DESY-12-047
TUM-HEP 830/12

Origin of family symmetries

Hans Peter Nilles[a], **Michael Ratz**[b], **Patrick K.S. Vaudrevange**[c]

[a] *Bethe Center for Theoretical Physics and Physikalisches Institut der Universität Bonn, Nussallee 12, 53115 Bonn, Germany*

[b] *Physik–Department T30, Technische Universität München, James–Franck–Straße, 85748 Garching, Germany*

[c] *Deutsches Elektronen–Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany*

Abstract

Discrete (family) symmetries might play an important role in models of elementary particle physics. We discuss the origin of such symmetries in the framework of consistent ultraviolet completions of the standard model in field and string theory. The symmetries can arise due to special geometrical properties of extra compact dimensions and the localization of fields in this geometrical landscape. We also comment on anomaly constraints for discrete symmetries.

1 Introduction

Discrete symmetries play an important role in particle physics. Apart from the fundamental space–time symmetries P, C and T, there are various well known examples such as the so–called matter or R parity in the minimal supersymmetric standard model (MSSM). There are good reasons for using discrete rather than continuous symmetries. Models with spontaneously broken global continuous symmetries exhibit Goldstone bosons which are typically phenomenologically unacceptable. Moreover, there are strong arguments that a continuous symmetry has either to be gauged or it will be broken by quantum gravity effects (see e.g. [1] for a recent discussion). In contrast to the fundamental symmetries, discrete symmetries are often just imposed by hand for phenomenological reasons. While introducing such symmetries can be a useful tool in bottom–up model building it appears worthwhile to clarify the origin of a given symmetry. Given a deeper understanding of how such symmetries arise, one might be able to obtain a more fundamental understanding of observations, such as the repetition of families and the flavor structure.

Discrete symmetries come in various classes. Various generation–dependent flavor symmetries, have been proposed in order to explain the pattern of quark and lepton Yukawa couplings, and to control higher–dimensional operators (see e.g. [2] for a quite recent review and other contributions of this special issue [3] for more references). Apart from these there are generation–independent symmetries, introduced in order to cure certain shortcomings of extensions of the standard model such as the MSSM. For example, dangerous proton decay operators are forbidden by matter parity [4, 5], baryon triality [6], proton hexality [7] and $\mathbb{Z}_4^R$ [8]. Further, discrete symmetries of high order can manifest themselves as accidental global U(1) symmetries in the (truncated) low–energy effective theory. Such accidental symmetries can be used for example in two ways: as (anomalous) Peccei–Quinn symmetry addressing the strong CP problem (cf. the discussion in [9, 10]) or as a $\mathrm{U}(1)_R$ explaining the hierarchy between the Planck and the electroweak scales [11].

The purpose of this review is to clarify the origin of discrete symmetries. They can be obtained from continuous symmetries by spontaneous breaking. But this is not the only possibility. In fact, here we mainly focus on alternative possibilities for the origin of discrete symmetries. In section 2 we discuss, in the framework of field theory, how discrete symmetries can be related to the geometry of extra dimensions. The discussion of higher–dimensional quantum field theories leaves certain questions unanswered. We therefore change gear and present a top–down derivation of discrete symmetries in section 3, focusing mainly on heterotic orb-

ifolds, as they provide us with explicit candidate models for a UV completion of the standard model, and, at the same time, allow for a CFT description and hence for a detailed understanding of the symmetries. As we shall see, the top–down settings are more restrictive than the bottom–up models. Some of the restrictions can be thought of as originating from the requirement of anomaly freedom, which we discuss separately in section 4. Finally, we summarize our discussion in section 5.

2 Geometrical origin of discrete symmetries

In this section we present three possible origins of discrete symmetries. After briefly summarizing the standard approach and its limitations in section 2.1, we discuss how to obtain a discrete symmetry from extra dimensions, either as the symmetry of compact space (section 2.2) or as a remnant of higher dimensional Lorentz symmetry (section 2.3).

2.1 Gauged discrete symmetries from continuous symmetries

The perhaps most straightforward possibility for obtaining a discrete symmetry is by spontaneous breaking of a continuous gauge symmetry. As a simple example, consider a U(1) gauge group broken by the VEV of a scalar φ with charge $q = 3$. Here we normalize the U(1) such that the charges are integer and have no common divisor. The unbroken symmetry is given by those U(1) transformations that leave the vacuum invariant, i.e.

$$\mathrm{e}^{\mathrm{i}\,\alpha(x)\,q}\,\langle\varphi\rangle \;=\; \mathrm{e}^{3\mathrm{i}\,\alpha(x)}\,\langle\varphi\rangle \;\stackrel{!}{=}\; \langle\varphi\rangle \quad \curvearrowright \quad \alpha(x) \;=\; \frac{2\pi\,n}{3}\,, \tag{2.1}$$

with $n = 0, 1, 2$. Hence, the (local) U(1) is broken to a (local) $\mathbb{Z}_3$ subgroup. The extension of this discussion to the case of multiple U(1) factors which get broken by several VEVs is given in [12].

One may also get non–Abelian discrete symmetries by spontaneous breaking (cf. e.g. [13–15]). However, this typically involves very large representations of the corresponding continuous symmetry, which often give rise to unwanted states in the broken phase. Therefore, arguably, this possibility appears not to be too attractive. In what follows, we therefore discuss alternative possibilities in which the discrete symmetries are related to the geometry of compact dimensions. As we shall see, this scheme does not suffer from the above problems, and is realized in explicit string–derived models of particle physics.

2.2 Repetition of families and symmetries

Discrete family symmetries can be motivated in settings with extra compact dimensions. It is not surprising that such models offer an explanation for the appearance of non–Abelian discrete flavor symmetries, because the latter are symmetries of certain geometrical solids, which describe the compact dimensions. The symmetries of internal space govern the interactions between fields that are localized in the compact dimensions and may eventually become flavor symmetries.

The purpose of this subsection is to explain that (non–Abelian) family symmetries can, to some extent, be understood geometrically. Let us start with a very simple example with one extra compact dimension, the orbifold $\mathbb{S}^1/\mathbb{Z}_2$ (figure 1). See appendix A.1 for a brief introduction to the construction of orbifolds. This

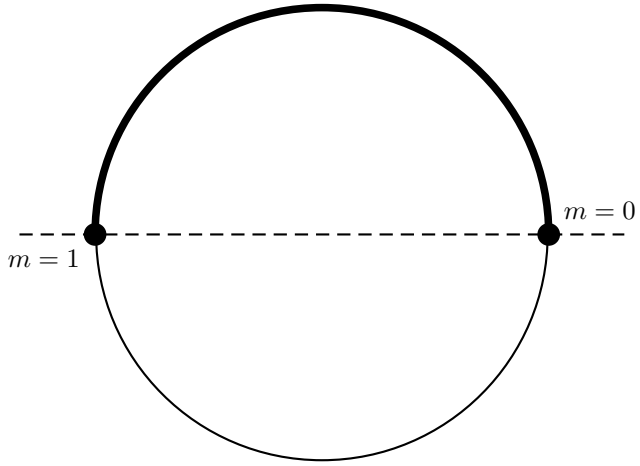
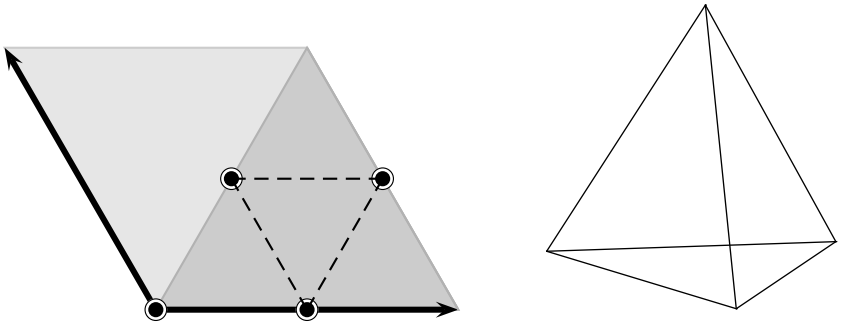


Figure 1: Example for one extra compact dimension: $\mathbb{S}^1/\mathbb{Z}_2$ orbifold. Points which are related by a reflection on the dashed line are identified. The fundamental region of the orbifold is an interval with the fixed points sitting at the boundaries.

orbifold possesses two geometrically equivalent fixed points. Suppose there are two states, i.e. two families of quarks and/or leptons, $\psi_{m=0}$ and $\psi_{m=1}$, with identical quantum numbers, one of them localized at each of the fixed points. Since the fixed points and the states ψ_m are geometrically indistinguishable, there is an S_2 permutation symmetry relating them, which manifests itself as a symmetry of the theory.

A somewhat more complex example is the tetrahedron in two extra compact dimensions (cf. [16]), which can be obtained from the $\mathbb{T}^2_{\mathrm{SU}(3)}/\mathbb{Z}_2$ orbifold (figure 2). Here the subscript SU(3) indicates that the basic translations defining the $\mathbb{T}^2$ torus

enjoy the same relations as the simple roots of the Lie algebra of SU(3), i.e. enclose 120° and have equal lengths.



(a) $\mathbb{T}^2_{\mathrm{SU}(3)}/\mathbb{Z}_2$. (b) Tetrahedron.

Figure 2: Example for two extra compact dimensions: If the $\mathbb{T}^2$ lattice vectors have equal length and enclose 120°, one can also fold the fundamental region of $\mathbb{T}^2/\mathbb{Z}_2$ (dark gray region in (a)) to the tetrahedron (b).

Clearly, the tetrahedron is invariant under a discrete rotation by 120° about an axis that goes through one corner and hits the opposite surface orthogonally. There are four operations of this type represented by

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} , \quad T\,S , \quad T\,S' , \quad T\,S\,S' , \tag{2.2}$$

in the basis where each of the four corners is represented by a four–dimensional vector e_i with $(e_i)_j = \delta_{ij}$ and $i, j = 1, 2, 3, 4$. Furthermore,

$$S = \mathbb{1}_{2\times 2} \otimes \sigma_1 , \tag{2.3a}$$

$$S' = \sigma_1 \otimes \mathbb{1}_{2\times 2} \tag{2.3b}$$

with the standard Pauli matrix σ_1. T generates a $\mathbb{Z}_3$ and S generates a $\mathbb{Z}_2$. In addition, one may allow for orientation–changing operations (with $\det = -1$), for example, generated by $S'' = \mathrm{diag}(\mathbb{1}_{2\times 2}, \sigma_1)$.

Since these generators do not commute, the multiplicative closure yields a non–Abelian discrete symmetry, being S_4. As mentioned in [16], if one restricts the

allowed operations to be contained in proper Lorentz transformations, one arrives at the non–Abelian flavor symmetry generated by T, S and S', which is A_4. We therefore arrive at the premature conclusion that, in a model in which each fixed point carries a state, the family symmetry will be A_4. However, as pointed out in [17] and as we shall see later in more detail the actual symmetry in UV complete settings is larger than that.

In summary, we see that extra dimensions offer a compelling explanation of non–Abelian discrete flavor symmetries. However, as the settings discussed here are based on gauge theories in more than four dimensions, one has to address the question of how to complete them in the UV. We will come back to this question in section 3, where we will see that string models indeed often exhibit non–Abelian discrete family symmetries.

2.3 Discrete R symmetries

In supersymmetric theories there are the so–called R symmetries which, by definition, do not commute with supersymmetries. Such symmetries can originate from extra dimensions as well. Specifically they are (discrete) remnants of the Lorentz symmetry of compact dimensions. The perhaps simplest way of seeing this is by recalling that under Lorentz rotations spinors, vectors and scalars transform differently such that different parts of superfields have different charges. This means, in particular, that R symmetries are deeply connected to the fundamental symmetries of space–time.

Let us illustrate this point in more detail by discussing toy–settings with two compact dimensions (without discussing SUSY breaking). If these dimensions were flat (and infinite) the setup would exhibit an SO(2) rotation symmetry. For instance, this symmetry can be defined by its action on the extra components of the gauge fields,

$$\begin{pmatrix} A_5 \\ A_6 \end{pmatrix} \rightarrow \begin{pmatrix} \cos\zeta & -\sin\zeta \\ \sin\zeta & \cos\zeta \end{pmatrix} \begin{pmatrix} A_5 \\ A_6 \end{pmatrix} . \tag{2.4}$$

Since such components get combined to the scalar component of a chiral superfield, describing a bulk field (or an untwisted sector field in string–derived orbifolds), it is more convenient to recast (2.4) in complex notation,

$$\mathrm{U}(1)_{56} \ : \ A_5 + \mathrm{i}A_6 \ \rightarrow \ \mathrm{e}^{\mathrm{i}\,\zeta}\,(A_5 + \mathrm{i}A_6) \, . \tag{2.5}$$

On the other hand, the spinor component of this 'untwisted superfield' turns out to transform differently under the Lorentz group. To understand this, note that the

4D spinor ρ is contained in the higher–dimensional one (Ψ) according to $\Psi = \rho \otimes \chi$, where χ is a spinorial zero mode in internal space. Recalling that spinors always rotate half as quickly as vectors under Lorentz transformations leads to the transformation law

$$\mathrm{U}(1)_{56} \; : \; \rho \; \to \; \mathrm{e}^{\mathrm{i}\,\zeta/2}\,\rho \; . \tag{2.6}$$

In the 4D superfield

$$\Phi \; = \; \frac{1}{\sqrt{2}}\,(A_5 + \mathrm{i}\,A_6) + \sqrt{2}\,\theta\rho + \theta\theta\,F \tag{2.7}$$

the superspace coordinates θ balance the transformations of the components (2.5) and (2.6), i.e.

$$\mathrm{U}(1)_{56} \; : \; \theta \; \to \; \mathrm{e}^{\mathrm{i}\,\zeta/2}\,\theta \; . \tag{2.8}$$

Hence, $\mathrm{U}(1)_{56}$ originating from the 6D Lorentz symmetry denotes an R symmetry.

It is also clear that typically a compact space does not possess the full Lorentz symmetry. For example, orbifolds can have discrete rotational symmetries and hence can naturally provide discrete R symmetries, see section 3.1.3 for more details in the case of string compactifications on orbifolds.

3 Orbifolds and string selection rules

So far, our discussion was purely bottom–up. It is, however, instructive to comment on the situation in top–down models. The geometrical repetition of families, as briefly discussed in section 2.2, is a common feature of most string compactifications.

1. In heterotic orbifolds, very often families come from so–called twisted sectors, which correspond to states localized at the orbifold fixed points in the extra dimensions. We will discuss the emergent family symmetries in more detail below.

2. In D–brane models (see e.g. [18] for a review) the repetition of families is due to the fact that branes can wrap cycles (i.e. some directions in the extra dimensions) multiple times. Therefore, one can have non–trivial intersection numbers between different branes, leading to otherwise equivalent chiral states localized at the intersections. Therefore such models also generically

exhibit non–trivial family symmetries. Also F theory models have non–trivial family symmetries, which often lead to the problem that the Yukawa couplings have rank one [19].

In what follows, we will focus on the heterotic string compactified on (toroidal) orbifolds. There are two main reasons for this choice. First of all, the heterotic framework gives rise to explicit globally consistent candidate models for physics beyond the standard model [20–23]. Second, at the same time, this scheme is simple enough to fully understand the symmetries. Discrete symmetries can appear mainly in two ways: (i) from the compacification to 4D as remnants of higher dimensional gauge/Lorentz symmetry and (ii) from going to a special vacuum configuration where some of the fields of the 4D effective theory obtain VEVs and hence induce further symmetry breaking. The situation is schematically illustrated in figure 3.

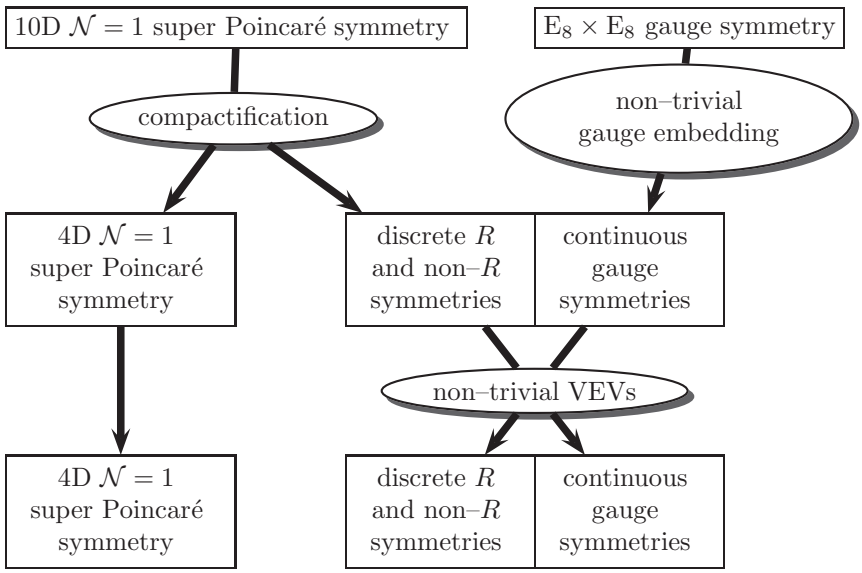


Figure 3: Origin of symmetries in heterotic orbifold compactifications. By compactification of six dimensions and appropriate gauge embedding the 10D super Poincaré and $E_8 \times E_8$ symmetries get broken to the 4D super Poincaré, a 4D gauge and various discrete R and non–R symmetries. The latter two get further broken to subgroups by non–trivial VEVs of certain charged fields.

Orbifolds are six–dimensional compact spaces which, in contrast to a general Calabi–Yau compactification, have additional discrete symmetries which manifest

themselves in the four-dimensional effective theory. Brief introductions to heterotic orbifolds and to the selection rules that govern the allowed terms of the superpotential are given in appendix A.

3.1 Discrete symmetries from string selection rules

3.1.1 Abelian symmetries

In general, there are two possible origins for Abelian (non–R) discrete symmetries in heterotic orbifold compactifications. Either they can arise from the space group selection rule discussed in appendix A.3 or as a discrete remnant of a spontaneously broken gauge symmetry. The second possibility was discussed in section 2.1, the first one will be presented in the following.

For the sake of concreteness, we consider the $\mathbb{S}^1/\mathbb{Z}_2$ orbicircle of section 2. In this case, the space group consists of the elements $\left(\theta^k, m\,e\right)$, where $\theta = -1$ and $k \in \{0,1\}$ describe the $\mathbb{Z}_2$ reflection, $m \in \mathbb{Z}$, $e = 2\pi\,R$ and R denotes the radius of the circle $\mathbb{S}^1$. As illustrated in figure 1, the integer m specifies the location of the twisted state (or 'brane field'), which have $k = 1$, as opposed to untwisted states (or 'bulk fields') which have $k = 0$. The space group selection rule requires the product of space–group elements of the states involved in a coupling to be congruent to identity (see appendix A.3). This gives rise to an Abelian $\mathbb{Z}_2^k \times \mathbb{Z}_2^m$ symmetry, i.e.

$$\prod_{r=1}^{L} g_r \;=\; \mathbb{1} \quad \curvearrowright \quad \begin{cases} \mathbb{Z}_2^k \;:\; \displaystyle\sum_{r=1}^{L} k^{(r)} \;=\; 0 \mod 2\,, \\ \mathbb{Z}_2^m \;:\; \displaystyle\sum_{r=1}^{L} m^{(r)} \;=\; 0 \mod 2\,. \end{cases} \tag{3.1}$$

We will refer to the condition on $k^{(r)}$ as the point group selection rule and to the second one on $m^{(r)}$ as the m–rule.

3.1.2 Non–Abelian symmetries

A particularly interesting situation arises if the two fixed points at $m = 0$ and 1 are equivalent, which happens to be the case unless one introduces a non–trivial background field (either a so–called discrete Wilson line [24] or the B-field (discrete torsion) [25]). In this case there is an additional S_2 permutation symmetry that interchanges $m = 0$ and $m = 1$. As we shall discuss now, together with the $\mathbb{Z}_2^k \times \mathbb{Z}_2^m$

symmetry discussed above in section 3.1.1, this leads to a non–Abelian discrete symmetry D_4 [17, 26].

We combine a state from the fixed point at $m = 0$ and a state from the one at $m = 1$ into a two–dimensional vector, i.e. a doublet. From equation 3.1 we see that the space group selection rule is generated in this basis by the elements

$$-\mathbb{1}_{2\times 2} \;=\; \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} , \quad \sigma_3 \;=\; \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{3.2}$$

i.e. the element $-\mathbb{1}_{2\times 2}$ generates $\mathbb{Z}_2^k$, i.e. the point group selection rule, and σ_3 generates $\mathbb{Z}_2^m$, i.e. the m–rule. The additional element that generates the permutation of the two states is given by

$$\sigma_1 \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{3.3}$$

The multiplicative closure of these three elements yields a non–Abelian group with eight elements $\{\mathbb{1}_{2\times 2}, -\mathbb{1}_{2\times 2}, \pm\sigma_1, \pm\sigma_3, \pm i\sigma_2\}$ and is known as the dihedral group D_4, associated with the symmetry of a square.

Similar to the $\mathbb{S}^1/\mathbb{Z}_2$ case, the $\mathbb{T}^2/\mathbb{Z}_2$ orbifold without Wilson lines (see figure 4) generically has a $(D_4 \times D_4)/\mathbb{Z}_2$ flavor symmetry which originates from the Abelian space group selection rule $\mathbb{Z}_2^3$ combined with the permutation symmetries S and S' of equation 2.3. It can be enhanced further for special values of the angle and the two radii of $\mathbb{T}^2/\mathbb{Z}_2$. For example, when the orbifold geometrically is a tetrahedron the naive geometrical S_4 symmetry obtained from field theory considerations in section 2.2 gets enhanced to SW_4, which has 192 elements, by the stringy space group selection rule [17]. If one allows only for proper Lorentz transformations, one obtains a group with 96 elements which is contained in SW_4. The string description allows us to clarify whether or not one should consider operations which are not contained in the proper Lorentz transformations. The couplings between states localized at different fixed points go like e^{-aT}, where T denotes the Kähler modulus of the corresponding orbifold plane. The real part of T is proportional to R^2, where R is the radius of the underlying torus. Clearly, R^2 does not change under these extra reflections, such that the absolute values of the coupling strengths will enjoy the larger symmetry SW_4. On the other hand, the imaginary part of T, the so–called T–axion, is related to the anti–symmetric tensor field in compact space, and does change its sign under the extra reflections. Hence, if the T–axion acquires a non–trivial VEV, the phases of the coupling strengths do no longer enjoy the larger symmetries. As is well known, breaking

the reflection symmetries in internal space can be related to CP violation in the effective 4D theory (cf. [27, 28]), and what we discussed here is just an example for this statement. Note that there are different possibilities to obtain non–trivial CP phases, also based on non–Abelian discrete symmetries (cf. [29]). It should be interesting to see if these also have an interpretation in terms of reflection symmetries in compact space.

Different lower–dimensional building blocks of orbifolds lead to other non–Abelian discrete symmetries (table 3.1).

orbifold	flavor symmetry	sector	string fundamental states
$\mathbb{S}^1/\mathbb{Z}_2$	D_4	U	$\mathbf{1}$
		T_1	$\mathbf{2}$
$\mathbb{T}^2/\mathbb{Z}_2$	$(D_4 \times D_4)/\mathbb{Z}_2$	U	$\mathbf{1}$
		T_1	$\mathbf{4}$
$\mathbb{T}^2/\mathbb{Z}_3$	$\Delta(54)$	U	$\mathbf{1}$
		T_1	$\mathbf{3}$
		T_2	$\bar{\mathbf{3}}$
$\mathbb{T}^2/\mathbb{Z}_4$	$(D_4 \times \mathbb{Z}_4)/\mathbb{Z}_2$	U	$\mathbf{1}$
		T_1	$\mathbf{2}$
		T_2	$\mathbf{1}_{A_1} + \mathbf{1}_{B_1} + \mathbf{1}_{B_2} + \mathbf{1}_{A_2}$
		T_3	$\mathbf{2}$
$\mathbb{T}^4/\mathbb{Z}_8$	$(D_4 \times \mathbb{Z}_8)/\mathbb{Z}_2$	U	$\mathbf{1}$
		T_1	$\mathbf{2}$
		T_2	$\mathbf{1}_{A_1} + \mathbf{1}_{B_1} + \mathbf{1}_{B_2} + \mathbf{1}_{A_2}$
		T_3	$\mathbf{2}$
		T_4	$4 \times (\mathbf{1}_{A_1} + \mathbf{1}_{B_1} + \mathbf{1}_{B_2} + \mathbf{1}_{A_2})$
$\mathbb{T}^4/\mathbb{Z}_{12}$	trivial		
$\mathbb{T}^6/\mathbb{Z}_7$	$S_7 \ltimes (\mathbb{Z}_7)^6$	U	$\mathbf{1}$
		T_k	$\mathbf{7}$
		T_{7-k}	$\bar{\mathbf{7}}$

Table 3.1: Survey of flavor symmetries arising from building blocks of orbifolds (from [17]). The T_k denote the various twisted sectors and U the untwisted sector.

The (non–Abelian) flavor symmetry could be broken in two ways: (i) explicitly: the presence of orbifold Wilson lines breaks the permutation symmetry, at least partially. If the permutation symmetry is completely broken, the remaining flavor

group is Abelian. (ii) spontaneously: by the VEV of some twisted field, since twisted fields necessarily transform in a non–singlet representation under the flavor group. For example, $\Delta(54)$ can be broken to S_3 by the VEV of a triplet $\mathbf{3}$ (e.g. $\langle \mathbf{3} \rangle = (v, v, v)$).

3.1.3 *R* symmetries

As already mentioned, discrete R symmetries could arise as discrete remnants of the Lorentz symmetry of compact dimensions. This is also true for string–derived orbifold models.

What are the R charges of states localized at the fixed points? In the framework of field theory one cannot answer this question unambiguously. For instance, in many field–theoretic analysis the profiles of these fields are taken to be δ–functions with support at the fixed points, from which one may conclude that the states transform trivially under the discrete R symmetries. It turns out that the naive field–theoretic expectation is incorrect. However, in string theory one can address this question. Specifically, in heterotic orbifolds the R symmetries derive from the H–momentum conservation law and one can determine the R charges unambiguously. We will discuss an explicit example in section 4.2.

3.2 Discrete symmetries in explicit models

Having seen how discrete symmetries arise in the effective field–theoretic description, we will now discuss which symmetries appear in explicit string models.

3.2.1 Flavor symmetries

In recent years, many MSSM candidate models have emerged from heterotic orbifolds [20, 30–32], known as the "heterotic mini–landscape". These models have a common flavor structure: focusing on the two–torus where a $\mathbb{Z}_2$ acts, the two light generations are localized on equivalent fixed points and the third one is in the bulk. Therefore, as discussed above, there is a D_4 flavor symmetry, under which the two light generations transform as a doublet whereas the the third family transforms trivially (Let us mention that there are also alternative models without this D_4 [21, 33]). This symmetry is broken in potentially realistic vacua by the VEVs of some localized singlets. Yet, using the D_4 symmetric situation as a starting point and then considering corrections can have certain advantages when discussing the (supersymmetric) flavor structure (cf. [34]). The emerging scheme

is somewhat similar to the one of 'minimal flavor violation' [35–37]. In particular, the structure of the soft masses is

$$\widetilde{m}^2 \;=\; \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{pmatrix} + \text{terms proportional to } D_4 \text{ breaking VEVs} \; . \qquad (3.4)$$

It is known that such an approximate form of the soft masses makes it possible to avoid the supersymmetric flavor problems. In addition, it naturally allows for scenarios in which the third family of squarks and sleptons is substantially lighter than the first two generations of superpartners (cf. the discussion in [38]).

3.2.2 Flavor–independent symmetries

In grand unified models, matter or R parity can be obtained from baryon–minus–lepton–number symmetry $\mathrm{U}(1)_{B-L}$ by spontaneous breaking, and the same is true in string–derived models [32], the only difference being that $\mathrm{U}(1)_{B-L}$ is not in GUT normalization and no large representations (such as $\overline{\mathbf{126}}$–plets of SO(10)) are required (nor available) to achieve the breaking $\mathrm{U}(1)_{B-L} \to \mathbb{Z}_2^{\mathcal{M}}$. That is, string theory avoids huge representations like the $\overline{\mathbf{126}}$–plets, but still allows us to derive matter parity from a local $B-L$ symmetry.

Similarly, proton hexality can be obtained from Pati–Salam (PS) times an extra U(1) symmetry [39]. Explicit orbifold models from $\mathbb{Z}_4 \times \mathbb{Z}_4$ compactifications using a local GUT approach,

$$\mathrm{E}_8 \text{ in 10D} \to \mathrm{SO}(12) \text{ in 6D} \to \mathrm{PS} \times \mathrm{U}(1) \to \mathrm{SM} \text{ in 4D} \; , \qquad (3.5)$$

revealed 850 heterotic MSSMs (i.e. three generations of quarks and leptons plus vector–like exotics), many of them with the correct proton hexality charge assignment for at least some quarks and leptons [39].

3.2.3 R symmetries

R symmetries play an important role in string models. In particular, approximate continuous R symmetries, which derive from exact discrete R symmetries, can explain the large hierarchy between the Planck, GUT and/or string scales on the one hand and the electroweak and/or supersymmetry breaking scales on the other hand. It has been demonstrated that, in the presence of a continuous R symmetry, at field configurations that satisfy the F–term constraints, the VEV of the superpotential vanishes [11]. If there is an approximate R symmetry that

gets explicitly broken at some high order N, the vacuum expectation value of the superpotential, or equivalently the gravitino mass $m_{3/2}$, goes like

$$\langle \mathscr{W} \rangle \ \sim \ \langle s \rangle^N \ , \tag{3.6}$$

where $\langle s \rangle$ denotes a typical size of a VEV of fields that break the symmetry spontaneously (in Planck units) and N is of the order 10 in explicit examples. Further, in the context of the MSSM it has been shown that in settings in which matter charges are consistent with grand unification, the only anomaly–free symmetries that can forbid the μ term are R symmetries [40]. Given that $\langle \mathscr{W} \rangle$, or, equivalently $m_{3/2}$, is the order parameter of R symmetry breaking, this yields a relation between μ and $m_{3/2}$ [11, 41], i.e. constitutes a solution to the μ problem. This solution does, unlike the Giudice–Masiero mechanism [42], not rely on a specific structure of the Kähler potential, rather it provides a holomorphic μ term of the right size, similar to the Kim–Nilles picture [43].

4 Anomaly Freedom

4.1 Anomaly constraints vs. embedding constraints

How can one derive anomaly constraints on discrete symmetries? It is instructive to review how they have been derived in the past. Ibáñez and Ross [44] have used the following strategy: they have obtained $\mathbb{Z}_N$ symmetries from U(1) by spontaneous breaking, as discussed in section 2.1. It is obvious that, if the U(1) is non–anomalous, and the spontaneous breaking is done consistently, then also $\mathbb{Z}_N$ is anomaly–free. However, one may question whether these are in general true anomaly constraints or rather embedding constraints, i.e. constraints that restrict the choice of the non–anomalous continuous gauge group into which the discrete group is supposed to be embedded.

Araki [45] proposed an alternative derivation of the anomaly constraints, which does not rely on embedding the discrete symmetry into a continuous one, but by using the path integral method [46]. This strategy has been applied to the $\mathbb{Z}_N$ case [47] with the result that all Ibáñez–Ross constraints apply except for the $\mathbb{Z}_N^3$ ones, which are known not to constitute true anomaly constraints [40, 48].

Also discrete anomaly constraints for non–Abelian discrete symmetries have first been derived by using the embedding strategy [49] (see [50] for a more recent discussion). While, again, these constraints ensure anomaly freedom, they turn out to be, in general, not true anomaly constraints but rather embedding constraints.

That is, if these constraints are satisfied, the symmetry is anomaly free, but the converse is not necessarily true. In particular, the constraints can depend on the choice of the continuous symmetry into which the discrete one is supposed to be embedded. The true constraints can be derived with the path integral method [45], and one finds that one only has to check anomaly freedom for the Abelian subgroups of a given non–Abelian symmetry [45,47]. For a discrete group D and a continuous gauge symmetry G one obtains the conditions that

$$\sum_{(\boldsymbol{r}^{(f)},\boldsymbol{d}^{(f)})} \delta^{(f)} \cdot \ell(\boldsymbol{r}^{(f)}) \stackrel{!}{=} 0 \mod \frac{N}{2} , \tag{4.1}$$

where the sum '$\sum_{(\boldsymbol{r}^{(f)},\boldsymbol{d}^{(f)})}$' is over representations which are non–trivial w.r.t. to both G and D. The discrete Abelian charge, denoted by $\delta^{(f)}$, can be expressed in terms of the group elements $U(\boldsymbol{d}^{(f)})$ as

$$\delta^{(f)} = N \frac{\ln \det U(\boldsymbol{d}^{(f)})}{2\pi \mathrm{i}} . \tag{4.2}$$

For the mixed gravitational–D anomaly one finds

$$\sum_{\boldsymbol{d}^{(f)}} \delta^{(f)} \stackrel{!}{=} 0 \mod \frac{N}{2} , \tag{4.3}$$

where the symbol '$\sum_{\boldsymbol{d}^{(f)}}$' means that the sum extends over all non–trivial representations $\boldsymbol{d}^{(f)}$ of D. What does it mean if a given discrete symmetry does not satisfy these constraints? In general, one may argue that in such a case the symmetry will be broken in an uncontrollable way and all the predictive power of the (discrete) symmetry will be lost. For useful applications in particle physics, reliable discrete symmetries should thus be anomaly free. There is, however, an exception: for the anomalous symmetry the anomalies might be cancelled (microscopically) by a discrete Green-Schwarz mechanism. In what follows, we shall discuss this possibility in detail.

4.2 Non–perturbative "violation" of discrete symmetries and discrete Green–Schwarz anomaly cancellation

As in the case of continuous symmetries, discrete anomalies can be cancelled by a Green–Schwarz (GS) mechanism (for a discussion in the path integral formalism see [40]). Also here this requires the presence of a scalar, the GS axion, which

multiplies some $F_{\mu\nu}\widetilde{F}^{\mu\nu}$ terms (with $F^{\mu\nu}$ denoting the field strength of some continuous gauge symmetry of the model), and shifts under the discrete symmetry. Once the axion acquires its vacuum expectation value, the discrete symmetry gets broken spontaneously. Effectively this leads to a situation in which the (anomalous part of the) discrete group appears to be broken by non–perturbative effects.[1]

As an example, consider the $\mathbb{Z}_4^R$ symmetry discussed in [8, 40]. It forbids the μ term and dimension 4 and 5 proton decay operators at the perturbative level. It appears to be broken by non–perturbative effects to its 'non–anomalous' subgroup, i.e. to $\mathbb{Z}_2^{\mathcal{M}}$ matter parity. The order parameter of this R symmetry breaking is the vacuum expectation value of the superpotential, i.e. the gravitino mass. One therefore has, in the context of gravity mediation, a μ term of the correct size (cf. the analogous discussion in section 3.2.3) while dimension five proton decay remains far below the experimental limits.

Similar to the case of R symmetries, also non–R symmetries can appear anomalous and hence be broken non–perturbatively. This, again, introduces a hierarchically small breaking of the discrete symmetry. It remains to be seen whether this mechanism can provide us with solutions to some of the open questions in flavor physics.

5 Summary

The flavor structure of the SM remains one of the greatest puzzles in particle physics. Flavor symmetries appear to be instrumental for solving this puzzle. Optimistically one may hope to find a compelling model that explains the observed flavor structure. In this case the question where the underlying family symmetries originate from is of greatest importance since given a deeper understanding may allow us to relate the observed fermion masses and mixing to some fundamental properties of our world.

In this paper we have reviewed the possible origin of discrete symmetries, paying particular attention to discrete flavor symmetries. Discrete symmetries can arise from continuous symmetries by spontaneous breaking or from extra dimensions. While for Abelian symmetries the first option is a very common tool in

[1] Non–perturbative effects generate couplings of the form $\exp(-\mathrm{i}a)\phi_1 \ldots \phi_n$, where a denotes the GS axion and the ϕ_i some (matter) fields of the theory. Such terms are invariant under the full discrete group when one takes the shift transformation of the GS axion a into account. But, when a obtains a vacuum expectation value, the ('anomalous' part of the) discrete group is broken spontaneously.

model building, we have argued that obtaining non–Abelian discrete symmetries from continuous ones (in four dimensions) does not lead to compelling models. On the other hand, non–Abelian discrete symmetries do arise in models with extra dimensions, where they are deeply connected to the explanation of the repetition of families. In particular, in stringy extensions of the standard model such symmetries often arise. Therefore they can play an important role in understanding or addressing the flavor puzzle in the standard model as well as in solving flavor problems in extensions such as the MSSM.

We have also commented on discrete anomalies, which constrain possible discrete symmetries in bottom–up model building. As we have pointed out, one should carefully distinguish between embedding constraints and true anomaly constraints. Discrete symmetries that appear anomalous open very attractive possibilities in model building as they appear to be broken non–perturbatively, i.e. the breaking can be hierarchically small. This observation has been applied to the μ parameter of the MSSM. It remains to be seen whether hierarchies in flavor physics can have a similar explanation.

Acknowledgments

One of us (M.R.) would like to thank Mu–Chun Chen for useful discussions and the UC Irvine, where part of this work was done, for hospitality. This work was partially supported by the SFB–Transregio TR33 "The Dark Universe" (Deutsche Forschungsgemeinschaft), the SFB 676, the European Union 7th network program "Unification in the LHC era" (PITN–GA–2009–237920) and the DFG cluster of excellence "Origin and Structure of the Universe" (Deutsche Forschungsgemeinschaft).

A Orbifolds

We give a brief introduction to orbifolds following [51, 52]. We start with the geometrical construction in appendix A.1. In appendix A.2 we depict how heterotic strings are compactified on orbifolds and appendix A.3 reviews string selection rules.

A.1 Construction of orbifolds

From the geometrical point of view, a d–dimensional (toroidal) orbifold is defined as the quotient of $\mathbb{R}^d$ divided by a discrete group S, called the space group. For $\mathbb{Z}_N$ orbifolds, the elements of the space group $g \in S$ are given by

$$g \;=\; \left(\theta^k, n_\alpha\, e_\alpha\right) \qquad \text{and act as} \qquad g\,X \;=\; \theta^k\, X + n_\alpha\, e_\alpha \;, \tag{A.1}$$

with sum over $\alpha = 1, \ldots, d$ and $X \in \mathbb{R}^d$. The d (linearly independent) vectors e_α generate a lattice Γ and hence define a torus $\mathbb{T}^d$. The rotation θ is of order N (i.e. $\theta^N = \mathbb{1}$) and is chosen to be an automorphism of Γ. Then the action of S is not free, i.e. there are fixed points $X_g \in \mathbb{R}^d$ with $gX_g = X_g$ for some $g \in S$. The space group element g associated to the fixed point X_g is called the constructing element, see figure 4. The resulting orbifold is written as $\mathbb{T}^d/\mathbb{Z}_N$.

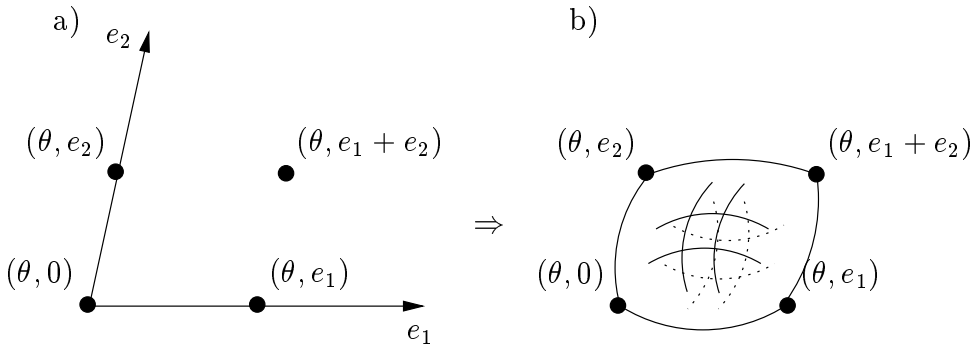


Figure 4: a) The four fixed points labeled by their constructing elements for a two–dimensional $\mathbb{T}^2/\mathbb{Z}_2$ example (i.e. with $\theta = -\mathbb{1}_{2\times 2}$). b) The orbifold of a) folded up to a pillow–like object with fixed points at the corners of the pillow.

A.2 Strings on orbifolds

Compactifying the heterotic string on six–dimensional orbifolds yields three different classes of closed strings: (i) untwisted strings with constructing element $(\mathbb{1}, 0)$ which would also close in uncompactified space, (ii) winding modes with constructing elements $(\mathbb{1}, n_\alpha e_\alpha)$ which would also close on the torus and (iii) twisted strings, localized at the fixed points, with constructing elements $\left(\theta^k, n_\alpha e_\alpha\right)$ with $k \neq 0$ which only close on the orbifold due to the θ rotation. The winding modes are massive with masses near the Planck scale. Since we are only interested in the low–energy effective action they are ignored in the following.

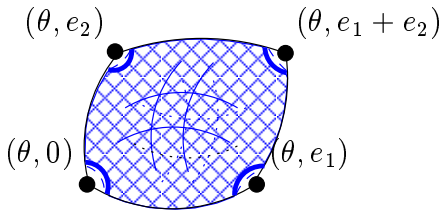


Figure 5: Visualization of the space group selection rule for four twisted states, indicated by the four bold strings at the four fixed points, see figure 4. The condition $(\theta,0)\,(\theta,e_1)\,(\theta,e_2)\,(\theta,e_1+e_2) = (\mathbb{1},(\mathbb{1}-\theta)\Gamma)$ is satisfied and hence the coupling is allowed.

The geometrical action of the space group has to be amended by an action on the gauge degrees of freedom of the heterotic string in order to fulfill the stringy consistency conditions of modular invariance. In the standard approach this is achieved by so called shifts and Wilson lines. Specifying these input parameters completely defines an orbifold compactification and allows one to compute the massless spectrum. An elegant way to obtain consistent orbifold models, for example MSSM–like models, to compute their massless spectra and to analyze their resulting four–dimensional effective theories is given by the public code "Orbifolder" [53].

A.3 String selection rules

The CFT description allows one to compute scattering amplitudes of strings on orbifolds. In the four–dimensional effective theory these amplitudes enter as coupling strengths of allowed terms in the superpotential. Their computation is technically involved. Hence, at a first step one is only interested in the string selection rules determining which coupling is allowed or forbidden. In many cases the string selection rules can be interpreted as a symmetry of the four–dimensional effective theory. The (standard) string selection rules are:

1. Gauge invariance

2. Space group selection rule: The space group selection rule reflects the geometrical possibility of orbifold strings to join. Consider L strings with constructing elements $g_r = \left(\theta^{k^{(r)}}, n_\alpha^{(r)} e_\alpha\right)$. Then the coupling is allowed if $\prod_{r=1}^{L} g_r = \mathbb{1}$, see figure 5.

3. R charge conservation: R charge conservation is a discrete remnant of ten–dimensional Lorentz symmetry. It arises whenever the orbifold $\mathbb{R}^6/S$ respects some additional rotational symmetry beside $\theta = \mathrm{diag}(\mathrm{e}^{2\pi\mathrm{i}\,v_1}, \mathrm{e}^{2\pi\mathrm{i}\,v_2}, \mathrm{e}^{2\pi\mathrm{i}\,v_3})$. For example, for a factorized orbifold, i.e. an orbifold whose lattice Γ is the direct product of three two–dimensional lattices $\Gamma = \Gamma_1 \times \Gamma_2 \times \Gamma_3$, a rotation in the sublattice Γ_i by $\mathrm{e}^{2\pi\mathrm{i}\,v_i}$ is a symmetry of the theory. The rotation by v_i is of order N_i (i.e. $N_i v_i \in \mathbb{Z}$) and results in a $\mathbb{Z}^R_{2N_i}$ symmetry

$$\sum_{r=1}^{L} -2R^i_r \;=\; 2 \mod 2N_i \,, \tag{A.2}$$

where $R^i_r = q^i_{\mathrm{sh},r} - \tilde{N}^i_r + \tilde{N}^{\bar{\imath}}_r$ with the oscillator numbers $\tilde{N}^i_r$ and $\tilde{N}^{\bar{\imath}}_r$ (see e.g. [31] for their definition), $q_{\mathrm{sh},r}$ are the bosonic right–moving momenta and the factor -2 originates from the normalization such that fermions have a shifted R charge by -1.

If the two–dimensional lattice Γ_i has a higher symmetry than N_i there is an additional string selection rule known as "rule 4". For example, the $\mathrm{SU}(3)^3$ root lattice of a $\mathbb{Z}_3$ orbifold allows for $\mathbb{Z}_6$ sublattice rotations. If all strings involved in a given interaction sit at the same fixed point they feel the higher symmetry and the R symmetry is enhanced to $\mathbb{Z}^R_{4N_i}$.

References

[1] T. Banks and N. Seiberg, Phys.Rev. **D83** (2011), 084019, `arXiv:1011.5120` [hep-th].

[2] H. Ishimori, T. Kobayashi, H. Ohki, Y. Shimizu, H. Okada, et al., Prog.Theor.Phys.Suppl. **183** (2010), 1, `arXiv:1003.3552` [hep-th].

[3] Origin of family symmetries, editors J. Valle et al., Fortschritte der Physik (2012 to appear).

[4] P. Fayet, Phys. Lett. **B69** (1977), 489.

[5] S. Dimopoulos, S. Raby, and F. Wilczek, Phys. Lett. **B112** (1982), 133.

[6] L. E. Ibáñez and G. G. Ross, Nucl. Phys. **B368** (1992), 3.

[7] H. K. Dreiner, C. Luhn, and M. Thormeier, Phys. Rev. **D73** (2006), 075007, hep-ph/0512163.

[8] H. M. Lee, S. Raby, M. Ratz, G. G. Ross, R. Schieren, K. Schmidt-Hoberg, and P. K. Vaudrevange, Phys.Lett. **B694** (2011), 491, arXiv:1009.0905 [hep-ph].

[9] K.-S. Choi, I.-W. Kim, and J. E. Kim, JHEP **03** (2007), 116, hep-ph/0612107.

[10] K.-S. Choi, H. P. Nilles, S. Ramos-Sánchez, and P. K. Vaudrevange, Phys.Lett. **B675** (2009), 381, arXiv:0902.3070 [hep-th].

[11] R. Kappl, H. P. Nilles, S. Ramos-Sánchez, M. Ratz, K. Schmidt-Hoberg, and P. K. Vaudrevange, Phys. Rev. Lett. **102** (2009), 121602, arXiv:0812.2120 [hep-th].

[12] B. Petersen, M. Ratz, and R. Schieren, JHEP **08** (2009), 111, arXiv:0907.4049 [hep-ph].

[13] A. Adulpravitchai, A. Blum, and M. Lindner, JHEP **0909** (2009), 018, arXiv:0907.2332 [hep-ph].

[14] C. Luhn, JHEP **1103** (2011), 108, arXiv:1101.2417 [hep-ph].

[15] A. Merle and R. Zwicky, JHEP **1202** (2012), 128, arXiv:1110.4891 [hep-ph].

[16] G. Altarelli, F. Feruglio, and Y. Lin, Nucl.Phys. **B775** (2007), 31, arXiv:hep-ph/0610165 [hep-ph].

[17] T. Kobayashi, H. P. Nilles, F. Plöger, S. Raby, and M. Ratz, Nucl. Phys. **B768** (2007), 135, hep-ph/0611020.

[18] R. Blumenhagen, B. Körs, D. Lüst, and S. Stieberger, Phys. Rept. **445** (2007), 1, hep-th/0610327.

[19] J. J. Heckman and C. Vafa, Nucl.Phys. **B837** (2010), 137, arXiv:0811.2417 [hep-th].

[20] O. Lebedev, H. P. Nilles, S. Raby, S. Ramos-Sánchez, M. Ratz, P. K. S. Vaudrevange, and A. Wingerter, Phys. Lett. **B645** (2007), 88, hep-th/0611095.

[21] O. Lebedev, H. P. Nilles, S. Ramos-Sánchez, M. Ratz, and P. K. S. Vaudrevange, Phys. Lett. **B668** (2008), 331, arXiv:0807.4384 [hep-th].

[22] L. B. Anderson, J. Gray, A. Lukas, and E. Palti, Phys.Rev. **D84** (2011), 106005, arXiv:1106.4804 [hep-th], 19 pages. References Added.

[23] L. B. Anderson, J. Gray, A. Lukas, and E. Palti, arXiv:1202.1757 [hep-th].

[24] L. E. Ibáñez, H. P. Nilles, and F. Quevedo, Phys. Lett. **B187** (1987), 25.

[25] C. Vafa, Nucl. Phys. **B273** (1986), 592.

[26] L. J. Dixon, D. Friedan, E. J. Martinec, and S. H. Shenker, Nucl. Phys. **B282** (1987), 13.

[27] M. Dine, R. G. Leigh, and D. A. MacIntire, Phys.Rev.Lett. **69** (1992), 2030, arXiv:hep-th/9205011 [hep-th].

[28] T. Kobayashi and O. Lebedev, Phys. Lett. **B565** (2003), 193, hep-th/0304212.

[29] M.-C. Chen and K. Mahanthappa, Phys.Lett. **B681** (2009), 444, arXiv:0904.1721 [hep-ph].

[30] W. Buchmüller, K. Hamaguchi, O. Lebedev, and M. Ratz, Phys. Rev. Lett. **96** (2006), 121602, hep-ph/0511035.

[31] W. Buchmüller, K. Hamaguchi, O. Lebedev, and M. Ratz, Nucl. Phys. **B785** (2007), 149, hep-th/0606187.

[32] O. Lebedev, H. P. Nilles, S. Raby, S. Ramos-Sánchez, M. Ratz, P. K. S. Vaudrevange, and A. Wingerter, Phys. Rev. **D77** (2007), 046013, arXiv:0708.2691 [hep-th].

[33] J. E. Kim, J.-H. Kim, and B. Kyae, JHEP **06** (2007), 034, hep-ph/0702278.

[34] P. Ko, T. Kobayashi, J.-h. Park, and S. Raby, Phys. Rev. **D76** (2007), 035005, arXiv:0704.2807 [hep-ph].

[35] R. S. Chivukula and H. Georgi, Phys. Lett. **B188** (1987), 99.

[36] A. J. Buras, P. Gambino, M. Gorbahn, S. Jäger, and L. Silvestrini, Phys. Lett. **B500** (2001), 161, hep-ph/0007085.

[37] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, Nucl. Phys. **B645** (2002), 155, hep-ph/0207036.

[38] S. Krippendorf, H. P. Nilles, M. Ratz, and M. W. Winkler, arXiv:1201.4857 [hep-ph].

[39] S. Förste, H. P. Nilles, S. Ramos-Sánchez, and P. K. Vaudrevange, Phys.Lett. **B693** (2010), 386, arXiv:1007.3915 [hep-ph].

[40] H. M. Lee, S. Raby, M. Ratz, G. G. Ross, R. Schieren, et al., Nucl.Phys. **B850** (2011), 1, arXiv:1102.3595 [hep-ph].

[41] F. Brümmer, R. Kappl, M. Ratz, and K. Schmidt-Hoberg, JHEP **04** (2010), 006, arXiv:1003.0084 [hep-ph].

[42] G. F. Giudice and A. Masiero, Phys. Lett. **B206** (1988), 480.

[43] J. E. Kim and H. P. Nilles, Phys. Lett. **B138** (1984), 150.

[44] L. E. Ibáñez and G. G. Ross, Phys. Lett. **B260** (1991), 291.

[45] T. Araki, Prog. Theor. Phys. **117** (2007), 1119, hep-ph/0612306.

[46] K. Fujikawa, Phys. Rev. Lett. **42** (1979), 1195.

[47] T. Araki et al., Nucl. Phys. **B805** (2008), 124, arXiv:0805.0207 [hep-th].

[48] T. Banks and M. Dine, Phys. Rev. **D45** (1992), 1424, hep-th/9109045.

[49] P. H. Frampton and T. W. Kephart, Int. J. Mod. Phys. **A10** (1995), 4689, hep-ph/9409330.

[50] C. Luhn and P. Ramond, JHEP **0807** (2008), 085, arXiv:0805.1736 [hep-ph].

[51] P. K. S. Vaudrevange, arXiv:0812.3503 [hep-th].

[52] S. Ramos-Sánchez, Fortsch.Phys. **10** (2009), 907, arXiv:0812.3560 [hep-th], Ph.D.Thesis (Advisor: H.P. Nilles).

[53] H. P. Nilles, S. Ramos-Sánchez, P. K. Vaudrevange, and A. Wingerter, Comput.Phys.Commun. **183** (2012), 1363, arXiv:1110.5229 [hep-th], 29 pages, web page http://projects.hepforge.org/orbifolder/.