$, so there will be conditions of the form 'z divides a_i at least n_i times', which are characteristic of the singularity type of the elliptic fiber over $z = 0$. These conditions have been worked out in [24,25] and are given in Table 1 which was taken from [25]. In retrospect, the table is perhaps better understood in terms of Higgs bundles, which we will discuss later. Now to get a local model from a global model, we assign scaling dimensions to (x, y, z) and drop the irrelevant terms. Physically, this should correspond to dropping certain higher order terms in the $8d$ gauge theory.

For phenomenological purposes the case of most interest is a surface S of I_5 singular fibers. Then according to Table 1, in order to have an $SU(5)$ singularity along $z = 0$, we need the leading terms near $z = 0$ to be

$$a_1 = -b_5, \quad a_2 = zb_4, \quad a_3 = -z^2 b_3, \quad a_4 = z^3 b_2, \quad a_6 = z^5 b_0, \tag{2.3}$$

where the b_i are generically non-vanishing, and we may have further subleading terms which vanish to higher order in z. The b_i are independent of z, so we may think of the b_i as sections of line bundles on the surface S. Now we assign scaling dimensions $(1/3, 1/2, 1/5)$ to (x, y, z) respectively. We throw out the 'irrelevant terms' whose scaling dimension is larger than one. The resulting equation we get is

$$y^2 = x^3 + b_0 z^5 + b_2 x z^3 + b_3 y z^2 + b_4 x^2 z + b_5 x y, \tag{2.4}$$

which is exactly the equation of an E_8 singularity unfolded to an $SU(5)$ singularity. The dimension one terms give the E_8 singularity and the terms with dimension smaller than one give a relevant deformation of this singularity. Thus we may extract an ALE fibration over S from a global model by taking a scaling limit. Note that $c_1(B_3)|_S = c_1(S) - t$, and so the above equation transforms as a section of $6c_1(S) - 6t$. Therefore the Chern classes of the sections b_i on S are given by

$$b_i \sim (6 - i)c_1(S) - t. \tag{2.5}$$

It seems rather remarkable that we have arrived an E_8-structure under some rather mild assumptions. This has recently been explained more rigorously using the notion of semi-stable degeneration [23]. The result is that if the elliptic fibration can be put in Tate form as above (and perhaps more generally) and if the order of vanishing of the a_i satisfies $n_i \leq i$, then we can define a degeneration limit in which a local E_8 model bubbles off from the rest of the Calabi–Yau.

Thus part of the attraction of local F-theory models is that almost all of the observable sector is described by this one eq. (2.4), plus a choice of G-fluxes. All the usual complications of global models can be hidden in the subleading corrections to this equation, and can be turned off by taking a suitable degeneration limit of the global model. This is equivalent to the statement that the local geometry is completely described by the $8d$ gauge theory. In the following we will analyze these local geometries in more detail.

2.2. Orientifold limits. In this section, we analyze IIb limits of F-theory vacua. Since the study of compactifications of perturbative IIb is a relatively well-developed subject, this should give a useful cross-check on our understanding of F-theory. However as we will discuss the regimes of validity are not overlapping and the IIb models we get look very different from any previously considered IIb GUT-like models. Thus there is still some work to be done to understand the relation between the two pictures.

Consider again the Weierstrass equation

$$y^2 = x^3 + fx + g \tag{2.6}$$

and its generalized form

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6. \tag{2.7}$$

As in [24], we define the following quantities:

$$\begin{array}{ll} \mathsf{b}_2 = a_1^2 + 4a_2, & \mathsf{b}_8 = \frac{1}{4}(\mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2), \\ \mathsf{b}_4 = a_1 a_3 + 2a_4, & \Delta = -\mathsf{b}_2^2\mathsf{b}_8 - 8\mathsf{b}_4^3 - 27\mathsf{b}_6^2 + 9\mathsf{b}_2\mathsf{b}_4\mathsf{b}_6. \\ \mathsf{b}_6 = a_3^2 + 4a_6 & \end{array} \tag{2.8}$$

Then f and g may be recovered as

$$f = -\frac{1}{48}(\mathsf{b}_2^2 - 24\mathsf{b}_4),$$
$$g = -\frac{1}{864}(-\mathsf{b}_2^3 + 36\mathsf{b}_2\mathsf{b}_4 - 216\mathsf{b}_6). \tag{2.9}$$

Now supposed that we want to take a limit in the complex structure moduli space so that the expectation value of the string coupling goes to zero almost everywhere in the IIb space-time. Since

$$j(\tau) = 4\frac{(24f)^3}{4f^3 + 27g^2}, \tag{2.10}$$

this will happen when

$$\frac{f^3}{4f^3 + 27g^2} \to \infty. \tag{2.11}$$

Inspecting (2.9), we see that the most evident way to achieve this is by scaling up b_2, or alternatively by scaling down b_4 and b_6. Therefore let us consider the following scaling limit:

$$a_3 \to \epsilon\, a_3, \qquad a_4 \to \epsilon\, a_4, \qquad a_6 \to \epsilon^2\, a_6. \tag{2.12}$$

Note that for our GUT models (2.4), in this limit b_i/b_0 scales like $1/\epsilon$ or $1/\epsilon^2$. Since $b_i/b_0 \sim \mathrm{Tr}(\Phi^i)$ are identified with Casimirs of the eight-dimensional Higgs field, this means that the VEV of the Higgs field is becoming large and we can no longer trust the $8d$ gauge theory/F-theory description. One may still hope to get a different weakly coupled description in terms of perturbative IIb string theory. As we will discuss, this is possible, but we have to push the model through a configuration with singularities that are neither well-described by F-theory nor by perturbative type IIb.

Continuing, one finds

$$f = -\frac{1}{48}(\mathsf{b}_2^2 - 24\epsilon\, \mathsf{b}_4),$$
$$g = -\frac{1}{864}(-\mathsf{b}_2^3 + 36\epsilon\, \mathsf{b}_2\mathsf{b}_4 - 216\epsilon^2\, \mathsf{b}_6). \tag{2.13}$$

The discriminant is given by

$$\Delta = \epsilon^2(-\mathsf{b}_2^2\mathsf{b}_8 - 8\epsilon\, \mathsf{b}_4^3 - 27\epsilon^2\mathsf{b}_6^2 + 9\epsilon\, \mathsf{b}_2\mathsf{b}_4\mathsf{b}_6)$$
$$\sim -\frac{1}{4}\epsilon^2\, \mathsf{b}_2^2(\mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2) + \mathcal{O}(\epsilon^3). \tag{2.14}$$

Therefore in the $\epsilon \to 0$ limit, all the roots are located at $\mathsf{b}_2 = 0$ and $\mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2 = 0$. The monodromies around these roots were analyzed in [26,27], with the result that

$$O7 : \mathsf{b}_2 = 0, \qquad D7 : \mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2 = 0. \tag{2.15}$$

Moreover, the j-function behaves as

$$j(\tau) \sim \frac{\mathsf{b}_2^4}{\epsilon^2\,(\mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2)}, \tag{2.16}$$

which means that the string coupling goes to zero almost everywhere. Therefore we get the following picture [26]: in the limit of complex structure moduli space that we discussed above, we can extract a Calabi–Yau threefold given by

$$\xi^2 = \mathsf{b}_2, \tag{2.17}$$

where $\mathsf{b}_2 \sim K_{B_3}^{-2}$, $\xi \sim K_{B_3}^{-1}$. That is, the emerging CY_3 is simply the double cover over B_3 with branch locus given by $\mathsf{b}_2 = 0$. The orientifolding acts as

$$\xi \to -\xi, \quad y \to -y, \tag{2.18}$$

and the positions of the branes on this threefold are given as above. There are two copies of the $D7$ locus $\mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2 = 0$ related by $\xi \to -\xi$.

There are unfortunately a number of shortcomings with this analysis. The relation of the IIb threefold to the F-theory fourfold is not clear, and the above prescription does not explain how the 7-brane worldvolume fields in IIb and in F-theory are related. Recently a semi-stable version of the Sen degeneration has been found [21], and these problems have been resolved. But here we will restrict ourselves to the old approach.

Now let's apply this picture to our local models. The Calabi–Yau threefold will be given by a double cover of the total space of the normal bundle $N_S \to S$, with branch locus given by $\mathsf{b}_2 = 0$. For $SU(5)$ models we get

$$\begin{aligned}
\mathsf{b}_2 &= b_5^2 + 4zb_4, \\
\mathsf{b}_4 &= z^2 b_3 b_5 + 2z^3 b_2, \\
\mathsf{b}_6 &= z^4 b_3^2 + 4z^5 b_0, \\
\mathsf{b}_2\mathsf{b}_6 - \mathsf{b}_4^2 &= z^5(4b_3^2 b_4 - 4b_2 b_3 b_5 + 4b_0 b_5^2 + z(16 b_0 b_4 - 4b_2^2)),
\end{aligned} \tag{2.19}$$

where z is a local coordinate on the normal bundle N_S. Hence we find a non-compact $O7$-plane along the branch locus $\mathsf{b}_2 = 0$, five gauge $D7$-branes wrapped on S, as well as a non-compact flavour $D7$-brane. The $O7$-plane intersects the gauge 7-branes along the matter curve,

$$\Sigma_{\mathbf{10}} = \{b_5 = 0\}, \tag{2.20}$$

which as expected carries an enhanced $SO(10)$ singularity. The flavour $D7$-brane intersects the gauge $D7$-brane along

$$\Sigma_{\mathbf{5}} = \{R = b_3^2 b_4 - b_2 b_3 b_5 + b_0 b_5^2 = 0\} \tag{2.21}$$

which carries an enhanced $SU(6)$ singularity. Finally the Yukawa couplings are localized at

$$\lambda_{\text{top}} \sim \{b_5 = b_4 = 0\}, \qquad \lambda_{\text{bottom}} \sim \{b_5 = b_3 = 0\}, \tag{2.22}$$

which carry enhanced E_6 and $SO(12)$ singularities, respectively.

Let us now look in more detail at the points of E_6 enhancement. The equation of the Calabi–Yau can be written as

$$\xi^2 = u^2 + zw, \tag{2.23}$$

where $u = b_5$ and $w = 4b_4$. Thus the E_6 points are conifold singularities of the Calabi–Yau threefold. We expect that the limiting model has zero B_{NS}-field through the vanishing S^2, so that it corresponds to a non-perturbative singularity of type IIb.

Perturbative string theory breaks down at such conifold singularities, and there are extra massless states. This should be a chiral field corresponding to the zero modes of B_2, C_2 on the 'resolved' picture, or to a $D3$ wrapped on the vanishing S^3 in the deformed picture. As we will see below, this chiral field is likely charged under an additional light $U(1)$ gauge symmetry appearing in the IIb limit.

In order to get a perturbative picture, we can try to resolve or deform the conifold singularity. Let us first discuss the resolutions. The two small $\mathbf{P}^1$'s are exchanged under the discrete symmetry $\sigma : \xi \to -\xi$, and thus the small resolution is projected out by the orientifold. The full orientifold action is given by $\Omega(-1)^{F_L}\sigma$, where Ω is worldsheet parity and $(-1)^{F_L}$ maps the RR fields to minus themselves. The NS B-field is odd under $\Omega(-1)^{F_L}$, so it is consistent to have a non-zero value of B through the vanishing $\mathbf{P}^1$. So one can 'resolve' the singularity by turning on the B-field. (C_2 may also be non-zero; it is paired with B_2 under SUSY). However there is no smooth geometric picture, and α' corrections would be important. The B-field may be tuned to the value $1/2$ which corresponds to the quiver locus. These models are very different from the IIb $SU(5)$ models that have been considered in the literature (see eg. [12] for a recent discussion and constructions), and more work needs to be done to connect the two pictures.

Now recall that up-type Yukawa couplings are perturbatively forbidden in the IIb theory, due to selection rules for the extra light gauge symmetry $U(1) \subset U(5)$. Thus we expect that in the resolved picture (by which we mean $\int_{\mathbf{P}^1} B \neq 0$), there will be a IIb description of the up-type Yukawa coupling using $D1$-instantons or worldsheet instantons (which are related by S-duality). Indeed if the $U(1)$-flux through the $\mathbf{P}^1$ is non-zero, then the action of such an instanton wrapping the $\mathbf{P}^1$ is not gauge invariant, as

$$\delta B_2 = \mathrm{Tr}(\lambda F). \tag{2.24}$$

This is just what one needs in order for an expression of the form $\mathbf{10} \cdot \mathbf{10} \cdot \mathbf{5}\, e^{-S}$ to be gauge invariant.

We may also ask what happens with the extra $U(1) \subset U(5)$ in the IIb model as we go to F-theory. In IIb its flux is related to the net amount of chiral matter. In F-theory, this $U(1)$ becomes part of the larger E_8 gauge symmetry, and is generically Higgsed by the adjoint field of the $8d$ gauge theory. In other words, it appears as a KK mode. In the IIb limit, the longitudinal part of this massive KK gauge boson should appear as a massless charged chiral field, and it seems plausible that is related to the charged massless modes of (B_2, C_2) that we saw appearing at a conifold point above. At any rate, the $U(1)$ symmetry is explicitly broken by the compactification, and it follows that upon integrating out the KK modes we should not expect the effective action in F-theory to respect the $U(1)$ selection rules appearing in IIb, just as we should not expect the effective action to satisfy E_8 selection rules.

Instead of trying to resolve the conifold points in IIb, one can also give the S^3 a finite size by deforming the branch locus to a generic section of $K_{B_3}^{-2}$. This is also compatible with the orientifold action and removes the conifold points. (Three-form fluxes through this S^3 are not compatible with the orientifold action and can not be turned on.) However this corresponds to breaking the $SU(5)$ GUT group by giving an expectation value to a field in the $\mathbf{10}$. So although one could get a smooth geometric background this way, it comes at the cost of breaking the GUT group.

It is amusing to ask what happens for local $SO(10)$ models when we take this limit. This corresponds to setting $b_5 \to 0$ identically in the above equations. Then the $O7$-plane is reducible and consists of a component wrapping S and a component wrapped on the curve $b_4 = 0$ in S and stretching in the normal direction. The spinors in the $\mathbf{16}$

live on the intersection of the non-compact orientifold plane with S and are partially made of non-perturbative (p, q) strings. The local equation of the Calabi–Yau threefold at these intersections is

$$\xi^2 = zw, \tag{2.25}$$

which means that they correspond to a curve of A_1 ALE singularities. Presumably again B_{NS} is zero here and they correspond to non-perturbative singularities of type IIb; indeed otherwise we would not expect massless modes of (p, q) strings here. Still this seems to be a very simple local model for producing spinor representations in the IIb language. The non-compact $D7$ brane intersects S along two curves, one of which is the curve above where the **16** lives, and the other is $b_3 = 0$ which is where the **10** of $SO(10)$ lives.

Finally we can ask what happens for E_6 models. This corresponds to setting both $b_5 \to 0$ and $b_4 \to 0$ identically in the above equations. Then b_2 vanishes identically so the limit we are trying to take does not correspond to a IIb limit (except for very special fibrations [28]).

2.3. Constraints from tadpole cancellation. From the local form of the singularity obtained above through the results of Tate's algorithm, we may immediately deduce the homology classes of the matter curves. Computing the discriminant of (2.2), one finds

$$\Delta = z^5 b_5^4(-b_0 b_5{}^2 + b_2 b_3 b_5 - b_4 b_3{}^2) + \mathcal{O}(z^6). \tag{2.26}$$

Thus the matter curves are given by

$$\Sigma_{\mathbf{10}} = \{b_5 = 0\}, \qquad \Sigma_{\mathbf{5}} = \{R = 0\}, \tag{2.27}$$

which yields the following homology classes:

$$[\Sigma_{\mathbf{10}}] = c_1 - t, \qquad [\Sigma_{\mathbf{5}}] = 8c_1 - 3t. \tag{2.28}$$

In particular it follows that

$$[\Sigma_{\mathbf{5}}] - 3[\Sigma_{\mathbf{10}}] - 5c_1 = 0. \tag{2.29}$$

Of course we also know the precise equation of the matter curves, but even these topological constraints are already quite restrictive. Mathematically, these are necessary conditions for the local geometry to be an elliptically fibered Calabi–Yau with section.

Although it is clear from our construction that these constraints have to be satisfied, it would be more satisfactory to give them a physical interpretation. In six dimensional compactifications of F-theory such constraints can be understood more physically as a consequence of anomaly cancellation [29]. For instance the relation (2.29) is then equivalent to cancellation of the $tr_f(F^4)$ anomaly. One expects such relations to hold also in more general F-theory settings [30]. We largely follow [29,30] in the remainder of this subsection.

Consider the worldvolume of a 7-brane S, intersecting another 7-brane S_a over a curve Σ_a. Under a gauge/Lorentz transformation, in the presence of (p, q) 7-branes we expect an additional contribution to the variation of the action given by

$$\delta_{\Lambda,\Theta} S \sim \int I^1_{adj,6}(\Lambda, \Theta) \wedge \delta^2(S) \wedge \delta^2(S) - \sum_{R_a} \int I^1_{R_a,6}(\Lambda, \Theta) \wedge \delta^4(\Sigma_a), \tag{2.30}$$

where Λ is a local gauge transformation and Θ is a local Lorentz transformation. Here $I^1_{R,6}$ is given through the descent procedure as

$$dI^1 = \delta I^0, \quad dI^0 = I_{R,8} = \Big[\mathbf{ch}_R(F) \wedge \hat{\mathbf{A}}(R)\Big]_8 \tag{2.31}$$

or more explicitly

$$\hat{I}_{R,8} = \frac{1}{24}\mathrm{Tr}_R(F^4) - \frac{1}{96}\mathrm{Tr}_R(F^2)\mathrm{Tr}(R^2) + \frac{rk}{128}\left(\frac{1}{45}\mathrm{Tr}(R^4) + \frac{1}{36}\mathrm{Tr}(R^2)^2\right), \tag{2.32}$$

where F is understood to be the gauge field on the gauge 7-brane wrapped on S, and $\hat{I} = (i(2\pi)^{d/2})I$. Further we have $\delta^2(S) \wedge \delta^2(S_a) = m_a\delta^4(\Sigma_a)$ and $\delta^2(S) \wedge \delta^2(S) = -c_1(S) \wedge \delta^2(S)$. Note that intersections are frequently not transverse in F-theory and $m_a \neq 1$. The above expression is the most straightforward generalization of the usual expression for D-branes [31,32]. There could be further contributions to δS in compact models, but here we will concentrate on the pieces that are associated to the gauge theory and have to be cancelled even in a local model.

In order to check anomaly cancellation we convert all the gauge traces to traces in the fundamental representation:

$$\mathrm{Tr}_R(F^4) = x_R\,\mathrm{Tr}_f(F^4) + y_R\,\mathrm{Tr}_f(F^2)^2, \qquad \mathrm{Tr}_R(F^2) = n_R\,\mathrm{Tr}_f(F^2). \tag{2.33}$$

In F-theory, the only massless tensor field available for the Green–Schwarz mechanism is the RR field C_4. Thus one would expect that the anomaly can be cancelled by mediation of C_4 if and only if the anomaly polynomial is factorizable, i.e. the matter representations occurring are such that

$$\hat{I}_{12} = \left[\sum_{0,a} n_a\delta^2(S_a) \wedge (2\mathrm{Tr}_f(F^2) - \frac{1}{2}\mathrm{Tr}(R^2))\right]^2. \tag{2.34}$$

The corresponding tadpole cancellation condition is the well-known constraint:

$$N_{D3} = \frac{\chi(Y_4)}{24} - \frac{1}{8\pi^2}\int_{Y_4} \mathsf{G} \wedge \mathsf{G}. \tag{2.35}$$

Since all three terms receive unknown contributions from infinity, we do not have to worry about this condition in a local model.

However this leaves a puzzle. The $\mathrm{Tr}_f(F^4)$ anomalies are non-zero and localized at different places in the internal space. So how do these pieces get cancelled exactly? There must be something mediating them. In perturbative type IIb, the $\mathrm{Tr}_f(F^4)$ and $\mathrm{Tr}(R^4)$ anomalies on branes are cancelled by mediation of the RR fields C_0/C_8. However in F-theory these fields are massive and do not appear as propagating fields in the effective action. Nevertheless it seems clear what must happen: in general F-theory compactifications integrating out the massive modes of the RR fields C_0 and C_8 leaves an effective interaction whose variation cancels the $\mathrm{Tr}_f(F^4)$ anomalies.

A similar issue in fact also arises in M-theory on G_2 manifolds and has been analyzed there [33] (see also [22] for a discussion). In the M-theory setting, chiral fermions are localized at points on the worldvolume of the gauge brane. In type IIa the corresponding anomalies would be mediated by the RR gauge field, but in M-theory this field is massive.

Nevertheless there is a residual interaction $\int K \wedge \omega^{(5)}$ which transforms under gauge transformations, and the Gauss law for $K \sim dA^{(1)}_{RR}$ is satisfied precisely when the $\mathrm{Tr}_f(F^3)$ anomalies are cancelled.

We have not precisely worked out the analogous statements in F-theory. The problem is that if we apply the analogous trick, rewriting $\int C_0 \wedge F^4 \sim -\int dC_0 \wedge \omega_7$, it does not yield an interaction that is invariant under $Sl(2, Z)$ transformations, so it is incomplete. However for our purposes we don't really need to work this out in detail, because we can use the IIb orientifold limit identified in Sect. 2.2 to show that the expected constraints have to be satisfied. In the IIb limit the anomaly is cancelled by C_0/C_8 exchange as usual, and we get the following modified Bianchi identity:

$$dF_1/2\pi = \sum_{D7} n_a \delta^2(S_a) - 8 \sum_{O7} \delta^2(O7). \tag{2.36}$$

Here we use the 'upstairs' picture, that is we write the relation on the covering space before taking the orientifold quotient. (F-theory corresponds more naturally to the 'downstairs' picture.)

Now the integral of dF_1 over any closed two-cycle is zero. Let us integrate over any curve Σ_b in S, and let us write (2.36) more suggestively as

$$dF_1/2\pi = 5\delta^2(S) + \delta^2(S_a) + 5\delta^2(S') - 8\delta^2(O7) + \text{other}, \tag{2.37}$$

where S' is the mirror of S under the orientifold action, the $O7$-plane is the one intersecting S over $\Sigma_{\mathbf{10}}$ (where it also intersects S'), and S_a is the part of the I_1 locus intersecting S over $\Sigma_{\mathbf{5}}$. Then we find

$$0 = -5c_1(S) \cdot \Sigma_b + \Sigma_{\mathbf{5}} \cdot \Sigma_b + (5-8)\Sigma_{\mathbf{10}} \cdot \Sigma_b, \tag{2.38}$$

or equivalently

$$[\Sigma_{\mathbf{5}}] - 3[\Sigma_{\mathbf{10}}] - 5c_1 = 0 \tag{2.39}$$

in $H_2(S, \mathbf{Z})$, which is what we wanted to show. More generally we expect the relation

$$\sum_{R_a} x_{R_a}[\Sigma_a] - \frac{1}{2} x_{\mathrm{adj}}\, c_1(S) = 0 \tag{2.40}$$

to be equivalent to cancelling the $\mathrm{Tr}_f(F^4)$ anomalies, but we have not been able to show this in full generality. As a special case, in six-dimensional compactifications of F-theory the above homology classes are all proportional to the class of a point, and this relation was verified in [29].

Following [30], we may get a second constraint by using a further relation in F-theory models:

$$\Delta = -12K_{B_3}. \tag{2.41}$$

This is also a kind of 7-brane tadpole cancellation (eg. on $K3$ it restricts the total number of 7-branes to be 24), but it differs from (2.36). Since we have an $SU(5)$ singularity along S, we may write

$$\Delta = 5[S] + \Delta'. \tag{2.42}$$

If we assume there are only matter curves for hypermultiplets in the **5** or **10**, as is generically the case, then by intersecting with S we obtain

$$-5t + 4\Sigma_{\mathbf{10}} + \Sigma_{\mathbf{5}} = -12K_{B_3}|_S. \tag{2.43}$$

Here we used $[S]\cdot[S] = c_1(NS)|_S = -t$. The intersection multiplicities can be read from the explicit form of the discriminant (2.26) (the coefficient of $[\Sigma_{\mathbf{10}}]$ can presumably be understood from the fact that the charge of an orientifold plane is -4 in the 'downstairs' picture). Further applying the adjunction formula $K_{B_3}|_S = K_S + t$, we find that

$$7t + 4\Sigma_{\mathbf{10}} + \Sigma_{\mathbf{5}} = -12K_S. \tag{2.44}$$

Together with the earlier constraint (2.39), it then follows that the homology classes of the matter curves are given by

$$[\Sigma_{\mathbf{10}}] = c_1(S) - t, \qquad [\Sigma_{\mathbf{5}}] = 8c_1 - 3t, \tag{2.45}$$

exactly as promised.

Above we discussed anomalies of the higher dimensional gauge theory, even though the higher dimensional gauge symmetry is broken through compactification. But the KK modes still transform under the higher dimensional gauge symmetry, albeit non-linearly. So the sum over KK modes remembers the higher dimensional anomalies, and does not makes sense unless those anomalies are cancelled.

2.4. Higgs bundles, spectral covers and ALE-fibrations. As advertized in (1.1), we claim there are several equivalent descriptions of the supersymmetric configurations of an $8d$ gauge theory. We may describe such a configuration as an ALE fibration, which is how it arises in F-theory in 'closed string' variables. However we may also think of it more intrinsically in terms of field configurations of the adjoint scalars and gauge field. This gives us the Higgs bundle picture. Finally we may replace the Higgs field by its eigenvalues and the bundle by the corresponding eigenvectors. This gives us the spectral cover picture, essentially a fibered weight diagram with one sheet for each weight of a representation. The latter yields conventional B-branes in an auxiliary non-compact Calabi–Yau threefold X. The description of B-branes in a Calabi–Yau is already a well-developed subject and so this picture is the most convenient for doing actual constructions and calculations. In this section, we spell out the spectral cover description and its relation to the other pictures in a bit more detail.

Much of the structure discussed here has been discussed in the heterotic setting, but the main point is that it is in fact *intrinsic* to the $8d$ supersymmetric Yang–Mills theory and therefore applies to an arbitrary local F-theory geometry, or any other UV completion of $8d$ Yang–Mills theory. Moreover the spectral cover description allows us to tie up some technical loose ends from our previous papers. A completely analogous construction can be made in $7d$ supersymmetric Yang–Mills theory [22] and leads to the construction of local models in M-theory, in the large volume limit where the Yang-Mills theory gives an accurate description. One can also apply the dictionary for ALE fibrations over a Riemann surface. This is essentially classic geometric engineering.

2.4.1. The dictionary. Let us start with the Higgs bundle picture. The conditions for supersymmetry in the $8d$ gauge theory are obtained by dimensional reduction. Namely we start with the Hermitian–Yang–Mills equations in $10d$, and assume fields are invariant under translation along a complex line. Then we can write the gauge field as

$$\mathsf{A}^{0,1} = A_{\bar{1}}(z^1, z^2)d\bar{z}^1 + A_{\bar{2}}(z^1, z^2)d\bar{z}^2 + \Phi_{\bar{3}}(z^1, z^2)d\bar{z}^3. \tag{2.46}$$

The gauge field $A^{0,1}$ is the (0, 1) component of a section of $\mathcal{A}^1(S, \mathrm{ad}(\mathcal{G}))$, where $\mathcal{G}$ is a principal G-bundle on S. We can think of $\Phi_{\bar{3}}$ as a (2, 0) form on S valued in $\mathrm{ad}(\mathcal{G})$, as it transforms in the same way under coordinate transformations; in other words, it is a section of $\mathrm{ad}(\mathcal{G}) \otimes K_S$.

The F-terms are given by

$$\mathsf{F}^{0,2} = 0 \quad \Rightarrow \quad F^{0,2} = 0, \qquad \bar{\partial}_A \Phi = 0, \tag{2.47}$$

and the D-terms are

$$\mathsf{g}^{i\bar{j}}\mathsf{F}_{i\bar{j}} = 0 \quad \Rightarrow \quad g^{i\bar{j}}F_{i\bar{j}} + g^{3\bar{3}}[\Phi_{\bar{3}}^\dagger, \Phi_{\bar{3}}] = 0. \tag{2.48}$$

Here $g_{i\bar{j}}$ is taken to be the Kähler metric on S, and $g_{3\bar{3}} = \det(g_{i\bar{j}})^{-1}$. By taking the Hodge star on S, we can also rewrite Eq. (2.48) as $J \wedge F + [\Phi^\dagger, \Phi] = 0$, where $\Phi_{ij} = \epsilon_{ij3} g^{3\bar{3}} \Phi_{\bar{3}}$. The dependence of the second term in this equation on $g_{3\bar{3}}$ actually cancels, and it is more convenient to only refer to the two-form Φ_{ij}, which as noted above is a section of $\mathrm{ad}(\mathcal{G}) \otimes K_S$.

The F- and D-term equations are called Hitchin's equations or the Yang–Mills–Higgs equations, and the data (A, Φ) subject to these equations defines a Higgs bundle [34,35]. The D-term is the moment map for gauge transformations acting on the pair (A, Φ) with respect to the Kähler form associated to the metric

$$g(\delta\mathsf{A}, \delta\mathsf{A}) = -\frac{i}{2\pi}\int_S J \wedge \mathrm{Tr}(\delta A^{0,1} \wedge (\delta A^{0,1})^\dagger) + \mathrm{Tr}(\delta\Phi \wedge \delta\Phi^\dagger), \tag{2.49}$$

where $J = \frac{1}{2}\sqrt{-1} g_{i\bar{j}} dz^i \wedge d\bar{z}^j$. The F-terms are critical points of a holomorphic Chern-Simons functional, which is informally written as:

$$W = \frac{1}{8\pi^2}\int_S \mathrm{Tr}\,(A + \Phi)\bar{\partial}(A + \Phi) + \frac{2}{3}(A + \Phi)^3 = \frac{1}{4\pi^2}\int_S \mathrm{Tr}\,\Phi \wedge F_A. \tag{2.50}$$

In other words, the F-terms are governed by a holomorphic version of a BF-theory. For a more precise discussion, see [23].

The Higgs bundle is the object of primary interest, because it describes supersymmetric solutions of the equations of motion of the weakly coupled $8d$ field theory. However it is not easy to handle directly, as in the non-abelian case the D-term equation is a complicated non-linear PDE, which we cannot solve in closed form. This leads us to the spectral cover picture, which attempts to give an abelianized description of the non-abelian Higgs bundle.

For convenience we temporarily focus on $U(n)$ and $SU(n)$ gauge groups, though analogous constructions exist for any gauge group. We let X denote the total space of the canonical bundle K_S, and we let s denote a coordinate on the fibers. Then the spectral sheaf $\mathcal{L}$ for a pair (A, Φ) is obtained as follows. We think of Φ as a map

$$p^*\Phi : p^*E \;\rightarrow\; p^*(E \otimes K_S), \tag{2.51}$$

where p is the projection $X \rightarrow S$. The spectral sheaf $\mathcal{L}$ is now defined as the cokernel of $\Psi = p^*\Phi - sI$, where I is the $n \times n$ identity matrix. In other words, $\mathcal{L}$ is defined through the short exact sequence

$$0 \;\rightarrow\; p^*E \;\xrightarrow{\Psi}\; p^*(E \otimes K_S) \;\rightarrow\; \mathcal{L} \;\rightarrow\; 0. \tag{2.52}$$

Sometimes slightly different definitions are used, for example the cokernel is sometimes defined as $\mathcal{L} \otimes p^* K_S$.[2]

Let us slowly unravel what this means. First of all, the spectral sheaf $\mathcal{L}$ is supported on a divisor C in X, which is called the spectral cover. Its equation is given by

$$\det \Psi = \det(sI - p^*\Phi) = 0. \tag{2.53}$$

By expanding in s, the polynomial coefficients are identified with invariant polynomials of Φ, the Casimirs. The map that sends the Higgs field Φ to its Casimirs is the Hitchin map.

For a generic point on S the roots λ_i of this polynomial give us n points on the fiber of K_S. The λ_i are interpreted as the eigenvalues of $p^*\Phi$, and by a complexified gauge transformation we can put $p^*\Phi$ in diagonal form:

$$g^{-1}(p^*\Phi)g \ \sim \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}. \tag{2.54}$$

Thus spectral cover C gives an n-fold cover over S which geometrizes how the λ_i vary over S. This is the spectral cover for the fundamental representation of $SU(n)$. Since we will be interested in non-compact covers, we should allow simple poles for the Higgs fields. We can get rid of the poles in (2.53) by multiplying with a suitable section. Thus instead of (2.53) we will write the degree n equation

$$0 = b_0 s^n + b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n. \tag{2.55}$$

Since $s = 0$ is marked, the only coordinate transformations allowed are rescaling. For $SU(n)$ gauge groups we further want to impose that all the roots add up to zero. Since we have

$$\lambda_1 + \cdots + \lambda_n = b_1, \tag{2.56}$$

therefore we set $b_1 = 0$. The surface C is non-compact. Along the locus $b_0 = 0$, two of the roots go off to infinity. Let us denote the divisor $b_0 = 0$ on S by η. Since s is a coordinate on K_S, the b_i are then seen to be sections of

$$b_i \sim \eta - i\, c_1(S). \tag{2.57}$$

We further have to describe the gauge field A in this picture. More precisely we have to describe how the Dolbeault operator on E up to complexified gauge transformations, or equivalently the holomorphic structure of the bundle E (since we only get an isomorphism of the holomorphic data), is encoded in the spectral cover picture. From our short exact sequence (2.52), we see that the fibers of $\mathcal{L}$ are identified with the eigenvectors of $p^*\Phi$. Generically the eigenvalues of $p^*\Phi$ are distinct, and so each fiber of E can be decomposed into one-dimensional eigenspaces $\oplus_i \mathbf{C}\,|i\rangle$. Let us denote coordinates on the total space $K_S \to S$ by pairs (p, s), where $p \in S$ and s is the coordinate on the fiber. The assignment

$$(p, \lambda_i) \to \mathbf{C}\,|i\rangle \tag{2.58}$$

[2] Another frequently used definition gives the spectral line bundle L, rather than the spectral sheaf $\mathcal{L}$. Let us denote by C the spectral cover, $R = K_{C/S}$ the ramification divisor, and $s \in H^0(C, p_C^* K_S)$ the tautological eigenvalue section, whose value at a point (p, s) is given by s. Then $L \otimes \mathcal{O}(-R)$ is the kernel of $p_C^*\Phi - sI : p_C^* E \to p_C^* E \otimes p_C^* K_S$.

yields a line bundle L on C called the spectral line bundle. Furthermore since $\bar{\partial}_A \Phi = 0$, $\bar{\partial}_A$ commutes with the action of Φ on E, and we can simultaneously diagonalize $\bar{\partial}_A$. Thus we get a Dolbeault operator on L, i.e. L is a holomorphic line bundle. The local sections of L which are annihilated by this Dolbeault operator define the sheaf $\mathcal{L}$.

Over a sublocus on the base, some of the eigenvalues of Φ will coincide, and we need to worry about the Jordan block structure. We can appeal to our general definition (2.52), and show that by picking the regular representative of Φ, we get a smooth global object. The spectral cover and line bundle in K_S satisfy the usual requirements of a B-brane in the large volume limit: a holomorphic cycle with a holomorphic bundle on it.

Conversely, given a spectral cover and a spectral line bundle, we may recover the Higgs field Φ and the bundle E. Namely given the spectral data (C, L), we may recover the Higgs bundle as:

$$E = p_{C*}L, \qquad \Phi = p_{C*}s. \tag{2.59}$$

Furthermore if we want an $SU(n)$ bundle rather than a $U(n)$ bundle, then we also need to require

$$\det(p_{C*}L) = \mathcal{O}, \tag{2.60}$$

where $\mathcal{O}$ is the trivial line bundle. This gives a topological constraint on the allowed spectral line bundles.

Now we discuss the ALE picture. The second homology group of an ALE space of type ADE is isomorphic to the corresponding root lattice of type ADE. We want to fiber this over a complex surface S. Locally on S we may choose a basis α_i of $H_2(\mathrm{ALE}, \mathbf{Z})$ corresponding to the fundamental roots of the corresponding ADE Lie algebra (obviously this depends on a choice of Weyl chamber). Similarly we may choose a dual basis ω^j of $H^2(\mathrm{ALE}, \mathbf{Z})$ satisfying

$$\int_{\alpha_i} \omega^j = \delta^{ij}. \tag{2.61}$$

In our local patch, the holomorphic volume form can now be expanded as

$$\Omega^{4,0} = \sum_j \Phi_j^{2,0} \wedge \omega^j, \tag{2.62}$$

where $\Phi_j^{2,0} = \int_{\alpha_j} \Omega^{4,0}$. Obviously, we want to interpret the $\Phi_j^{2,0}$ as the components of Φ proportional to the Cartan generators. Similarly, we expand the three-form field as

$$\mathsf{C}_3 = \sum_j A_j \wedge \omega^j \tag{2.63}$$

and we want to interpret the A_j as the Cartan components of the gauge field. Globally, our local patches should be glued using the natural large diffeomorphism symmetries of the ALE (which are in one-to-one correspondence with elements of the Weyl group).

We may go back and forth between the spectral cover and the ALE fibration. For $SU(n)$ gauge groups the A_{n-1}-ALE fibration is defined by the following equation

$$y^2 = x^2 + b_0 s^n + b_2 s^{n-2} + \cdots + b_n. \tag{2.64}$$

As far as the variation of Hodge structure is concerned, the quadratic terms x^2 and y^2 are irrelevant and may be dropped, recovering our previous equation. This argument is well known from Landau–Ginzburg models, where we 'integrate out' the fields with a quadratic potential. For other groups there is an analogous but slightly more complicated relation.

The effective four-dimensional theory describes the deformations of the holomorphic data. We can give a precise map between the spectral cover and the ALE fibration pictures of such variations using the notion of cylinder mappings. Let us think of (2.64) as a conic bundle fibered over the complex plane parametrized by s. We have a map

$$p_R : R \to C, \tag{2.65}$$

where R is obtained from C by attaching a line (with equation $y = x$) to each point in the fiber of the covering $C \to S$. The variety R is called the cylinder. (More precisely, the cylinder consists of the pair of lines fibered over the discriminant locus. This leads to a relation between the Hodge structure of Y_4 and a double cover of the discriminant, branched over the singular locus of the discriminant. In the A_{n-1} case there is no monodromy, so the cylinder splits in two and it suffices to consider half the cylinder. In the D_n case we need both the lines). We also have a map

$$i : R \to Y_4 \tag{2.66}$$

which embeds these lines in the ALE (2.64), each line sitting at the corresponding point $s = \lambda_i$ in the s-plane. Then we get a map

$$i_* p_R^* : \; H^{i,j}(C) \;\to\; H^{i+1,j+1}(Y_4). \tag{2.67}$$

It gives an explicit isomorphism of the Hodge structures appearing on both sides. To define this properly, one should consider certain compactified version, and sometimes one needs a small correction to subtract a singlet of the Weyl group. We will not discuss this explicitly here, see Appendix C of [1] and [21] for further discussion.

Now we see that deformations of the spectral cover, which live in $H^{2,0}(C)$, get mapped to generators of $H^{3,1}(Y_4)$, which correspond to complex structure deformations of the ALE fibration. Furthermore in terms of the ALE fibration Y_4, the spectral line bundle is encoded as G-flux. Let us decompose the flux of the spectral line bundle as

$$c_1(L) = \frac{1}{2} c_1(K_{C/S}) + \gamma, \tag{2.68}$$

where $K_{C/S} = K_C - p_C^* K_S$ is the ramification divisor, and $p_{C*}\gamma = 0$. Then the spectral line bundle and the G-flux are related by

$$\mathsf{G} = i_* p_R^* \gamma - q\,[\text{ALE}] \;\in \mathrm{H}^{2,2}(\mathrm{Y}_4), \tag{2.69}$$

where [ALE] is the class of an ALE fiber, and q is determined by requiring that $\int_S G = 0$, which gives $q = \gamma \cdot_C \Sigma_E$. Given this explicit expression it is not too hard to check that such fluxes are primitive, i.e. satisfy $J \wedge G = 0$ on Y_4, if $p_{C*}\gamma = 0$. For $U(n)$ bundles, we need to make sure that if J contains a piece $\pi^* J_S$ pulled-back from S, then $p_{C*}\gamma \cdot J_S = 0$.

We may state the correspondence somewhat more intrinsically as follows. We have a local system $\pi : \Lambda_{\text{ADE}} \to S$ over S, whose fiber $\pi^{-1}(p)$ over a point $p \in S$ is given by an ADE root lattice. We also consider the complexification $\Lambda \otimes \mathcal{O}_S$. Then we are interested in the cohomology groups $H^k(S, \Lambda)$ and $H^{i,j}(S, \Lambda \otimes \mathcal{O}_S)$. They are acted on by the Weyl group and decompose into various pieces. The cohomology groups encountered above in the spectral cover or ALE picture all correspond to specific pieces of this decomposition, and do not care how the ADE lattice is 'realized.'

2.4.2. Other associated spectral covers. The $SU(n)$ spectral cover we have considered so far should really be called C_E, to indicate that it corresponds to the fundamental representation. We can also construct spectral covers for other representations, which typically describe equivalent data. One important cover that we will need is the spectral cover $C_{\Lambda^2 E}$ for the anti-symmetric representation of $SU(n)$. This has $\frac{1}{2}n(n-1)$ sheets. Each sheet intersects a fiber of K_S in the points

$$\Lambda^2 E: \qquad \lambda_i + \lambda_j, \quad i < j, \tag{2.70}$$

where addition is defined in the obvious way in each fiber. In fact it is not hard to write down an explicit equation using Mathematica. For the case $n = 5$, the cover is defined by the degree 10 equation,

$$\begin{aligned} 0 = {} & s^{10} + 3s^8 c_2 - s^7 c_3 + s^6(3c_2^2 - 3c_4) + s^5(-2c_2c_3 + 11c_5) + s^4(c_2^3 - c_3^2 - 2c_2c_4) \\ & + s^3(-c_2^2 c_3 + 4c_3c_4 + 4c_2c_5) + s^2(-c_2c_3^2 + c_2^2 c_4 - 4c_4^2 + 7c_3c_5) \\ & + s(c_3^3 + c_2^2 c_5 - 4c_4c_5) - c_3^2 c_4 + c_2c_3c_5 - c_5^2, \end{aligned} \tag{2.71}$$

where $c_i = b_i/b_0$ and the whole equation should be multiplied with b_0^3 in order to remove the denominators. We denote the intersection of $C_{\Lambda^2 E}$ with the zero section $s = 0$ by $\Sigma_{\Lambda^2 E}$.[3] The surface $C_{\Lambda^2 E}$ is singular when two of the eigenvalues coincide, i.e. $\lambda_i + \lambda_j = \lambda_k + \lambda_l$ for some i, j, k, l. This happens in codimension one, so the matter curve $\Sigma_{\Lambda^2 E}$ is also singular at isolated points. The spectral line bundle on this cover is given fiberwise by

$$L_{\Lambda^2 E}: \qquad (p, \lambda_i + \lambda_j) \to \mathbf{C}\,|i\rangle \wedge |j\rangle\,. \tag{2.72}$$

It is not really a line bundle but a (torsion-free) sheaf, its rank jumping up at the singular locus, and one has to desingularize in order to define things unambiguously. Still this data is determined uniquely by the spectral line bundle for the cover of the fundamental representation, as follows.

In order to write an unambiguous formula it is more natural to think about unembedded covers [36]. We take pairs of points $(q_1, q_2) \in C_E \times_S C_E$, and remove the diagonal where $q_1 = q_2$. Then we define the quotient[4]

$$\tilde{C}_{\Lambda^2 E} = \{(q_1, q_2) \in C_E \times_S C_E \mid q_1 \neq q_2\}/\mathbf{Z}_2, \tag{2.73}$$

where the $\mathbf{Z}_2$ action interchanges $(q_1, q_2) \to (q_2, q_1)$. This cover is embedded in $X \times_S X/\mathbf{Z}_2$, but not in X, and provides a resolution of $C_{\Lambda^2 E}$. There is a natural map

$$C_E \times_S C_E - \mathrm{diag}(C_E) \quad \to \quad \tilde{C}_{\Lambda^2 E} \quad \to \quad C_{\Lambda^2 E}. \tag{2.74}$$

The last map is given fiberwise by sending $(\lambda_i, \lambda_j) \to \lambda_i + \lambda_j$. The pairs (λ_i, λ_j) and (λ_k, λ_l) are distinct in $\tilde{C}_{\Lambda^2 E}$ even when $\lambda_i + \lambda_j = \lambda_k + \lambda_l$ in $C_{\Lambda^2 E}$. The inverse image

[3] Note that the subscript here indicates the representation of the holonomy group, not the unbroken gauge group. In our discussion later however we will instead use the subscript to denote the representation under the GUT group, as in our previous papers. Thus in our $SU(5)$ examples later we will have $\Sigma_{\Lambda^2 E} = \Sigma_{\mathbf{5}}$ and $\Sigma_E = \Sigma_{\mathbf{10}}$.

[4] Strictly we have to take the closure and then take the quotient. We oversimplified this issue here and in the remainder in order to avoid too much notation.

of $\Sigma_{\Lambda^2 E}$ in $\tilde{C}_{\Lambda^2 E}$ is its normalization $\tilde{\Sigma}_{\Lambda^2 E}$. The spectral line bundle L_E on C_E gets mapped to a smooth line bundle on $\tilde{C}_{\Lambda^2 E}$:

$$L_E \boxtimes L_E \quad \to \quad \tilde{L}_{\Lambda^2 E} \quad \to \quad L_{\Lambda^2 E}. \tag{2.75}$$

It only gets mapped to a sheaf $L_{\Lambda^2 E}$ on $C_{\Lambda^2 E}$ because the map $\tilde{C}_{\Lambda^2 E} \to C_{\Lambda^2 E}$ is two-to-one at the singular locus, but this is irrelevant since we should work with the non-singular surface $\tilde{C}_{\Lambda^2 E}$. This construction should be interpreted as follows. The spectral line bundle on $C_{\Lambda^2 E}$ is the set of eigenlines $|i\rangle \wedge |j\rangle$ of $\Lambda^2 E$ under the action of the Higgs field. When $\lambda_i + \lambda_j = \lambda_k + \lambda_l$ the cover $C_{\Lambda^2 E}$ is singular, so there is an ambiguity in assigning eigenlines of $\Lambda^2 E$ to eigenvalues of $\Phi_{\Lambda^2 E}$ in a neighbourhood of the singular locus. This ambiguity is naturally resolved by recalling that the assignment of eigenlines to eigenvalues was unambiguous for E (assuming C_E is smooth), in other words it is naturally resolved by requiring that $L_{\Lambda^2 E}$ descends from a smooth line bundle on $\tilde{C}_{\Lambda^2 E}$. As emphasized in [3], this means that keeping track of the gauge indices implies that the hypermultiplet at the intersection really couples to $\tilde{L}_{\Lambda^2 E}$. Thus the hypermultiplet propagates on the normalized matter curve $\tilde{\Sigma}_{\Lambda^2 E}$ rather than on $\Sigma_{\Lambda^2 E}$ itself.

Similarly we may construct spectral covers for other representations. For instance the spectral cover for the symmetric representation $C_{S^2 E}$ is given fiberwise by

$$S^2 E: \qquad \lambda_i + \lambda_j, \quad i \leq j. \tag{2.76}$$

We will not have any need for these other coverings in this paper.

2.4.3. Fermion zero modes. Now that we have a description of configurations in the $8d$ gauge theory in terms of holomorphic cycles and bundles on them, we would like to describe the zero modes of the Dirac operator. In holomorphic geometry the Dirac operator splits into a Dolbeault operator

$$\bar{D} = \bar{\partial} + A^{0,1} + \Phi^{2,0} \tag{2.77}$$

and its adjoint $\bar{D}^\dagger$.

The spinor configuration space together with the $\bar{D}$ operator yield a complex, and we may consider its cohomology. In fact we may interpret this as the cohomology of a double complex $(\Omega^\bullet_S(\mathrm{ad}(E)) \otimes \Lambda^\bullet K_S, \bar{\partial}_A, \Phi)$, since $\bar{\partial}_A^2 = \Phi \wedge \Phi = [\bar{\partial}_A, \Phi] = 0$ by the F-term equations. These cohomology groups are thus usually referred to as the hypercohomology groups of the Higgs bundle, and denoted by

$$\mathbb{H}^p(\mathcal{E}^\bullet), \tag{2.78}$$

where $\mathcal{E}^\bullet$ is the two step complex $\mathrm{ad}(E) \to \mathrm{ad}(E) \otimes K_S$. Since the operator $\not{D} = \bar{D} + \bar{D}^\dagger$ is elliptic, by arguments familiar from Hodge theory its zero modes are in one-to-one correspondence with the generators of these hypercohomology groups. We might call them the 'harmonic' representatives.

As usual, the index p correlates with the $4d$ chirality as $(-1)^p$. For $p = 1, 2$ the $4d$ part of the wave function belongs to a chiral (anti-chiral) superfield, and for $p = 0, 3$ we get four-dimensional gauginos (or possibly ghosts if suitable stability conditions are not satisfied).

Thus to find the spectrum and interactions, we are interested in computing $\mathbb{H}^p(\mathcal{E}^\bullet)$. By the usual arguments of deformation theory, they describe the symmetries, the tangent space to the deformation space, and the obstructions. For example, using the long exact sequence for hypercohomology, we find the exact sequence

$$\cdots \to H^0(\mathrm{ad}(E) \otimes K_S) \to \mathbb{H}^1(\mathcal{E}^\bullet) \to H^1(\mathrm{ad}(E)) \to \cdots . \tag{2.79}$$

Here $H^0(\mathrm{ad}(E) \otimes K_S)$ describes deformations of the Higgs field Φ, and $H^1(\mathrm{ad}(E))$ describes deformations of the bundle E.

Note that $\not{D}$ and its zero modes depend explicitly on the hermitian metric on E solving the D-term equations. Therefore, like the hermitian metric, the harmonic representatives are impossible to write down exactly in closed form. To find the spectrum, it will be crucial to use the fact that unlike the harmonic representatives, the hypercohomology groups do not depend on the choice of hermitian metric.

Now let us consider the spectral picture. Recall that we have a short exact sequence relating the Higgs bundle to the spectral sheaf,

$$0 \to p^*E \longrightarrow p^*(E \otimes K_S) \to \mathcal{L} \to 0. \tag{2.80}$$

One may use a spectral sequence argument to show that the hypercohomology groups of the Higgs bundle are isomorphic to Ext groups of the spectral sheaf $\mathcal{L}$, i.e.

$$\mathbb{H}^p(\mathcal{E}^\bullet) \simeq \mathrm{Ext}^p_X(\mathcal{L}, \mathcal{L}) \tag{2.81}$$

again still assuming $\mathrm{GL}(n, \mathbf{C})$ Higgs bundles. This is exactly as it should be, because the hypercohomology groups classify symmetries, deformations and obstructions of the Higgs bundle, whereas Ext groups do the same for coherent sheaves. So if the two pictures are equivalent, these had better match.

Like the hypercohomology groups of the Higgs bundle, the Ext-groups of the spectral sheaf naturally give a unified description of the spectrum for all possible configurations. However for explicit computations, it is convenient to go one step further. In applications to model building, the spectral sheaf $\mathcal{L}$ often decomposes into multiple pieces. Let us denote by i, j the embedding of two such components D_1, D_2 into X, and assume that $\mathcal{L} = i_*L_1 \oplus j_*L_2$. Then $\mathrm{Ext}^p_X(\mathcal{L}, \mathcal{L})$ decomposes into $\mathrm{Ext}^p_X(i_*L_1, j_*L_2)$, $\mathrm{Ext}^p_X(i_*L_1, i_*L_1)$, et cetera. It is not hard to show that $\mathrm{Ext}^p(A, B)$ can be localized on the intersection of the supports of A and B.

Now we can further simplify by relating these Ext-groups to various Dolbeault cohomology groups on the intersection of the supports. The idea is to use the short exact sequence $0 \to \mathcal{O}_X(-D) \to \mathcal{O}_X \to \mathcal{O}_D \to 0$ in order to resolve $\mathcal{O}_D$. Assuming that $\Sigma = D_1 \cap D_2$ is a curve and L_1 is a line bundle, one can then show that

$$\mathrm{Ext}^p_X(i_*L_1, j_*L_2) \simeq H^{p-1}(\Sigma, L_1^\vee \otimes L_2 \otimes K_{D_1}|_\Sigma). \tag{2.82}$$

In applications to model building, one would typically take $i : S \to X$ to be the zero section, $i_*L_1 = i_*\mathcal{O}_S$ to be the 'gauge brane', and j_*L_2 to be a 'flavour brane.'

Instead of intersecting branes, we could also consider coincident branes, which corresponds to taking $D_1 = D_2$. Then one finds that

$$\mathrm{Ext}^p_X(i_*L_1, i_*L_2) \simeq H^{p-1}(D_1, L_1^\vee \otimes L_2 \otimes K_{D_1}) \oplus H^p(D_1, L_1^\vee \otimes L_2). \tag{2.83}$$

More precisely, there is a spectral sequence relating the right to the left and possibly killing some generators on the right, as was actually noted in [1]. At any rate, we see that this agrees with the spectrum that was deduced in [1,2], which relied on writing down approximate zero modes in the Higgs bundle picture. The argument above is perhaps less intuitive, but the pay-off is that it is more rigorous and can be applied to arbitrary Higgs field configurations, including for example the more subtle configurations discussed in [37].

By further specializing the above expressions, we see that the number of moduli of a smooth spectral cover (C_E, L_E) is given by

$$N_{\text{mod}} = \dim \text{Ext}^1_X(j_* L_E, j_* L_E) \;\simeq\; h^{2,0}(C_E) \oplus h^{0,1}(C_E). \tag{2.84}$$

The number of moduli is independent of the spectral line bundle on C, which just cancels in the formula. Applying this instead to the gauge brane corresponds to computing the number of adjoint fields of the unbroken gauge group in the compactified theory:

$$N_{\text{adj}} = \dim \text{Ext}^1_X(i_* \mathcal{O}_S, i_* \mathcal{O}_S) \;\simeq\; h^{2,0}(S) \oplus h^{0,1}(S). \tag{2.85}$$

We can also apply these methods to spectral covers that are not smooth, such as the cover $C_{\Lambda^2 E}$ encountered in Sect. 2.4.2, although some more care has to be taken. Thinking of $i_* \mathcal{O}_S$ as the gauge brane and $C_{\Lambda^2 E}$ as the flavour brane, the formula for the amount of chiral matter on $\Sigma_{\Lambda^2 E} = i(S) \cap C_{\Lambda^2 E}$ is given by

$$\text{Ext}^1_X(i_* \mathcal{O}_S, j_* L_{\Lambda^2 E}) \;\simeq\; H^0(\Sigma_{\Lambda^2 E}, L_{\Lambda^2 E} \otimes K_S|_{\Sigma_{\Lambda^2 E}}). \tag{2.86}$$

The right-hand side is a sheaf cohomology group but not a Dolbeault cohomology group, as $L_{\Lambda^2 E}$ is not a smooth line bundle. We can further relate it to a Dolbeault cohomology group by lifting to the line bundle $\tilde{L}_{\Lambda^2 E}$ on the normalization of $\Sigma_{\Lambda^2 E}$, which was introduced in Sect. 2.4.2. Geometrically this replaces each of the double points of $\Sigma_{\Lambda^2 E}$ by two distinct points. Using the Leray sequence for the normalization $\nu : \tilde{\Sigma}_{\Lambda^2 E} \to \Sigma_{\Lambda^2 E}$, one easily shows that the above is also equal to $H^0(\tilde{\Sigma}_{\Lambda^2 E}, \tilde{L}_{\Lambda^2 E} \otimes \nu^* K_S|_{\Sigma_{\Lambda^2 E}})$. Since $\tilde{L}_{\Lambda^2 E}$ is actually a smooth line bundle, this can now be interpreted as a Dolbeault cohomology group, and recovers the answer found in [3]. (For earlier work, see [38,39]). As noted previously it has a nice physical interpretation: the hypermultiplet behaves as if it propagates on the normalization $\tilde{\Sigma}_{\Lambda^2 E}$, rather than on the singular matter curve $\Sigma_{\Lambda^2 E}$.

Apart from the matter content, the spectral cover description also allows us to give a precise mathematical definition of the classical Yukawa couplings (and higher dimension couplings as well), at least up to field redefinitions. The basic point is that the Yoneda product $\text{Ext}^1(i_{1*}L_1, i_{2*}L_2) \times \text{Ext}^1(i_{2*}L_2, i_{3*}L_3) \to \text{Ext}^2(i_{1*}L_1, i_{3*}L_3)$ yields an obstruction class, but under the trace map we have that $\text{Ext}^2(i_{1*}L_1, i_{3*}L_3)$ is dual to $\text{Ext}^1(i_{3*}L_3, i_{1*}L_1)$, which represents matter fields. Thus the Yukawa couplings are given by a triple Yoneda product followed by a trace map:

$$\text{Ext}^1_X(i_{1*}L_1, i_{2*}L_2) \times \text{Ext}^1_X(i_{2*}L_2, i_{3*}L_3) \times \text{Ext}^1_X(i_{3*}L_3, i_{1*}L_1) \;\to\; \mathbf{C}. \tag{2.87}$$

Again this expression summarizes all the possibilities, with wave functions either localized in the bulk or on 7-brane intersections, and again it manifestly only depends on holomorphic data. One should be careful about drawing conclusions from such computations however. The usual warnings about the relation with the physical Yukawa couplings (which depend on the Kähler potential and may receive loop corrections) apply. By considering Yoneda products on Ext^p for other values of p, one can also deduce the algebra of the unbroken gauge group and the charges of the matter fields.

To complete the picture, we should further ask how to compute the spectrum and interactions in the ALE fibration picture. This is a bit less elegant. We have already seen that part of the spectrum comes from deformations of the metric and the tensor field, but

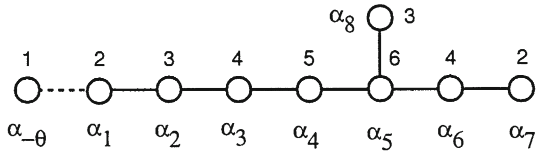

Fig. 1. The extended E_8 Dynkin diagram and Dynkin indices

in addition one gets degrees of freedom from wrapped $M2$-branes. To treat this properly, one first has to resolve the singularities of the ALE fibration, preserving the Calabi-Yau condition. In the M-theory picture, this corresponds to going out on a new branch of the $3d$ theory, the Coulomb branch. In the limit that the exceptional cycles are large, one can treat $M2$-branes wrapped on such cycles as solitons and quantize them as such. In this way one can recover the remaining Dolbeault cohomology groups above. However we do not know how to formulate this in the clean universal form we saw for Higgs bundles and spectral covers above, since the cohomology groups obtained from quantizing $M2$-branes correspond to massive modes in this description, so the deformation theoretic arguments don't apply. Given that the regime of validity and the physics of the M-theory soliton approach is really quite different, one may doubt that such a universal formulation is actually possible.

2.4.4. E_8 Higgs bundles. Now we return to the case of primary interest. In local F-theory models we are dealing with fibrations by E_8 ALE spaces, or equivalently with E_8 Higgs bundles. The relevant spectral cover is the one for the adjoint representation, which we will simply call 'the' spectral cover. The adjoint representation is 248 dimensional, of which eight are Cartan generators. Thus the full spectral cover will have 248 sheets. In order to break to an $SU(5)$ GUT group, we turn on an $SL(5, \mathbf{C})$ Higgs bundle. The adjoint representation of E_8 decomposes as

$$\mathbf{248} = (\mathbf{24}, \mathbf{1}) + (\mathbf{1}, \mathbf{24}) + (\mathbf{5}, \mathbf{10}) + (\bar{\mathbf{5}}, \overline{\mathbf{10}}) + (\mathbf{10}, \bar{\mathbf{5}}) + (\overline{\mathbf{10}}, \mathbf{5}). \tag{2.88}$$

Thus the E_8 spectral cover breaks up into several pieces, which can be labelled by representations of the holonomy group of the Higgs bundle. Clearly the relevant spectral covers are those for the fundamental representation and for the anti-symmetric representation of $SU(5)$.

Referring back to the general form of a local $SU(5)$ model as derived using Tate's algorithm:

$$y^2 = x^3 + b_0 z^5 + b_2 x z^3 + b_3 y z^2 + b_4 x^2 z + b_5 x y. \tag{2.89}$$

The parameters can be identified with the following Casimirs of a meromorphic $Sl(5, \mathbf{C})$ Higgs bundle:

$$C_i(\Phi) = \mathrm{Tr}(\Phi^i) + \cdots = b_i / b_0. \tag{2.90}$$

To see this, the singularity (2.89) is generically of type A_4, but by sequentially tuning the b_i to zero we get successively $SO(10)$, E_6, E_7 and an E_8 singularity. Since the holonomy group of the Higgs bundle is the commutant of the gauge group in E_8, then the parameters must correspond to the indicated Casimirs. (A more precise way to see this [1] is by using the F-theory/heterotic duality map.) We see that there exists a canonical map between the parameters in the ALE fibration, and an $SU(5)$ spectral cover in $K_S \to S$ defined by

$$b_0 s^5 + b_2 s^2 + \cdots + b_5 = 0. \tag{2.91}$$

Note that η is related to our earlier t by $\eta = 6c_1 - t$. The five roots $\{\lambda_1, \ldots, \lambda_5\}$ of this polynomial determine the sizes of all the cycles of the E_8 ALE space. Recall that the λ_i are eigenvalues corresponding to certain eigen-lines,

$$\Phi \, |i\rangle = \lambda_i \, |i\rangle \, . \tag{2.92}$$

As discussed in Sect. 5.1 of [5], we can take the five roots of the polynomial to correspond to the periods of the following cycles (up to Weyl permutations)

$$\begin{aligned} |1\rangle &= \alpha_4, & |4\rangle &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4, \\ |2\rangle &= \alpha_3 + \alpha_4, & |5\rangle &= \alpha_{-\theta} + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4. \\ |3\rangle &= \alpha_2 + \alpha_3 + \alpha_4, \end{aligned} \tag{2.93}$$

The sizes of the cycles $\{\alpha_5, \ldots, \alpha_8\}$ are taken to be zero, generating an $SU(5)$ GUT group, and all other cycles are obtained as linear combinations. The matter curve $\Sigma_{\mathbf{10}}$ corresponds to $\lambda_i = 0$ for some i, the matter curve $\Sigma_{\mathbf{5}}$ corresponds to $\lambda_i + \lambda_j = 0$ for some i, j, etc. The top Yukawa is localized at $\lambda_i = \lambda_j = \lambda_i + \lambda_j = 0$, the bottom at $\lambda_i + \lambda_j = \lambda_k + \lambda_l = \epsilon_{ijklm}\lambda_m = 0$, and the $\mathbf{5} \cdot \bar{\mathbf{5}} \cdot \mathbf{1}$ (at least formally) at $\lambda_i + \lambda_j = \lambda_j + \lambda_k = \lambda_i - \lambda_k = 0$.

2.5. Construction of fluxes. Let us briefly recap what we saw above. Local models in F-theory correspond to ALE fibrations over a surface S with G-flux. Physically we expect that this data, the ALE fibration and the G-flux, can be described as configurations in an $8d$ supersymmetric Yang–Mills theory compactified on S, i.e. a Higgs bundle. This is indeed the case, and moreover this data is also equivalent to a covering C_E of the zero section in an auxiliary non-compact Calabi–Yau threefold $X = (K_S \to S)$, together with a holomorphic line bundle. In a IIb-like language, we can call this covering a non-compact flavour brane, whose intersection with the gauge brane (which is wrapped on the zero section of X) yields the matter curve $\Sigma_{\mathbf{10}}$. The group theory of E_8 implies that there is a second flavour-brane $C_{\Lambda^2 E}$, completely determined by the first covering, whose intersection with the gauge brane yields the matter curve $\Sigma_{\mathbf{5}}$.

Constructing the branes, or equivalently the ALE fibration, is easy: we only need to specify the b_i, which are sections of line bundles on S with Chern classes given by $\eta - ic_1$. In order to get chiral matter however we must actually turn on a flux on C_E, which will determine a unique flux on $C_{\Lambda^2 E}$ by group theory. In this subsection we discuss the issue of constructing such fluxes.

In order to facilitate the analysis we will compactify the local Calabi–Yau to

$$\bar{X} = \mathbf{P}(\mathcal{O} \oplus K_S). \tag{2.94}$$

$\bar{X}$ is certainly not a Calabi–Yau; the Calabi–Yau metric diverges at infinity. We denote by $\mathcal{O}(1)$ the line bundle on $\bar{X}$ which restricts to the eponymous line bundle on each $\mathbf{P}^1$-fiber. We may choose homogeneous coordinates (u_1, u_2) on the $\mathbf{P}^1$-bundle which are sections of $\mathcal{O}(1)$ and $\mathcal{O}(1) \otimes K_S$ respectively. The coordinate s used previously is identified with u_2/u_1.

The spectral cover $C \subset X$ is compactified to a compact surface $\bar{C} \subset \bar{X}$ by adding a divisor η_∞ at infinity. The equation

$$b_0 s^5 + b_2 s^3 + b_3 s^2 + b_4 s + b_5 = 0 \tag{2.95}$$

has a double zero at $u_1 = 0$. Therefore η_∞ covers η exactly once and is isomorphic to it. We denote the cohomology class dual to the zero section (the Poincaré dual of S in $\bar{X}$) by s_0 and the class of the section at infinity by $s_\infty = c_1(\mathcal{O}(1))$. Then we have $s_\infty = s_0 + c_1(TS)$ and $0 = s_0 \cdot s_\infty = s_0 \cdot (s_0 + c_1(TS))$.

We would like to lift line bundles on C to line bundles on the compact surface $\bar{C}$ which are easier to study. If the genus of η_∞ is non-zero, then there may be line bundles on C that cannot be lifted to $\bar{C}$. However any algebraic line bundle on C can be lifted. To see this, any algebraic line bundle on C is of the form $\mathcal{O}(D)_C$ for some divisor D. Let $\bar{D}$ be the closure of D in $\bar{C}$. Then $\mathcal{O}(\bar{D})_{\bar{C}}$ gives a lift of $\mathcal{O}(D)_C$ as desired. Moreover global G-fluxes are algebraic and yield an algebraic class in the local model. Therefore from here on we may restrict our attention to extendable line bundles.

Now consider a spectral line bundle L on $\bar{C}$. The corresponding Higgs bundle is given by $E = p_{C*}L,\ \ \Phi = p_{C*}s$. We have

$$c_1(p_{C*}L) = p_{C*}c_1(L) - \frac{1}{2}p_{C*}r, \tag{2.96}$$

where r is the ramification divisor, $r = -c_1(\bar{C}) + p_C^* c_1(S)$. Explicitly we find that

$$r = (n-2)s_0 + p_C^*(\eta - c_1(S)). \tag{2.97}$$

If $c_1(p_{C*}L)$ is not zero, then we have a $GL(n, \mathbf{C})$ Higgs bundle rather than an $SL(n, \mathbf{C})$ Higgs bundle. For phenomenological applications we want the latter, so we need to impose the restriction $c_1(p_{C*}L) = 0$ on the allowed spectral line bundles. Then it is convenient to decompose

$$c_1(L) = \frac{1}{2}r + \lambda\gamma, \tag{2.98}$$

where λ is a parameter. The condition $c_1(p_{C*}L) = 0$ is then equivalent to $p_{C*}\gamma = 0$. The class $r/2$ is generally not integer quantized. Since $c_1(L)$ must be integer quantized, γ must compensate and can generally not be an integer class either, but it will always be a rational linear combination of integer classes.

So our task is to find integer classes γ with $p_{C*}\gamma = 0$ in $H^2(\bar{C}, \mathbf{Z})$. To get a supersymmetric configuration, γ must further be of Hodge type $(1, 1)$. As we will now argue, for generic complex structure moduli there exists only one such class (up to multiplication by an integer), and we can write it down explicitly.

As for line bundles on $\bar{C}$, we can use the Lefschetz–Noether theorem. $\bar{C}$ is the zero locus of a section of $\mathcal{O}(n) \otimes L_{\eta - nc_1}$, which is usually an ample line bundle since $\mathcal{O}(n)$ is ample and $L_{\eta - nc_1}$ is effective and non-zero. Therefore there is an injective map $i^* : H^{1,1}(\bar{X}) \to H^{1,1}(\bar{C})$. As a result, $H^{1,1}(\bar{C})$ splits into two pieces, the classes inherited from $\bar{X}$ and the primitive classes. The Noether–Lefschetz theorem says that when $\bar{C}$ is ample, then for 'generic' complex structure moduli there are no primitive classes in $H^{1,1}(\bar{C})$.

Let us write down the inherited class explicitly. The cohomology group $H^2(\bar{X}, \mathbf{Z})$ is spanned by s_0 and $\pi^* H^2(S, \mathbf{Z})$, i.e. the pull-back of classes on S to $\bar{X}$. Therefore the inherited classes in $H^{1,1}(\bar{C})$ are spanned by the class of the matter curve $\Sigma_{\mathbf{10}}$, as well as any class on S pulled back to $\bar{C}$. (In particular, $c_1(\bar{C})$ is in this span, by the adjunction formula, and so is η_∞.) Our class γ will be a linear combination of those, but it also needs to satisfy $p_{C*}\gamma = 0$, which is clearly not satisfied by any class pulled back from S. Thus we can single out the class $[\Sigma_E]$ and subtract the 'trace', i.e. we single out the following unique linear combination:

$$\gamma_u = n[\Sigma_E] - p_C^* p_{C*}[\Sigma_E] = n[\Sigma_E] - p_C^*(\eta - nc_1). \tag{2.99}$$

We used the subscript u on γ_u to indicate that this class is universal, i.e. it always exists in a local model. In the last equality we just use the fact that $p_{C*}[\Sigma_E]$ is just the class $[\Sigma_E]$ sitting inside $H^2(S)$, and since it is given by $b_n = 0$ it follows that it can also be written as $\eta - nc_1$.

Let us define a line bundle using this class. Its first Chern class will be given by

$$c_1(L) = \frac{1}{2}r + \lambda\gamma_u, \tag{2.100}$$

with γ_u as defined above and λ a parameter. From our explicit expressions for r and γ_u, we see that $c_1(L)$ is an integer class when λ is an integer and n is even, or when $\lambda = \frac{1}{2}+$ integer and n is odd. For this corresponding choice of spectral line bundle, we can deduce the net amount of chiral matter. It is given by

$$N_{\text{chiral}} = -\chi(i_*\mathcal{O}_S, j_*L) = +\chi(L \otimes K_S|_{\Sigma_{\mathbf{10}}}) = \lambda \int_\Sigma \gamma_u, \tag{2.101}$$

where in the last equality we used the Riemann–Roch formula and the fact that $(r/2 + c_1(K_S))|_\Sigma = -c_1(\Sigma)/2$. We have

$$\Sigma \cdot_{\bar{C}} \Sigma = S_0 \cdot_{\bar{X}} S_0 \cdot_{\bar{X}} \bar{C} = -c_1(S_0) \cdot_{S_0} \Sigma. \tag{2.102}$$

Further we have $\Sigma \cdot_{\bar{C}} p^*\alpha = \alpha \cdot_{S_0} \Sigma$ for any $\alpha \in H^2(S, \mathbf{Z})$. Applying this with $\alpha = \eta - nc_1$, we see that

$$\gamma_u \cdot_{\bar{C}} \Sigma = -\eta \cdot_{S_0} \Sigma. \tag{2.103}$$

Therefore we find

$$N_{\text{chiral}} = \lambda \int_\Sigma \gamma_u = -\lambda\eta(\eta - nc_1). \tag{2.104}$$

This is of course the same formula as encountered in spectral cover constructions in the heterotic string [40].

Therefore the only fluxes available for general complex structure moduli will give the conventional chirality formula known from the heterotic string. We do not see more general options in the local F-theory set-up. We will call such fluxes *inherited* or *universal*. However, there do exist more general fluxes, both in the heterotic setting and in the F-theory setting. The point is that general fluxes are not supersymmetric for generic Higgs bundle moduli, and thus are not among the fluxes that we found above. For special values of the moduli (which is called the Noether–Lefschetz locus) there are additional supersymmetric fluxes available, and turning on such fluxes would therefore automatically stabilize some of the moduli. We will call them Noether–Lefschetz fluxes or *non-inherited* fluxes. Generic fluxes are non-inherited. They exist for both F-theory spectral covers and heterotic spectral covers, where they give rise to rigid bundles on the CY_3 after Fourier–Mukai transform. But they are harder to write down and have not really been analyzed in either context.

2.6. Further constraints. In the previous sections we encountered a number of constraints that must be satisfied for consistency of the local model. Now we would like to consider imposing a few further constraints, that are not needed for consistency but are likely needed to get a realistic and calculable four-dimensional model. We will concentrate on $SU(5)$ models, so there is a matter curve

$$[\Sigma_{\mathbf{10}}] = c_1 - t, \tag{2.105}$$

which must be effective and non-zero.

The Kähler class J is a generator of $H^2(B_3, \mathbf{R})$ that must be positive on all the effective cycles of the geometry. Modulo these positivity constraints, there is an independent scale in the geometry for every generator of $H^2(B_3, \mathbf{R})$. In order to get a model that is calculable and predictive, we need some small parameters that we can expand in. The main requirement that we want to make is that M_{GUT}/M_{Pl} is unbounded from below, where $M_{GUT} \sim V_S^{-1/4}$ and $M_{Pl} \sim V_{B_3}^{1/2}/\kappa \sim V_{B_3}^{1/2}/\alpha_{GUT} V_S$.

Now although $M_{GUT}/M_{Pl} \to 0$ should make gravity subleading, it may not be enough to decouple the local model. Depending on the geometry of B_3, there may be additional scales in the global model with physics that cannot be decoupled from the visible sector. In order to say that the visible sector is largely independent of the rest of B_3, we will want to require that one can take $V_S \to 0$ while keeping V_{B_3} *as well as* the volume of any other cycles not inside S finite. This implies that the kinetic terms of fields localized on any cycle not contained in S are large compared to the kinetic terms of the fields localized on S. Moreover this yields a local constraint that can be checked without knowing the compactification manifold B_3.

There are two ways in which we could take $V_S \to 0$ while keeping other cycles fixed. The first is that we could require S to contract to a point. This will turn out to be a very strong condition which will essentially single out a unique model. We could also require S to contract to a curve of singularities. This is a less stringent condition, but together with some other physical constraints will still rule out a good deal of models.

Thus our first assumption is as follows:

1. *Contractibility.* S can be contracted to a point. By Grauert's criterion [41], this means that the class t must be ample, in particular $t \cdot C > 0$ for any curve C in S.[5]

We can draw some immediate conclusions from this assumption. Since $c_1 - t$ is effective and non-zero, and t is ample, c_1 must be effective and non-zero. Therefore K_S^{-n} cannot have sections for any positive n and the Kodaira dimension is $-\infty$. From the classification of surfaces, we then know that S is related to $\mathbf{P}^2$ or a ruled surface (i.e. a $\mathbf{P}^1$-fibration over a Riemann surface of genus g) by a sequence of blow-ups and blow-downs.

The ruled surfaces have $h^{1,0}(S) = g$, which would lead to massless adjoint fields in the effective four-dimensional theory if $g > 0$. This looks phenomenologically undesirable, so we will exclude this possibility with our second assumption:

2. *No adjoint scalars.* The Hodge numbers of S must satisfy $h^{0,1}(S) = 0$ and $h^{2,0}(S) = 0$.

Then S is either $\mathbf{P}^2$ or can be obtained by a sequence of blow-ups from a Hirzebruch surface $\mathbf{F}_r$. Note this includes all the del Pezzo surfaces. The divisors on $\mathbf{F}_r$ are generated by b, f and E_i, with the intersection numbers

[5] Even though we have seen that for many purposes S can be regarded as living inside the total space of the canonical bundle, this criterion has nothing to do with contractibility in the auxiliary local Calabi-Yau. We will see this more explicitly in the global examples later.

$$b\cdot b=-r,\qquad b\cdot f=1,\qquad f\cdot f=0,\qquad b\cdot E_i=f\cdot E_i=0,\qquad E_i\cdot E_j=-\delta_{ij}. \tag{2.106}$$

By exchanging b and $b+f$, we may take $r\geq 0$. Further we have

$$c_1(\mathbf{F}_r)=2b+(r+2)f-\sum_{i=1}^{k}E_i. \tag{2.107}$$

Similarly we may write

$$t=n_b b+n_f f-\sum_{i=1}^{k}n_i E_i. \tag{2.108}$$

Let us first assume we are on $\mathbf{F}_r$, with no blow-ups. From ampleness of t we get $n_b>0$, $-n_b r+n_f>0$. Since c_1-t is effective and non-zero, we also get $n_b\leq 2$ and $n_f\leq r+2$, with strict inequality for n_f if $n_b=2$ or vice versa. Then we either have $n_b=1$ and $r<n_f\leq r+2$, or we have $n_b=2$, $n_f=r+1$ and $r=0$ or 1.

We may add a further reasonable assumption which eliminates most of these models. Currently, there is only one known mechanism for breaking the GUT group while preserving the standard GUT relations at leading order [5,6]. This mechanism requires a -2-class on S (i.e. a class with $x\cdot x=-2$) in order to avoid massless lepto-quarks. This class must further be orthogonal to any classes that are inherited from B_3 in order to avoid Higgsing hypercharge or losing the standard $SU(5)$ relations between the gauge couplings at leading order. Let us take this as our third assumption:

3. *GUT breaking using fluxes.* There must exist a -2-class $x\in H^2(S,\mathbf{Z})$ which is orthogonal to any class inherited from $H^2(B_3,\mathbf{Z})$. In particular, $x\cdot c_1=x\cdot t=0$.

Let us again consider the Hirzebruch surfaces. Then $h^2(\mathbf{F}_r)=2$ so by condition (3) it follows that t must be a rational multiple ac_1 of c_1. Since $t\cdot f>0$, the coefficient a must be positive, and since c_1-t must be effective, the coefficient a must be ≤ 1. But this happens only for r even in which case c_1 is divisible by 2, so this leaves

$$S=\mathbf{F}_r\quad\text{with } r \text{ even},\qquad \Sigma_{\mathbf{10}}=\frac{1}{2}c_1. \tag{2.109}$$

But now by condition (1) we get $t\cdot b=-r+2>0$, so this leaves only $S=\mathbf{F}_0$ and $t=\frac{1}{2}c_1$. Note that $\mathbf{P}^2$ is also ruled out by condition (3).

Now we consider the case of Hirzebruch surfaces with at least one blow-up. Again we have the constraints above from $t\cdot b>0$, $t\cdot f>0$ and c_1-t effective. However we also get $t\cdot E_i=n_i>0$ and $t\cdot(f-E_i)=n_b-n_i>0$. Hence we must have

$$t=2b+(r+1)f-\sum_{i=1}^{k}E_i. \tag{2.110}$$

From $t\cdot b=-r+1>0$ we find that $r=0$. Moreover, $\mathbf{F}_r$ with one blow-up is actually the same surface as $\mathbf{F}_1$ with one blow-up, so $r=0$ is ruled out as well, and therefore all cases with blow-ups are ruled out. So we conclude that Assumptions (1)–(3) leave a unique possibility for S and t:

$$(1)+(2)+(3)\;\Rightarrow\quad S=\mathbf{F}_0,\qquad t=\frac{1}{2}c_1,\qquad \Sigma_{\mathbf{10}}=\frac{1}{2}c_1. \tag{2.111}$$

We will study this case in more detail later in the paper. In particular we will show how to engineer three-generation models and how to embed it in a global model.

It is evident by now that condition (1) in particular is quite strong. In order to have $V_S \to 0$ while keeping other cycles fixed, we can also replace assumption (1) by:

$1'$. *Contractibility.* S can be contracted to a curve, i.e. S admits a fibration $F \to S \to B$ where the fibers F can be contracted to a curve B of singularities.

In this case t is not necessarily ample, but t must be ample when restricted to the components that are being contracted [42,43]. Therefore, $t \cdot C > 0$ when C is the general fiber F or an irreducible component of the singular fibers.

A priori the base B of the fibration can have any genus g, in which case we would have $h^{1,0}(S) = g$. Again by assumption (2) the base B is restricted to be $\mathbf{P}^1$. Likewise, the fibers must be rational: the curve $c_1 - t = \Sigma_{\mathbf{10}}$ is effective and the fiber F moves, so there must be some copy of F that is not contained in $\Sigma_{\mathbf{10}}$. Therefore we must have $(c_1 - t) \cdot F \geq 0$. Since F gets contracted, we have $t \cdot F > 0$, and it follows that $c_1 \cdot F > 0$. But by the adjunction formula we have $c_1 \cdot F = 2 - 2g(F)$, hence $c_1 \cdot F = 2$ and F is a $\mathbf{P}^1$ as promised.

So our S is a ruled surface with rational base and fibers. From the classification of surfaces, we know that S is rational and can be obtained by blowing up some points on a Hirzebruch surface $\mathbf{F}_r$. The most general possibility is obtained by blowing up some points on a conic bundle, which allows for the possibility of a multiple fiber. Our argument below is actually even more constraining when there are multiple fibers, so in what follows we will focus on the case that none of the fibers is multiple.

If S is a Hirzebruch surface then we can run the previous argument. Using condition (3) it follows that $t = \frac{1}{2}c_1$ and r is even. Apart from these we must consider possible blow-ups of $\mathbf{F}_r$. Again we write

$$t = n_b b + n_f f - \sum_{i=1}^{k} n_i E_i. \tag{2.112}$$

Under assumption ($1'$) we can no longer conclude that $t \cdot b$ must be positive, but we still know that t must be positive on f, E_i and $f - E_i$. From $t \cdot f > 0$, $t \cdot E_i > 0$ and $t \cdot (f - E_i) > 0$ we get $n_b > 0$, $n_i > 0$, and $n_b - n_i > 0$. From $c_1 - t$ effective and non-zero we get $n_b \leq 2$ and $n_f \leq r + 2$, with strict inequality for n_f if $n_b = 2$ and $n_i = 1$. Therefore the only possibility is

$$S = B_k(\mathbf{F}_r), \qquad t = 2b + n_f f - \sum_{i=1}^{k} E_i, \qquad \Sigma_{\mathbf{10}} = (r + 2 - n_f) f \tag{2.113}$$

with $n_f < r+2$. Here we used B_k to denote blowing-up k times. These possibilities also satisfy condition (3), since there are classes of the form $f - E_i - E_j$ and $E_i - E_j$ which are orthogonal to c_1 and t. Moreover we can't do too many blow-ups. Recall that the sections b_i specifying an $SU(5)$ model live in $c_1 - t, \ldots, 6c_1 - t$, so these line bundles need to admit sections.

So we conclude that under assumptions ($1'$), (2) and (3), we get the following possibilities for S and t:

$$S = \mathbf{F}_r \quad \text{with } r \text{ even}, \qquad t = \frac{1}{2}c_1, \qquad \Sigma_{\mathbf{10}} = \frac{1}{2}c_1 \tag{2.114}$$

or

$$S = B_k(\mathbf{F}_r), \qquad t = 2b + n_f f - \sum_{i=1}^{k} E_i, \qquad n_f < r + 2. \tag{2.115}$$

The remaining possibility $S = \mathbf{P}^2$ is still ruled out by assumption (3).

We may consider adding one final assumption. We will soon see though that this assumption has a very important loophole, so it will be weakened significantly.

4. *Three generations.* The net number of generations is given by

$$-\lambda(6c_1 - t) \cdot_{dP} (c_1 - t), \tag{2.116}$$

where $\lambda \in \mathbf{Z} + \frac{1}{2}$. As we argued, this is the only universal formula one can write down. However this does not represent the most general configuration of local F-theory models and will be revisited in Sect. 2.8.

Let's apply this to all the possibilities we found. For the Hirzebruch surfaces with $t = \frac{1}{2}c_1$ we find that the minimal number of generations is eleven. For the blow-ups of Hirzebruch surfaces with t as in (2.115), we find that the minimal number of generations is $5 \times (r + 2 - n_f)$, and in general it is always divisible by 5. We conclude it is not possible to make a local three generation $SU(5)$ model under these assumptions.

If we drop condition (1) or (1′) it is not hard to find three-generation models. For instance, the dP_8 example in [1] with $\eta = 6c_1(S)$ is consistent and satisfies assumptions (2) and (3), but it does not satisfy assumption (1) or (1′) since it has $t = 0$, and is therefore not a truly local model.

It may be interesting to point out that the three generation $SO(10)$ models in [1] (which have $\eta = 4c_1 + E$, where E is any -1-curve, and $\Sigma_{\mathbf{16}} = [E]$) do satisfy the conditions (1),(2) and (3) for dP_k with $2 \leq k \leq 7$. However a fully satisfactory way of breaking the $SO(10)$ GUT group in these models while preserving gauge coupling unification has not yet been identified. (This has recently been revisited in [44], where it was shown that one can actually use hypercharge flux without introducing exotics.)

To summarize this subsection, under conditions (1) – (3) we only found one possibility for S and t, listed in (2.111). Under assumptions (1′) – (3) we only found the possibilities listed in (2.114) and (2.115). Using the inherited fluxes (assumption (4)), none of these models could account for three generations. In the following subsections, we will examine some possible loopholes in our assumptions.

2.7. Another way to break the GUT group. In [5,6] the GUT group was broken to the Standard Model gauge group by turning on an abelian flux. A priori there exists a second possibility: one may also break the GUT group to the Standard Model by turning on an abelian Higgs field.

To do this, we take a degree six spectral cover

$$b_0 s^6 + b_2 s^4 + \cdots + b_6 = 0 \tag{2.117}$$

which generically breaks E_8 to $SU(3) \times SU(2)$. Note that b_1 must vanish if the structure group is to be in $SU(6)$ rather than $U(6)$. Now if the b_i are such that this equation factorizes

$$(c_0 s + c_1)(d_0 s^5 + d_1 s^4 + \cdots + d_5) = 0, \tag{2.118}$$

where $c_0 d_1 + c_1 d_0 = 0$, then the structure group of the Higgs field commutes with $SU(3) \times SU(2) \times U(1)$.

The matter curves are easy to find. Consider first an irreducible degree 6 spectral cover (this was worked out in [3]). One uses the following decomposition of the adjoint of E_8 under $SU(3) \times SU(2) \times SU(6)$:

$$\begin{aligned}\mathbf{248} = &(\mathbf{8}, \mathbf{1}, \mathbf{1}) + (\mathbf{1}, \mathbf{3}, \mathbf{1}) + (\mathbf{1}, \mathbf{1}, \mathbf{35})\\ &+(\mathbf{3}, \mathbf{2}, \mathbf{6}) + (\bar{\mathbf{3}}, \mathbf{2}, \bar{\mathbf{6}}) + (\mathbf{3}, \mathbf{1}, \mathbf{15}) + (\bar{\mathbf{3}}, 1, \bar{\mathbf{15}}) + (\mathbf{1}, \mathbf{2}, \mathbf{20}).\end{aligned} \tag{2.119}$$

We have the following matter curves:

$$\begin{aligned}(\mathbf{3}, \mathbf{2}) &\to \{b_6 = 0\},\\ (\mathbf{3}, \mathbf{1}) &\to \{b_0 b_5^3 - b_2 b_3 b_5^2 + b_4 b_5 b_3^2 - b_3^3 b_6 = 0\},\\ (\mathbf{1}, \mathbf{2}) &\to \{b_6(b_2^2 - 4 b_4 b_0) + b_0 b_5^2 - b_2 b_5 b_3 + b_4 b_3^2 = 0\}.\end{aligned} \tag{2.120}$$

In the reducible case we simply substitute the b_i for the appropriate bilinears in c_i and d_j.

In addition there can be hypercharged scalars. To see this, recall that the moduli of the spectral cover are counted by

$$\mathrm{Ext}^1_X(i_* L, i_* L). \tag{2.121}$$

When the spectral cover is reducible, this decomposes as

$$\sum_{m,n} \mathrm{Ext}^1_X(i_{m*} L_m, i_{n*} L_n), \tag{2.122}$$

where m, n run over the irreducible components. The off-diagonal zero modes are clearly charged under the extra $U(1)$'s since their VEVs smooth the spectral cover and break these $U(1)$'s. These modes are localized at the intersection $c_0 s + c_1 = d_0 s^5 + d_1 s^4 + \ldots + d_5$. In order to get a realistic model, such hypercharged scalars must be absent, and we have to pick a spectral line bundle which ensures that.

There are unfortunately at least two possible issues with this scenario. The polynomial b_6 is a section of N_S. Thus not only $c_1 - t$ should have a section, but also $-t$. By our classification of the possible pairs (S, t), even with condition (3) dropped, such an S can not be contractible, which means it is not a true local model. Thus this mechanism can only be used if we drop the requirement that M_{GUT}/M_{Pl} can be made parametrically small.

The second issue is that since we are in some sense directly breaking E_8 to $SU(3) \times SU(2) \times U(1)_Y$, it is not guaranteed that the gauge couplings will unify. This could perhaps be addressed by requiring c_1/c_0 to be small, as the spectral cover degenerates and the breaking of $SU(5)_{GUT}$ to the Standard Model gauge group should be small in the limit $c_1/c_0 \to 0$.

2.8. Noether–Lefschetz fluxes. The results of Sect. 2.6 clarify our options. Dropping the first assumption is a priori possible and leads to consistent models, but it would mean we can not make an expansion in M_{GUT}/M_{Pl} and therefore would greatly diminish the predictiveness of F-theory GUTs. Dropping assumption (3) means that we need an alternative mechanism to break the GUT group while preserving the standard $SU(5)$ relations at leading order. We made such an alternative proposal in Sect. 2.7 but it did not seem to be compatible with contractibility of S. Therefore we are led to drop assumption (4) and investigate the possibilities of non-inherited G-fluxes.

The argument in Sect. 2.6 that the inherited fluxes are not sufficient by no means rules out three generation models. Rather it means that we need to look at more general fluxes that are not critical points of the holomorphic Chern–Simons superpotential for generic moduli, and we have to do more work to show that there exists a stable supersymmetric minimum. In mathematics circles this would be called a Noether–Lefschetz problem. On the other hand, the first three assumptions already ruled out all but a handful of 7-brane configurations, listed in (2.111), (2.114) and (2.115) in Sect. 2.6. Thus in contrast to eg. heterotic model building, we have a very restricted set of possibilities to start with and we know all the continuous parameters.

Let us first ignore the requirement of supersymmetry, and simply ask if there are any fluxes available, not necessarily of type (1, 1), which might give three generations. This will be the case if we can show there exists a class $\gamma_2 \in H^2(\bar{C}, \mathbf{Z})$ which is orthogonal to $p_C^* H^2(S, \mathbf{Z})$ and satisfies $\gamma_2 \cdot \Sigma_{\mathbf{10}} = 1$, since then we can add some integer multiple of it to $\frac{1}{2}r + \frac{1}{2}\gamma_u$ and get any number of generations we want. Equivalently, γ_2 must satisfy

$$\gamma_2 \cdot \gamma_u = \gamma_2 \cdot (5\Sigma_{\mathbf{10}} - p_C^* p_{C*} \Sigma_{\mathbf{10}}) = 5. \tag{2.123}$$

Now the lattice $H^2(\bar{C}, \mathbf{Z})$ modulo torsion is unimodular by Poincaré duality. Thus if γ_u is primitive in the sense that it is not an integer multiple of a smaller integer class, then there exists an $\alpha \in H^2(\bar{C}, \mathbf{Z})$ such that

$$\alpha \cdot \gamma_u = 1. \tag{2.124}$$

Then defining $\gamma_2 = 5\alpha - p_C^* p_{C*}\alpha$, we have $\gamma_2 \cdot \gamma_u = 5$ and $p_{C*}\gamma_2 = 0$ as required. Therefore we are guaranteed that the required fluxes exist if γ_u is primitive in the sense above. In a unimodular lattice, a sufficient condition is that $\gamma_u \cdot \gamma_u$ is not divisible by any square. It is easy to see that $\gamma_u \cdot \gamma_u = 5\gamma_u \cdot \Sigma_{\mathbf{10}}$ which we have already computed in Sect. 2.6. For $S = \mathbf{P}^1 \times \mathbf{P}^1$ with $t = \frac{1}{2}c_1$, it is equal to 5×22 which does not have any squares. So we conclude from a purely topological argument that it is possible to obtain three generations, although this argument cannot establish that there is a supersymmetric minimum for finite values of the Higgs bundle moduli.

In the following we would like to give a fairly general construction of *algebraic* classes that satisfy $\alpha \cdot_{\bar{C}} \gamma = 1$. We will apply it to $S = \mathbf{P}^1 \times \mathbf{P}^1$, but it should be clear that with some simple substitutions it can also be applied to the other cases. Thus we will finally establish some examples of supersymmetric $SU(5)$ models with three generations and S contractible.

The strategy is as follows: we first take a curve $\alpha_0 \in H^2(S, \mathbf{Z})$ such that $\alpha_0 \cdot_S \Sigma_{\mathbf{10}} = 1$. Then we will construct a curve $\alpha \in \bar{X}$ which does not intersect $\Sigma_{\mathbf{10}}$, and which covers α_0 exactly once. Finally we will require $\bar{C}$ to contain α by tuning the complex structure moduli. The result is an algebraic class in $H^2(\bar{C}, \mathbf{Z})$ with $\alpha \cdot_{\bar{C}} \gamma = 1$. We can then construct an additional flux γ_2 as above by subtracting the trace, and define a spectral line bundle with

$$c_1(L) = \frac{1}{2}r + \frac{1}{2}\gamma_u + n\gamma_2. \tag{2.125}$$

By adjusting n, we then get any number of generations we want.

In the case of interest, we have $S = \mathbf{P}^1 \times \mathbf{P}^1$. We denote the coordinates on S by $(z_1, z_2; w_1, w_2)$, and the two rulings by H_1 and H_2, with intersection numbers $H_1^2 = H_2^2 = 0$, $H_1 \cdot H_2 = 1$. As we deduced above, the matter curve should be given by $\Sigma_{\mathbf{10}} = [H_1 + H_2]$, and we need a class with $\alpha_0 \cdot \Sigma_{\mathbf{10}} = 1$. Thus a simple choice is to pick $\alpha_0 = H_1$, though clearly there are additional options. In equations it is given by (say) $w_1 = 0$.

Now we need to construct α. As before we use coordinates (u_1, u_2) on the $\mathbf{P}^1$-fibers of $\bar{X} = \mathbf{P}(\mathcal{O} \oplus K_S)$. Then we define α by the following two equations:

$$\alpha:\ w_1 = 0, \qquad u_1 = P_2(z, w)u_2, \tag{2.126}$$

where $P_2(z, w)$ is a section of $\mathcal{O}(2, 2)$. Note that $u_2 = 0 \Rightarrow u_1 \neq 0$, so α does not intersect the zero section $u_2 = 0$. This is needed if we want $\bar{C}$ to contain α, because the intersection of $\bar{C}$ with $u_2 = 0$ is by definition $\Sigma_{\mathbf{10}}$, and we promised to construct a class which does not intersect $\Sigma_{\mathbf{10}}$. Also, α covers α_0 precisely once.

Therefore we now need to show that we can tune the complex structure moduli so that $\bar{C}$ contains α. The equation of $\bar{C}$ is given by

$$b_0(z, w)u_2^5 + b_2(z, w)u_1^2u_2^3 + \cdots + b_5(z, w)u_1^5 = 0. \tag{2.127}$$

We simply substitute the equations for α in order to get a restriction on the coefficients of $\bar{C}$. Clearly we find that

$$w_1 \quad \text{divides} \quad b_0(z, w) + b_2(z, w)P_2(z, w)^2 + \cdots + b_5(z, w)P_2(z, w)^5. \tag{2.128}$$

This can be satisfied by leaving $b_2, \ldots, b_5$ arbitrary, and putting

$$b_0(z, w) = -\Big[b_2(z, w)P_2(z, w)^2 + \cdots + b_5(z, w)P_2(z, w)^5\Big]_{w_1 \to 0} + \mathcal{O}(w_1). \tag{2.129}$$

The only thing left to check is that $\bar{C}$ is generically smooth, so that our calculations of the chiral spectrum apply. But this is fairly obvious because generically the derivatives of Eq. (2.127), even with (2.129), give independent equations.

Therefore we have constructed a (1, 1) class $\gamma_2 = 5\alpha - p_C^* p_{C*}\alpha$ with the desired properties. Defining a spectral line bundle as in (2.125) with $n = 8$, we find precisely three chiral generations on $\Sigma_{\mathbf{10}}$.

2.9. 'Flux' landscapes in the heterotic string and F-theory. As we already remarked, much of the structure of F-theory vacua is identical with that of the heterotic string. BPS instantons effects, branes and flux superpotentials, which are some of the main ingredients inducing potentials for the moduli, can be related under the duality. In particular, semi-realistic heterotic models appear to have an enormous number of 'flux' vacua as well. We put 'flux' in quotation marks here because after Fourier–Mukai transform, we get a smooth non-abelian bundle on the Calabi–Yau threefold without any $U(1)$ fluxes. These extra 'flux' vacua are obtained by using spectral line bundles that are not inherited, and generically should stabilize all vector bundle moduli. There is a landscape of such vacua and a priori it is not clear why we should exclude these possibilities. The method we used for constructing such more general fluxes in Sect. 2.8 can also be used in the heterotic string.

Thus landscapes seem to be quite generic properties of superpotentials in string theory. It would be interesting to study these vacua microscopically. In the heterotic setting there are no RR fields so one could try to use conventional CFT techniques. Perhaps one may then find a reason to exclude most of them, although it is currently not clear why that would be the case.

As in type II settings, this leads to philosophical problems: we don't really understand moduli stabilization and the cosmological constant problem very well, it is practically impossible to enumerate all the vacua that seem to exist at the effective field theory level, and one of the solutions that have been proposed to solve the cosmological constant problem is NP hard [45]. A possible way out was promoted in [46]: if M/M_{Pl} can be parametrically small, where M is some scale relevant for particle physics like the GUT scale, then we can prevent the unknown physics responsible for solving gravity-related problems from feeding back into physics at the scale M. This may allow us to discuss phenomenology without having to solve the cosmological constant problem and other problems related to gravity. But combining this principle with GUTs leads us to F-theory; this idea cannot be implemented in the heterotic string.

2.10. Conclusions. We have clarified the rules for constructing local models in F-theory. Such models can be defined by specifying suitable spectral data (a type of B-brane) in an auxiliary Calabi–Yau geometry. We classified the possible matter curves for local $SU(5)$ models. We have constructed the first truly local $SU(5)$ models with three generations. We did not compute the exact matter content in our model, so it is possibly still an open problem to construct local $SU(5)$ models with exactly the MSSM spectrum, or some acceptable extension. However since it is now clear that there is a landscape already in local models, and since there is no index theorem that prevents chiral/anti-chiral pairs from lifting, we should expect the number of models with the MSSM spectrum to be enormous.

We explained how to connect F-theory models to a IIb picture by taking an orientifold limit. For $SU(5)_{GUT}$ models, it was found that the E_6 points on S where the up-type $\mathbf{10} \cdot \mathbf{10} \cdot \mathbf{5}$ Yukawa couplings are localized became singular points in the perturbative IIb limit. However a number of general issues pertaining to the IIb limit remained unresolved. A much better approach to this problem will be presented in [21].

We also found that it seems to be impossible for a local $SU(5)$ model with completely unstabilized Higgs field moduli to have three generations. From a physical perspective, this is good news since there are many indications that we do not want a generic model, such as problems associated to dimension four and five proton decay. Thus requiring a three-generation model automatically stabilizes some of the moduli. Requiring the precise MSSM spectrum will likely stabilize even more moduli. On the other hand, this also makes the problem of constructing realistic local models much more challenging.

Along the way we have encountered a number of constraints that the matter curves and fluxes on the matter curves must satisfy in a consistent local model, from topological and integral (such as anomaly cancellation (2.29) and the fact that the flux must lift to the integral class of a line bundle L with $c_1(p_{C*}L) = 0$), to analytic (such as the forced singularities on $\Sigma_{\Lambda^2 E}$ and constraints on the moduli entering the matter curves for non-inherited fluxes, so that L is a holomorphic line bundle). Probably we have not found them all, since some of these constraints look very non-trivial from the point of view of the brane carrying the unbroken gauge group. However we have shown that there is an isomorphism relating a configuration (E, Φ) in the $8d$ gauge theory to its spectral data (C_E, L_E), and the constraints on the spectral data are few and simple to understand.

In the next section we will make the first strides towards embedding our local models in a global model.

3. Compactification

In order to get a finite four-dimensional Planck scale we should embed our local models into a compact elliptically fibered CY fourfold. In the philosophy of local model building, the goal of this pursuit is not to find ‘the’ UV completion of the local model. Indeed as we reviewed earlier, it is not even clear that this is an answerable question. Rather it is to ascertain that all the ingredients used can in principle be combined in a UV complete model, and there are no obvious constraints from UV completion that would rule them out.

In doing so there are many issues to be addressed. Our aim here is rather modest; we would like to discuss a simple set of compactifications which implement a few of the requirements for viable local GUTs, and which make clear how such constructions work in general. In particular we would like to construct compactifications in which GUT breaking by fluxes can be implemented, and in which dangerous proton decay channels can be avoided.

Our discussion will have one important caveat. We will freely assume that appropriate fluxes may be found which give precisely the Standard Model spectrum on the matter curves we engineer. The point of our discussion is not to understand the fluxes, but rather some of the constraints on the geometry of the four-fold arising from phenomenological requirements.

3.1. First example: cubic surface in $\mathbf{P}^3$. Let us discuss simple compactifications of local toy models with $SU(5)$ GUT group. Our GUT brane should be wrapped on a Del Pezzo surface $S_2 \subset B_3$, such that some homology classes in the Del Pezzo become boundaries when embedded in B_3. For simplicity we will take B_3 to be $\mathbf{P}^3$ in our first example, although much of what we will say can clearly be adapted to more general Fano three-folds. Then we can take S_2 to be a quadric surface $Q_2(z_1, z_2, z_3, z_4) = 0$ (i.e. $\mathbf{P}^1 \times \mathbf{P}^1$) or a cubic surface $Q_3(z_1, z_2, z_3, z_4) = 0$ (i.e. a Del Pezzo 6). For definiteness we take the cubic.

Recall again the Tate form of the Weierstrass equation,

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6, \tag{3.1}$$

where the a_i are sections of $K_{B_3}^{-i}$. In the case of $B_3 = \mathbf{P}^3$, the a_i are polynomials of degree $4i$. As we discussed, in order to get an I_5-locus along $Q_3 = 0$, as well as matter curves and Yukawa coupling localized along certain prescribed submanifolds, we must impose certain restrictions on the a_i which can be read from Table 1.

Now let us try to impose various constraints.

1. *$SU(5)$ gauge group on* $Q_3 = 0$. According to Table 1 this implies the following leading form for the a_i:

$$a_1 = P_4, \quad a_2 = sP_5, \quad a_3 = s^2 P_6, \quad a_4 = s^3 P_7, \quad a_6 = s^5 P_9, \tag{3.2}$$

where $s = Q_3$ and the P_n are generic polynomials of degree n on $\mathbf{P}^3$. They are identified with the sections b_{9-n} that appeared in the general discussion. In fact we are clearly allowed to add further subleading terms, eg. $a_1 = P_4 + sT_1$, $a_2 = sP_5 + s^2 T_2$, et cetera. Such additional subleading terms do not affect behaviour near the I_5 locus, but they do provide additional flexibility in building a global model. To keep things simple, we will set them to zero. Then the discriminant is computed to be

$$\Delta = s^5 P_4^4 (-P_9 P_4{}^2 + P_6 P_7 P_4 - P_5 P_6{}^2) + \mathcal{O}(s^6), \tag{3.3}$$

which vanishes to 5^{th} order along $Q_3 = 0$, as required.

2. *Matter curves.* The discriminant vanishes to higher order along $\Sigma_{\mathbf{10}} = \{Q_3 = P_4 = 0\}$ and $\Sigma_{\mathbf{5}} = \{Q_3 = R_{17} = 0\}$, where

$$R_{17} \equiv P_9 {P_4}^2 - P_6 P_7 P_4 + P_5 {P_6}^2. \tag{3.4}$$

Note that

$$\Lambda^2 T_{Q_3} = \mathcal{O}(1)|_{Q_3}, \qquad N_{Q_3} = \mathcal{O}(3)|_{Q_3} \tag{3.5}$$

and hence the cohomology classes dual to $\Sigma_{\mathbf{10}}$ and $\Sigma_{\mathbf{5}}$ on Q_3 are given by $c_1 - t$ and $8c_1 - 3t$. Of course this all fits in the general discussion in Sects. 2.1 and 2.3.

3. *Yukawa couplings and dimension four proton decay.* The up-type Yukawa couplings are localized at $\{Q_3 = P_4 = P_5 = 0\}$ and the down type Yukawas are localized at $\{Q_3 = P_4 = P_6 = 0\}$. Methods for suppressing proton decay in F-theory were discussed in [5,6,30]. Here we will see how they can be implemented in global models. In order to prevent dimension four proton decay, we want to make sure that

$$\mathbf{10}_m \cdot \bar{\mathbf{5}}_m \cdot \bar{\mathbf{5}}_m \Rightarrow \text{absent}, \qquad \mathbf{10}_m \cdot \bar{\mathbf{5}}_m \cdot \bar{\mathbf{5}}_{h_d} \Rightarrow \text{present}. \tag{3.6}$$

The coupling $\mathbf{10}_m \cdot \bar{\mathbf{5}}_h \cdot \bar{\mathbf{5}}_h$ will be absent automatically due to the anti-symmmetry. We will try to implement the above by splitting $\Sigma_{\mathbf{5}}$ into two pieces, one supporting the matter fields and another supporting the Higgses. This means we have to tune the P_n so that the polynomial R factorizes modulo Q_3. In terms of ideals, we require a decomposition

$$\langle Q_3, R\rangle = I_{\bar{\mathbf{5}}_m} \cap I_h. \tag{3.7}$$

To secure the absence of R-parity violating down-type Yukawa couplings we must make sure that whenever we have an intersection of $\Sigma_{\mathbf{10}_m}$ with $\Sigma_{\bar{\mathbf{5}}_m}$, there is also a branch of $\Sigma_{\bar{\mathbf{5}}_h}$ intersecting at that point. Since $R = 0$ has a double point at such intersections, we can also say that whenever $\Sigma_{\mathbf{10}_m}$ intersects with $\Sigma_{\bar{\mathbf{5}}_m}$, then $\Sigma_{\bar{\mathbf{5}}_m}$ is not allowed to have a double point (the second order vanishing of R instead being due to a branch of $\Sigma_{\bar{\mathbf{5}}_h}$ coming in and intersecting there). Similarly in order to avoid the couplings $\mathbf{10}_m \cdot \bar{\mathbf{5}}_h \cdot \bar{\mathbf{5}}_h$, we want to avoid double points on $\Sigma_{\bar{\mathbf{5}}_{h_d}}$ which also meet $\Sigma_{\mathbf{10}_m}$. We don't know the general solution to this algebraic problem. But to see that it can be achieved, we will exhibit one simple solution that exists for more general $SU(5)$ models as well. We take

$$\begin{aligned} P_6 &= H_6^d \mod Q_3, \qquad & P_9 &= H_1^u T_2 H_6^d \mod Q_3, \\ P_7 &= H_1^u T_6 \mod Q_3, \qquad & P_5 &= H_1^u T_4 \mod Q_3, \end{aligned} \tag{3.8}$$

for some T_i and H_i of the appropriate degree, but otherwise arbitrary. Then we take $\Sigma_{\mathbf{5}_h} = \{Q_3 = H_1^u H_6^d = 0\}$, i.e. the Higgs curve is actually reducible, with only up type Yukawa couplings on $H_1^u = 0$ (since $H_1^u = 0$ implies $P_5 = 0$) and only down type Yukawa couplings on $H_6^d = 0$. When we discuss dimension five proton decay we will see why that is a good thing to have. Now we can factorise $\Sigma_{\mathbf{5}}$ as

$$R_{17} = (H_1^u H_6^d) \cdot M_{10} \mod Q_3, \qquad M_{10} = T_2 P_4^2 - T_6 P_4 + T_4 P_6 \tag{3.9}$$

and M_{10} has no double points at $Q_3 = P_4 = P_6 = 0$. Moreover the up and down-type Yukawa's are still present. For instance the up-type Yukawa's come from $Q_3 = P_4 = H_1^u = 0$, which consists of $3 \cdot 4 \cdot 1 = 12$ points.

There are additional cubic couplings of the form

$$\bar{\mathbf{5}}_m \cdot \mathbf{5}_{h_u} \cdot \mathbf{1}, \qquad \bar{\mathbf{5}}_{h_d} \cdot \mathbf{5}_{h_u} \cdot \mathbf{1}. \tag{3.10}$$

The singlets correspond to Higgs field moduli (which are complex structure moduli of the Calabi–Yau fourfold). At least three of them should give rise to right-handed neutrinos, with the first coupling in (3.10) corresponding to the usual Yukawa couplings for neutrinos. The number of moduli appearing in such couplings is the difference between the number of moduli describing $\Sigma_{\bar{\mathbf{5}}_m}$ and $\Sigma_{\mathbf{5}_{h_u}}$ separately or as a single smooth curve, which yields 65 singlets in our example. The problem of getting Majorana masses of the right order of magnitude is a problem of moduli stabilization for the Higgs fields. The couplings on the right give rise to the minimal extension of the MSSM with a dynamical μ-parameter. There are additional constraints from dimension five proton decay however, as we discuss next.

4. *Dimension five proton decay.* We further want to eliminate dimension five proton decay. This proceeds through mediation of massive KK triplets T_u, T_d propagating on curves supporting a hypermultiplet in the $\mathbf{5}$. The possible channels are given by

$$\begin{array}{ccccccc} Q\,Q & \stackrel{\lambda_u}{\longleftrightarrow} & T_u & \stackrel{m^{ab}}{\longleftrightarrow} & T_d & \stackrel{\lambda_d}{\longleftrightarrow} & Q\,L \\ \Sigma_{\mathbf{10}_m} \times \Sigma_{\mathbf{10}_m} & & \Sigma^a_{\mathbf{5}} & & \Sigma^b_{\mathbf{5}} & & \Sigma_{\mathbf{10}_m} \times \Sigma_{\bar{\mathbf{5}}_m} \end{array} . \tag{3.11}$$

In order to prevent such processes, we have to shut off at least one of the interactions in this chain. If H_u and H_d propagate on the same matter curve, and if we assume the existence of classical up-type and down-type Yukawa couplings for the Standard Model, then such decays are unavoidable. Since we want to keep the classical Yukawa couplings, we require the existence of a decomposition

$$I_h = I_u \cap I_d, \tag{3.12}$$

so that we can shut off the coupling m^{ab}. If we allow Σ_u and Σ_d to intersect, then there could either be a branch of $\Sigma_{\mathbf{10}_m}$ also intersecting there; or it can correspond to a $\mathbf{5}_u \cdot \bar{\mathbf{5}}_d \cdot \mathbf{1}$ coupling. As long as the VEV of the singlet vanishes we do not have the troublesome mass terms linking triplets localized on Σ_u and Σ_d. In either case the existence of a classical μ-term is excluded. Our example corresponds to the latter case: there are $3 \cdot 1 \cdot 6 = 18$ intersection points on $H^u_1 = H^d_6 = 0$ corresponding to the couplings $\mathbf{5}_u \cdot \bar{\mathbf{5}}_d \cdot \mathbf{1}$.

There are several possible alternate channels for dimension five proton decay (3.11). The most dangerous are cases where $\Sigma^a_{\mathbf{5}} = \Sigma^b_{\mathbf{5}}$, because then mass terms m^{ab} between T_u and T_d cannot be avoided. The case $\Sigma^a_{\mathbf{5}} = \Sigma^b_{\mathbf{5}} = \Sigma_{\bar{\mathbf{5}}_m}$ is harmless by the solution to dimension four proton decay, which shuts off the interactions λ_d. The case $\Sigma^a_{\mathbf{5}} = \Sigma^b_{\mathbf{5}} = \Sigma_{\bar{\mathbf{5}}_{h_d}}$ requires shutting off the interactions λ_u. The curve $\Sigma_{\mathbf{10}_m}$ is positive in our example and therefore certainly intersects $\Sigma_{\bar{\mathbf{5}}_{h_d}}$. However in our solution to the dimension four problem, by design any such intersection has $P_4 = P_6 = 0$ and therefore corresponds to a λ_d coupling, not a λ_u coupling, so this channel is not available. Finally there is the case $\Sigma^a_{\mathbf{5}} = \Sigma^b_{\mathbf{5}} = \Sigma_{\mathbf{5}_{h_u}}$, which requires shutting off the interaction λ_d. In our example, by design any intersection point between $\Sigma_{\mathbf{5}_{h_u}}$ and $\Sigma_{\mathbf{10}}$ yields an up-type Yukawa, so the potentially troublesome interactions are again absent.

The remaining possible channels have $\Sigma^a_{\mathbf{5}} \neq \Sigma^b_{\mathbf{5}}$. Assuming both the λ_u and λ_d couplings are present (which they need not necessarily be), the problem is to shut

off the interactions m^{ab}. This depends on the existence of intersections of $\Sigma^a_{\mathbf{5}}$ and $\Sigma^b_{\mathbf{5}}$ which give rise to a coupling $\mathbf{5}\cdot\bar{\mathbf{5}}\cdot\mathbf{1}$. If such intersections are present, m^{ab} is proportional to the VEV of the singlet, which is a complex structure modulus. As long as the dynamics of moduli stabilization is such that the VEV of this field remains zero, there will be no proton decay through this channel.
Consider for instance $\Sigma^a_{\mathbf{5}} = \Sigma_{\bar{\mathbf{5}}_m}$ and $\Sigma^b_{\mathbf{5}} = \Sigma_d$ (the other cases being similar). The curves $\Sigma_{\bar{\mathbf{5}}_m}$ and Σ_d can intersect in two ways. Either there is also a branch of $\Sigma_{\mathbf{10}}$ intersecting there, which corresponds to the down type Yukawas that we want to have; or it corresponds to a $\mathbf{5}\cdot\bar{\mathbf{5}}\cdot\mathbf{1}$ coupling. In our example with $I_{h_d} = \langle Q_3, H^d_6\rangle$, intersection points where $H^d_6 = 0$ and $M_{10} = 0$ have either $P_4 = 0$ or $T_2P_4 - T_6 = 0$. In the former case it meets with $\Sigma_{\mathbf{10}}$, and there are $3\cdot 6\cdot 4 = 72$ such intersection points; in the latter case it corresponds to the coupling to singlets whose VEV must remain zero, and this accounts for $3\cdot 6\cdot 6 = 108$ intersection points.

Hence we see in this simple example that there seems to be enough room in complex structure moduli space to implement our geometric requirements for absence of dimension four and five proton decay. This is perhaps a little misleading, because we have not carefully examined the fluxes. When we take the hypercharge flux into account, we find the following problem. In order to get just the Higgses on Σ_{H_u} and Σ_{H_d} and no light triplets, the hypercharge flux through these curves should be non-zero and opposite. However the homology classes of the curves above are always pull-backs of homology classes in the bulk, and so the hypercharge flux through them must vanish, if we are to avoid giving a mass to the hypercharge gauge boson through a Stückelberg coupling to axions.

This problem could be circumvented in general by a slight variation of the factorization above. With the notation in Sect. 2, the matter curves are given by

$$\Sigma_{\mathbf{10}} = \{b_5 = 0\}, \qquad \Sigma_{\bar{\mathbf{5}}} = \{b_0 b_5^2 - b_2 b_3 b_5 + b_3^2 b_4 = 0\}. \tag{3.13}$$

Now we will factorize $\Sigma_{\bar{\mathbf{5}}}$ as

$$\Sigma_{H_u} = \{h = 0\}, \qquad \Sigma_{H_d} = \{\beta_1 b_5 + \gamma b_3 = 0\}, \qquad \Sigma_{\bar{\mathbf{5}}_m} = \{\beta_2 b_5 + b_3 = 0\}, \tag{3.14}$$

where $b_4|_{Q_3=0} = h\cdot\gamma$. Then it is not hard to see that in general one may have non-zero and opposite hypercharge flux through Σ_{H_u} and Σ_{H_d}. If moreover there are no simultaneous zeroes for b_5 and γ, then the above arguments for avoiding proton decay still go through.

An alternative approach would be to implement the approach of [30], which requires decomposing the $SU(5)$ Casimirs into those of a smaller holonomy group, so that one has additional $U(1)$ symmetries available.

3.2. Second example: a contractible $\mathbf{P}^1\times\mathbf{P}^1$. In our previous example the del Pezzo was not contractible in B_3. The main purpose of this subsection is to give a simple example of a del Pezzo S which has two-cycles not inherited from B_3 (necessary for allowing GUT breaking fluxes), and which is also contractible in B_3. This is an explicit realization of case (2.111) discussed in Sect. 2.6.

The example is as follows. We will take B_3 to be the blow-up of $\mathbf{P}^3$ along a curve C defined by

$$C = \{Q_2 = 0\}\cap\{Q_3 = 0\}. \tag{3.15}$$

The corresponding ideal is denoted as I_C and the blow-up along this ideal as $B_3 = \tilde{\mathbf{P}}^3$. We have

$$K_{\tilde{\mathbf{P}}^3} = i^* K_{\mathbf{P}^3} + \tilde{C}, \tag{3.16}$$

where $\tilde{C}$ is the exceptional divisor (a $\mathbf{P}^1$-fibration over C, given by projectivising the normal bundle). Sections of the anti-canonical bundle $K^{-1}_{\tilde{\mathbf{P}}^3}$ are sections of $K^{-1}_{\mathbf{P}^3}$ which are also in the ideal I_C. In particular there are non-trivial sections in $K^{-4}_{\tilde{\mathbf{P}}^3}$ and $K^{-6}_{\tilde{\mathbf{P}}^3}$, and so we can write a Weierstrass equation and construct elliptic fibrations over $\tilde{\mathbf{P}}^3$ which are Calabi-Yau.

In this example, the del Pezzo on which the gauge branes are wrapped will be the Hirzebruch surface $\mathbf{F}_0 = \mathbf{P}^1 \times \mathbf{P}^1$, here defined by $Q_2 = 0$. As mentioned, the reason for picking this model as our next example is that the proper transform of $S = \{Q_2 = 0\}$ is contractible. To see this, let us first check that the normal bundle is indeed negative. The normal bundle of S in $\mathbf{P}^3$ is $\mathcal{O}(2)|_S$. After blowing up C, the new normal bundle is $\mathcal{O}(2)|_S \otimes \mathcal{O}(-\tilde{C})|_S$, where $\tilde{C}$ is the exceptional divisor. But the intersection of $\tilde{C}$ with S is in $\mathcal{O}(3)|_S$, so

$$\mathcal{O}(2) \otimes \mathcal{O}(-\tilde{C})|_S \sim \mathcal{O}(-1)|_S. \tag{3.17}$$

Hence the normal bundle is negative, a necessary condition for being contractible.

By looking at this example slightly differently, one may establish that S is indeed contractible. Consider a cubic hypersurface Q in $\mathbf{P}^4$ vanishing to second order at a point p. Let T be the tangent space to $\mathbf{P}^4$ at p. We identify T with an open subset of $\mathbf{P}^4$, and write the Taylor expansion of Q at p as:

$$Q = Q_2 + Q_3, \tag{3.18}$$

with $Q_2,\ Q_3$ as in (3.15). If we also identify the $\mathbf{P}^3$ of (3.15) with the projectivization of this T, we see that the set of lines in Q through its singular point p can be identified with the curve C.

Consider the projection $\widetilde{\mathbf{P}^4} \to \mathbf{P}^3$ with center p, where $\widetilde{\mathbf{P}^4}$ is the blowup of $\mathbf{P}^4$ at p. It restricts to a surjective morphism $\pi : \widetilde{Q} \to \mathbf{P}^3$, where $\widetilde{Q}$ is the blowup of Q at p. The exceptional divisor S_0 in $\widetilde{Q}$ is mapped by π isomorphically to a quadric surface in $\mathbf{P}^3$ that can be identified with our surface S. On the other hand, the inverse image of each point of $C \subset \mathbf{P}^3$ is the corresponding line in Q. We thus have an identification of $\widetilde{Q}$ with $B_3 = \widetilde{\mathbf{P}^3}$, showing that S can indeed be blown down to the singular point p of the cubic threefold Q.

In order to break the GUT group without generating a mass for hypercharge, we need a class in S which is topologically trivial in B_3. For $S = \mathbf{P}^1 \times \mathbf{P}^1$ there is a unique candidate, the difference between the two rulings. It's not hard to see that the two $\mathbf{P}^1$'s yield equivalent classes in B_3: they have the same intersection number with the transform of the hyperplane class in $\mathbf{P}^3$, as well as with the exceptional divisor.

Now let us write explicitly the elliptic fibration. Once again we recall the Tate form of the Weierstrass equation

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6 \tag{3.19}$$

and repeat the exercise of the previous subsection. The a_i are sections of $K^{-i}_{\tilde{\mathbf{P}}^3}$, which means they are polynomials of degree $4i$ on $\mathbf{P}^3$ which vanish to ith order along C. Let us define $s = Q_2$ and $u = Q_3$. A polynomial which vanishes to ith order along C is a sum of terms each of which is of degree at least i in s and u.

1. *$SU(5)$ gauge group on $Q_2 = 0$.* According to Table 1 this implies the following form for the a_i:

$$\begin{aligned} a_1 &= P_1 u + P_2 s, & a_4 &= s^3(P_7 u + P_8 s), \\ a_2 &= s(P_3 u + P_4 s), & a_6 &= s^5(P_{11} u + P_{12} s). \\ a_3 &= s^2(P_5 u + P_6 s), \end{aligned} \tag{3.20}$$

Here we included various subleading terms ($P_2, P_4, P_6, P_8, P_{12}$) which are not directly needed, and which will not show up in the analysis of the matter curves and Yukawa couplings.

2. *Matter curves.* The discriminant is given by

$$\Delta = s^5 u^7 P_1^4 R_{13} + \mathcal{O}(s^6), \tag{3.21}$$

where

$$R_{13} = -P_1^2 P_{11} + P_1 P_5 P_7 - P_3 P_5^2. \tag{3.22}$$

If we recall that $\tilde{C}$ is a $\mathbf{P}^1$-fibration over C, then the intersection of $\tilde{C}$ with $s = 0$ gives the section at 'zero' and with $u = 0$ gives the section at 'infinity.' Hence after blowing up along C, the surfaces $s = 0$ and $u = 0$ no longer intersect; instead they intersect $\tilde{C}$ along two disjoint curves. Thus from the discriminant we read off that the matter curves are given by

$$\Sigma_{\mathbf{10}} = \{Q_2 = P_1 = 0\}, \tag{3.23}$$

which is generically a rational curve, and

$$\Sigma_{\mathbf{5}} = \{Q_2 = R_{13} = 0\}. \tag{3.24}$$

Recall we showed above that $N_S = \mathcal{O}(-1)|_S$, and it is not hard to see that $c_1 \sim \mathcal{O}(2)|_S$. Therefore the homology classes of the matter curves are given by $\mathcal{O}(1)|_S \sim c_1 - t$ and $\mathcal{O}(13)|_S \sim 8c_1 - 3t$, in full agreement with the general discussion.

3. *Yukawa couplings and proton decay.* As is familiar by now, the Yukawa couplings are localized at $\lambda_{up} \sim \{Q_2 = P_1 = P_3 = 0\}$ and $\lambda_{down} \sim \{Q_2 = P_1 = P_5 = 0\}$. The discussion of the first example goes through if we choose the analogous factorization:

$$P_{11} = P_5 T_5 H_1^u \mod Q_2, \qquad P_7 = T_6 H_1^u \mod Q_2, \qquad P_3 = T_2 H_1^u \mod Q_2. \tag{3.25}$$

With this factorization we have

$$R_{13} = (P_5 H_1^u)\, M_7, \qquad M_7 = -P_1^2 T_5 + P_1 T_6 - T_2 P_5, \tag{3.26}$$

and we identify $\Sigma_{\bar{\mathbf{5}}_{h_d}} = \{Q_2 = P_5 = 0\}$, $\Sigma_{\mathbf{5}_{h_u}} = \{Q_2 = H_1^u = 0\}$, and $\Sigma_{\bar{\mathbf{5}}_m} = \{Q_2 = M_7 = 0\}$. We refer to the discussion in the first example for why this eliminates the classical dimension four and five proton decay.

Again, we encounter the problem with the hypercharge flux that we saw in the previous example, and a slightly different factorization along the lines suggested there is needed to solve the problem.

The local version of this model was discussed in Sect. 2.8. We demonstrated that it is possible to engineer three chiral generations by putting a mild restriction on P_{11}, but we did not calculate the exact spectrum for the Noether–Lefschetz flux constructed there. One may have to engineer a different flux to yield the exact Standard Model spectrum, but we have also seen there there is an enormous landscape of such Noether–Lefschetz fluxes available.

Our second example is really a special case of a more general construction. Consider blowing up a Fano threefold along a curve C, and assume that the blow-up still admits a CY T^2-fibration. The proper transform of the surface S of minimal degree containing C is usually contractible, by the reasoning around Eq. (3.17). Moreover such a surface will typically have homology classes which are not inherited from the ambient space, as in the example above, if the surface had such classes before blowing up. If the degree of the surface is not too large, we can prescribe an I_5 fibration along it.

Acknowledgements. We would like to thank V. Alexeev, F. Denef, M. Gross and T. Pantev for helpful discussions, and J. Heckman for patiently explaining many details of his works [2,6,8]. The research of R.D. is supported by NSF grant DMS 0612992 and Research and Training Grant DMS 0636606.

References

1. Donagi, R., Wijnholt, M.: Model building with F-theory. Adv. Theor. Math. Phys. **15**, 1237. http://arXiv.org/abs/0802.2969v3 [hep-th], 2011
2. Beasley, C., Heckman, J.J., Vafa, C.: GUTs and exceptional branes in F-theory—I. JHEP **0901**, 058 (2009)
3. Hayashi, H., Tatar, R., Toda, Y., Watari, T., Yamazaki, M.: New aspects of heterotic—F theory Duality. Nucl. Phys. B **806**, 224 (2009)
4. Vafa, C.: Evidence for F-theory. Nucl. Phys. B **469**, 403 (1996)
5. Donagi, R., Wijnholt, M.: Breaking GUT groups in F-theory. Adv. Theor. Math. Phys. **15**, 1523. http://arXiv.org/abs/0808.2223v1 [hep-th], 2011
6. Beasley, C., Heckman, J.J., Vafa, C.: GUTs and exceptional branes in F-theory—II: experimental predictions. JHEP **0901**, 059 (2009)
7. Tatar, R., Watari, T.: GUT relations from string theory compactifications. Nucl. Phys. B **810**, 316 (2009)
8. Heckman, J.J., Vafa, C.: Flavor hierarchy from F-theory, http://arXiv.org/abs/0811.2417v3 [hep-th], 2009
9. Heckman, J.J., Kane, G.L., Shao, J., Vafa, C.: The footprint of F-theory at the LHC, http://arXiv.org/abs/0903.3609v1 [hep-ph], 2009
10. Hayashi, H., Kawano, T., Tatar, R., Watari, T.: Codimension-3 singularities and Yukawa couplings in F-theory, http://arXiv.org/abs/0901.4941v3 [hep-th], 2009
11. Andreas, B., Curio, G.: From local to global in F-theory model building, http://arXiv.org/abs/0902.4143v1 [hep-th], 2009
12. Blumenhagen, R., Braun, V., Grimm, T.W., Weigand, T.: GUTs in type IIB orientifold compactifications. Nucl. Phys. B **815**, 1–94 (2009)
13. Font, A., Ibanez, L.E.: Yukawa structure from $U(1)$ fluxes in F-theory grand unification. JHEP **0902**, 016 (2009)
14. Blumenhagen, R.: Gauge coupling unification in F-theory grand unified theories. Phys. Rev. Lett. **102**, 071601 (2009)
15. Bourjaily, J.L.: Local models in F-theory and M-theory with three generations, http://arXiv.org/abs/0901.3785v2 [hep-th], 2009
16. Jiang, J., Li, T., Nanopoulos, D.V., Xie, D.: F-SU(5). Phys. Lett. B **677**, 322 (2009)
17. Collinucci, A.: New F-theory lifts. JHEP **0908**, 076 (2009)
18. Collinucci, A., Denef, F., Esole, M.: D-brane deconstructions in IIB orientifolds. JHEP **0902**, 005 (2009)
19. Heckman, J.J., Marsano, J., Saulina, N., Schafer-Nameki, S., Vafa, C.: Instantons and SUSY breaking in F-theory, http://arXiv.org/abs/v20808.1286 [hep-th], 2009
20. Marsano, J., Saulina, N., Schafer-Nameki, S.: An instanton toolbox for F-theory model building. JHEP **1001**, 128 (2010)

21. Clingher, A., Donagi, R., Wijnholt, M.: The Sen limit, http://arXiv.org/abs/1212.4505v1 [hep-th], 2012
22. Pantev, T., Wijnholt, M.: Hitchin's equations and M-theory phenomenology. J. Geom. Phys. **61**, 1223 (2011)
23. Donagi, R., Katz, S., Wijnholt, M. Weak coupling, degeneration and log Calabi–Yau spaces, http://arXiv.org/abs/1212.0553v1 [hep-th], 2012
24. Tate, J.: *Algorithm for determining the type of a singular fiber in an elliptic pencil*. In: Modular functions of one variable IV, Lecture Notes in Mathematics, Vol. **476**, Berlin, Springer, 1975, pp. 33–52
25. Bershadsky, M., Intriligator, K.A., Kachru, S., Morrison, D.R., Sadov, V., Vafa, C.: Geometric singularities and enhanced gauge symmetries. Nucl. Phys. B **481**, 215 (1996)
26. Sen, A.: Orientifold limit of F-theory vacua. Phys. Rev. D **55**, 7345 (1997)
27. Sen, A.: F-theory and orientifolds. Nucl. Phys. B **475**, 562 (1996)
28. Dasgupta, K., Mukhi, S.: F-theory at constant coupling. Phys. Lett. B **385**, 125 (1996)
29. Sadov, V.: Generalized Green–Schwarz mechanism in F theory. Phys. Lett. B **388**, 45 (1996)
30. Tatar, R., Watari, T.: Proton decay, Yukawa couplings and underlying gauge symmetry in string theory. Nucl. Phys. B **747**, 212 (2006)
31. Cheung, Y.K., Yin, Z.: Anomalies, branes, and currents. Nucl. Phys. B **517**, 69 (1998)
32. Minasian, R., Moore, G.W.: K-theory and Ramond–Ramond charge. JHEP **9711**, 002 (1997)
33. Witten, E.: Anomaly cancellation on G(2) manifolds, http://arXiv.org/abs/hepth-th/0108165v1, 2001
34. Hitchin, N.J.: The self-duality equations on a Riemann surface. Proc. Lond. Math. Soc. (55) **3**, 59–126 (1987)
35. Simpson, C.T.: Higgs bundles and local systems. Publications Mathématiques de l'IHES, **75**, 5–95 (1992)
36. Donagi, R.: Spectral covers, http://arXiv.org/abs/alg-geom/9505009v1, 1995
37. Donagi, R., Wijnholt, M.: Gluing Branes- I. JHEP **1305**, 068. http://arXiv.org/abs/arXiv:1104.2610 [hep-th], 2013
38. Donagi, R., He, Y.H., Ovrut, B.A., Reinbacher, R.: The particle spectrum of heterotic compactifications. JHEP **0412**, 054 (2004)
39. Blumenhagen, R., Moster, S., Reinbacher, R., Weigand, T.: Massless spectra of three generation U(N) heterotic string vacua. JHEP **0705**, 041 (2007)
40. Curio, G.: Chiral matter and transitions in heterotic string models. Phys. Lett. B **435**, 39 (1998)
41. Grauert, H.: Ueber Modifikationen und exzeptionelle analytische Mengen. Math. Ann. **146**, 331–368 (1962)
42. Ancona, V.: Faisceaux amples sur les espaces analytiques. Trans. Am. Math. Soc. **274**(1), 89–100 (French English summary) [Ample sheaves on analytic spaces] (1982)
43. Peternell, T.: Manuscripta Math. **37**, 19–26 (1982); Erratum, ibid. **42**, 259–263 (1983)
44. Donagi, R., Wijnholt, M.: Gluing branes II: flavour physics and string duality. JHEP **1305**, 092. http://arXiv.org/abs/arXiv:1112.4854v1 [hep-th], 2013
45. Denef, F., Douglas, M.R.: Computational complexity of the landscape I. Ann. Phys **322**, 1096 (2007)
46. Verlinde, H., Wijnholt, M.: Building the standard model on a D3-brane. JHEP **0701**, 106 (2007)

Communicated by N. A. Nekrasov