**Dissertation**

**submitted to the**

**Combined Faculties of the Natural Sciences and Mathematics**

**of the Ruperto-Carola-University of Heidelberg, Germany**

**for the degree of**

**Doctor of Natural Sciences**


**Put forward by**

*Philipp Henkenjohann*

*born in: Bielefeld, Germany*


*Oral examination: 22 May 2019*

# Quantum Gravity and Axions

*Referees:*

Prof. Dr. Arthur Hebecker

Prof. Dr. Jörg Jäckel

## Quantengravitation und Axione

In dieser Arbeit untersuchen wir die *Weak Gravity Conjecture* (WGC) und die *Swampland Distance Conjecture* (SDC) im Zusammenhang mit Axionen. Sie legen nahe, dass Theorien von Axionen mit super-Planck'schen Feldbereichen, wie zum Beispiel Modelle von *Large-Field* Inflation, nicht kompatibel mit Quantengravitation sind.

Wir stellen diese Aussage in Frage, indem wir Axione mit marginal super-Planck'schen Feldbereichen in einer Kompaktifizierung von Typ IIB Stringtheorie mit Flüssen konstruieren. In der zugehörigen 4d effektiven Theorie sind große Felddistanzen nur zum Preis eines exponentiell niedrigen Cutoffs möglich. Dies ist qualitativ ähnlich zur SDC.

Des Weiteren betrachten wir effektive Theorien von Axionen mit super-Planck'scher Zerfallskonstante, welche daher die WGC verletzen. Axionische Strings führen zu topologischer Inflation, was als Pathologie dieser Theorie aufgefasst werden kann. Dies legt nahe die naive magnetische WGC ernst zu nehmen. Jedoch können wir keine endgültigen Schlüsse ziehen.

Schließlich untersuchen wir axionische Verschiebungssymmetrien, die unserer Erwartung nach von Effekten der Quantengravitation gebrochen werden sollten. Wir stellen fest, dass bestimme fermionische Operatoren eine solche Brechung verursachen, und quantifizieren das Verbot von exakten axionischen Verschiebungssymmetrien durch eine spekulative untere Schranke an Axionmassen. Außerdem diskutieren wir 3-Formen als effektive Beschreibungen von Instantonen und Fermionwechselwirkungen, welche von gravitationellen Instantonen generiert werden.

## Quantum Gravity and Axions

In this thesis we study the Weak Gravity Conjecture (WGC) and the Swampland Distance Conjecture (SDC) in the context of axions. They suggest that theories containing axions with super-Planckian field ranges, such as models of large field inflation, are incompatible with quantum gravity.

We challenge this statement by constructing axions with mildly super-Planckian field ranges in a compactification of type IIB string theory with fluxes. Large field distances in the corresponding 4d effective theory are only possible at the expense of an exponentially low cutoff. This is in spirit consistent with the SDC.

Furthermore we study effective theories of axions with super-Planckian decay constants which hence violate the WGC. Axionic strings lead to topological inflation which may be considered a pathology of such theories. This suggests to take the naive magnetic WGC for axions seriously but we can not draw definite conclusions.

Finally we investigate axionic shift symmetries which are expected to be broken by quantum gravitational effects. We point out that certain fermion operators break such symmetries and quantify the censorship of axionic shift symmetries via

a conjectured lower bound on axion masses. Besides we discuss 3-forms as effective descriptions of instantons and fermion interactions generated by gravitational instantons.

# Acknowledgements

First of all I would like to thank my supervisor Arthur Hebecker who did an exceptionally good job in the last three years. I have learned so much from his intuitive but precise way of thinking about physics which becomes particularly clear to me whenever I compare the physicist I was three years ago and the physicist I am today. His feedback on my work was invaluable to me and I am especially grateful for this. It was always a pleasure and inspiring to work with him.

Besides Arthur, I thank my collaborators Lukas Witkowski and Gia Dvali. My first two projects have greatly benefited from the numerous discussions with Lukas as well as from his creative ideas and expertise. Gia contributed crucial ideas to my third project and the conversations with him were important for our understanding of the involved physics.

I am grateful to Martin Bies, Sascha Leonhardt, Patrick Mangat, Christian Reichelt, Fabrizio Rompineve, Pablo Soler and Fengjun Xu for many helpful discussions and for patiently answering even my most naive questions. Special thanks go to my officemates Martin, Fabrizio, Christian and Fengjun whose fellowship I enjoyed very much. Moreover, I thank all the other people who contributed to the great atmosphere at Philosophenweg 19: Florent Baume, Julius Eckhard, Daniel Junghans, Daniel Kläwer, Craig Lawrie, Ling Lin, Thomas Mikhail, Jakob Moritz, Eran Palti, Andreas Schachner, Moritz Schöne and Timo Weigand.

When it comes to organizational and administrative issues I could always count on the support of our secretaries Sonja Bartsch, Cornelia Merkel and Melanie Steiert. I always enjoyed to work with them, thank you!

I would like to express my gratitude to Timo Weigand and Jörg Jäckel who were my second and third supervisors. In particular, Jörg Jäckel was willing to be the second referee of this thesis. Furthermore I would like to thank Björn Malte Schäfer and Ulrich Uwer for being part of my examination committee.

I thank Sascha Leonhardt for proofreading and helpful comments on parts of this thesis.

I am grateful for the financial support of the Heidelberg Graduate School of Fundamental Physics and the Institute for Theoretical Physics at Heidelberg University. I especially thank Sandra Klevansky and Eduard Thommes for their help in these matters.

Finally, I thank my family, Ulla, Christian, Mark and Lotte, for their constant and invaluable support during my PhD. I am especially grateful to Lotte, for her patience and all the compromises she had to make for me.

# Contents

# 1 Introduction

The $20^{\text{th}}$ century was an exceptionally fruitful period of time for fundamental physics, both experimentally and theoretically. From the theoretical point of view one can barely overestimate the importance of the development of quantum field theory (QFT) and general relativity (GR). Indeed, the standard model of particle physics (SM) as well as the $\Lambda$CDM model of cosmology are formulated within the framework of QFT and GR. They correctly describe a large fraction of observed phenomena on scales ranging from $10^{26}$ m, the scale of the observable universe, down to microscopic scales of $10^{-20}$ m, which are probed at the Large Hadron Collider at CERN. This range covers impressive 46 orders of magnitude.

However, the utility of QFT and GR goes beyond their application in the SM and $\Lambda$CDM model. In a sense they capture the essence of our knowledge about the most fundamental principles of nature. This is one reason why most of the current research on physics beyond the SM (BSM) is done in the framework of QFT and GR. If one is not directly interested in a truly fundamental description of nature and is ignorant of physics beyond some energy scale $\Lambda$, the concept of effective field theories (EFTs) has proven to be very flexible and powerful to describe low energy physics. While any QFT that is supposed to be valid at all energy scales has to be renormalizable in order to be predictive, an EFT is by definition only meant to be an approximate description of nature which breaks down beyond a characteristic energy scale $\Lambda$, the cutoff. Even though the SM is in fact renormalizable it is oftentimes only interpreted as an EFT.

The principle of naturalness applied to the electroweak hierarchy problem of the SM has been the primary guideline in model building for decades [1–3]. This has led to many different BSM models such as technicolor [4], low-scale supersymmetry (SUSY) [5], extra dimensions [6, 7], and the little Higgs [8]. However, with the absence of new BSM particles at the LHC these models are considerably constraint by experimental data [9] and in particular do not solve the electroweak hierarchy problem anymore. These negative results have led some theorists to reexamine the principle of naturalness itself [10–12]. In any case, right now it seems like the most reliable principles we can refer to in model building are simplicity and mathematical consistency. This is not much and without good experimental or theoretical guidance at hand it is very difficult to favor one particular model over the other if both correctly reproduce experimental data.

## 1.1 String Theory

Besides concrete phenomenological problems there is the more fundamental question how QFT and GR can be combined to give a consistent theory of quantum gravity (QG). One of the best understood and promising candidate theories of QG is (super-)string theory which is based on the quantum theory of a string propagating in spacetime. So far five different consistent string theories are known which are called type I, type IIA, type IIB, heterotic SO(32) and heterotic $E_8 \times E_8$ string theory (see [13–16] for classic introductions to string theory). Consistency of these theories requires spacetime to have 10 dimensions. Furthermore the five known string theories are thought to emerge from one single fundamental theory, called M-theory.

### 1.1.1 The String Landscape

The fact that string theory lives in nine spatial dimensions obviously contradicts the fact that we see only three of them. One way to resolve this contradiction is to assume that the additional six dimensions are compact and so small that we are not able to observe them with current experiments. The full 10-dimensional spacetime is then taken to be a direct product $\mathcal{M}_{10} = \mathcal{M}_4 \times X_6$, where $\mathcal{M}_4$ is the 4-dimensional spacetime we observe and $X_6$ is a compact 6-dimensional manifold. There is an extremely large number of possible topologies for $X_6$ each of which gives rise to a different (metastable) string vacuum and a corresponding low energy theory in four dimensions. Initial hopes that string theory does not only provide a unifying framework for QFT and GR but also uniquely gives rise to the SM at low energies have been quickly upset by this observation.

Nevertheless, there is still the exciting possibility that at least one of the many topologies gives rise to the SM at low energies. To answer the question whether this is possible or not theoretical physicists have extensively studied string compactifications on a special type of manifold, so-called Calabi-Yau manifolds, which have a vanishing Ricci-tensor $R_{MN} = 0$. Although Calabi-Yau manifolds lead to phenomenologically attractive 4d EFTs they also have serious problems which need to be addressed in order to obtain realistic 4d physics. One of these problems is the presence of many massless or light scalar fields, so-called moduli, which is in contradiction to fifth force experiments [17] and leads to the cosmological moduli problem [18–21].

A possible solution to this problem is provided by so-called flux compactifications. The 10-dimensional EFT of string theory, supergravity, contains generalized versions of the electromagnetic field strength $F_{\mu\nu}$. These generalized field strengths are $p$-form field strengths or, equivalently, totally antisymmetric tensor fields $F_{\mu_1 \ldots \mu_p}$ with $p$ indices and are naturally integrated over $p$-dimensional (sub)manifolds. If these (sub)manifolds are homologically non-trivial, so-called $p$-cycles, the integral of a $p$-form field over this $p$-cycle takes on only quantized values. This is simply a higher-dimensional version of the Dirac quantization condition for magnetic monopoles. In the case of non-zero integral one says that there is non-trivial $p$-form flux on the cycle.

Since the compact manifold $X_6$ in general contains many different cycles each of them can be equipped with different amounts of flux. These fluxes cannot be chosen completely arbitrary but are subject to the tadpole cancellation condition which in particular forbids arbitrarily large fluxes. These fluxes introduce a contribution to the 4d energy density which depends on the moduli fields and in this way generate a potential for them. This procedure is called moduli stabilization and can solve the phenomenological problems of moduli but also increases the number of string vacua because each choice of fluxes can give rise to a different potential for the scalars. For a detailed discussion of flux compactifications see for example [22, 23].

The freedom in the choice of topology and fluxes eventually leads to a vast number of 4d low energy theories in string theory [23–25]. Although enormous, this number is thought to be finite and has been estimated to be at least of the order of $10^{272000}$ [26]. The set of all these 4d EFTs is referred to as the string landscape. Even though the number of relevant stringy 4d EFTs seems to be finite this does not really improve our situation from a practical point of view. We know nothing specific about the vast majority of those models and even if we would, it is obviously completely unfeasible to analyze them all in order to determine which of them may fit best to experimental observations.

### 1.1.2 Moduli in String Compactifications

When supergravity is compactified to four dimensions each 10-dimensional physical degree of freedom gives rise to an infinite tower of Kaluza-Klein (KK) states and the corresponding fields. Of particular interest are the massless KK zero modes as they necessarily have to be included in the 4d EFT. Among these are the moduli fields we have mentioned in the last subsection already and their vacuum expectation values parametrize the so-called moduli space. Physically, the moduli space parametrizes all possible vacuum field configurations of the 10-dimensional fields for a given compact manifold. This is similar to the vacuum expectation value of Nambu-Goldstone bosons. However, Nambu-Goldstone bosons are protected by an exact symmetry from quantum effects while the stringy moduli in general are not. Hence they can receive a non-vanishing potential via perturbative and non-perturbative quantum corrections. SUSY can protect some or all moduli against such corrections. In the following we discuss two important types of stringy moduli, axions and geometric moduli, in some more detail. More exhaustive discussions can be found in [23, 27, 28].

Let us start with axions which are particularly well-known from the Peccei-Quinn solution to the strong CP problem. Axions are pseudoscalars with a perturbative shift symmetry and therefore massless at the perturbative level. Often this continuous shift symmetry is broken by non-perturbative effects to a discrete shift symmetry. In this case the axion is a periodic field with a purely non-perturbative potential. In string compactifications axions typically arise when $p$-form gauge fields $A_p$ are integrated over $p$-cycles in order to obtain the corresponding 4d scalar field. Such gauge fields are naturally present in the 10-dimensional EFT of string theory and transform under a gauge symmetry according to $A_p \to A_p + \mathrm{d}C_{p-1}$, where $C_{p-1}$ is some arbitrary $(p-1)$-form field. This higher-dimensional gauge symmetry eventually gives rise to the shift symmetry of the corresponding axion in four dimensions.

Geometric moduli correspond to certain smooth deformations of the compact manifold's metric. More specifically, consider a string compactification on a Calabi-Yau manifold $X_6$ with metric $g_{MN}$ and Ricci-tensor $R_{MN}[g]$. By definition we have $R_{MN} = 0$. Now consider a smooth deformation $\delta g$ of the original metric $g$ such that the resulting metric can be written as $g' = g + \delta g$. Then the geometric moduli correspond to those deformations $\delta g$ which satisfy $R_{MN}[g + \delta g] = 0$. That means in particular that the vacuum expectation value of these moduli parametrize the vacuum solutions of the internal components of the 10d Einstein equation.

In order to gain some intuition for axions and geometric moduli let us illustrate them in the example of Einstein-Maxwell theory in five dimensions compactified on a circle $S^1$. The field content is given by the metric tensor $g_{MN}$ and the electromagnetic 1-form potential $A = A_M dx^M$ with $M, N \in \{0, 1, 2, 3, 4\}$. In the 4d EFT each of these fields gives rise to a massless KK state. Among these massless fields is the Wilson line $\phi = \int_{S^1} A$ of the 5d 1-form potential $A$. As mentioned above this 4d scalar inherits a perturbative continuous shift symmetry $\phi \to \phi + \text{const.}$ from the 5d gauge symmetry which reveals its axionic nature.

Next consider the component $g_{44}$ of the metric tensor. It determines the volume, i.e. circumference, of the $S^1$ to be $2\pi R = \int_{S^1} dx^4 \sqrt{g_{44}}$ where $R$ corresponds to the radius of the $S^1$ which, in general, depends on the 4d coordinates. Thereby this defines a scalar field $R$ in the 4d EFT which turns out to be massless at tree level. $R$ is an example for a geometric modulus the vacuum expectation value of which parametrizes different solutions of the vacuum Einstein equations, in this case circles of different radii. Since our toy model is not supersymmetric we expect quantum corrections to provide a potential for $R$. Indeed the Casimir energy associated with the circle generates a potential $\sim 1/R^4$ for $R$ which has a minimum in the decompactification limit $R \to \infty$ [29]. Such unstable behavior can be avoided by introducing suitable fluxes on the $S^1$.

## 1.2 The Swampland Paradigm

We have seen that the string landscape contains extremely many 4d EFTs. This raises the impression that essentially anything goes and that there is no hope to definitely answer the question whether the SM is contained in the landscape or not. From a conceptual point of view, this situation may be improved tremendously by the swampland paradigm introduced by Vafa in [30]. He pointed out that not all EFTs which are consistent from a field theoretical point of view are also necessarily consistent with QG. Such theories are hence not part of the string landscape but instead are said to be in the swampland. The motivation for this picture comes from the observation that all 4d EFTs obtained in string theory seem to have certain properties in common that can never be violated. Those properties are frequently called swampland constraints. The intriguing power of the swampland paradigm is that the corresponding constraints can be formulated quite generally so that they can be easily checked for any given EFT. In this way they may give novel guidelines for model building and we will discuss an example for this in the following paragraphs.

Before we continue to review two important swampland constraints let us add a word of caution: All of these constraints are speculative and merely conjectures. Most of the evidence for them is based on examples in string theory. Nonetheless, if true, they provide new insights into the nature of QG and help distinguishing promising paths from dead ends in model building.

### 1.2.1 The Swampland Distance Conjecture

Now we are ready to state the first swampland constraint, the so-called Swampland Distance Conjecture (SDC), as first formulated in [30] and made more precise in [31]. Consider a given string compactification and the corresponding moduli space $\mathcal{M}$. Recall that different points in $\mathcal{M}$ correspond to different vacuum expectation values of the moduli fields. The kinetic terms of all moduli can be compactly written as $g_{ij}(\Phi_k)\partial_\mu\Phi^i\partial^\mu\Phi^j$ where $i, j, k$ run over all moduli. If we interpret the $\Phi^i$ as coordinates on the moduli space, it is given the structure of a manifold with metric $g_{ij}(\Phi_k)$. Consequently, one can calculate the distance $d(p_1, p_2)$ between two points $p_1, p_2 \in \mathcal{M}$ in the moduli space. In the following we assume $d(p_1, p_2)$ to be measured in terms of the Planck mass $M_4$ of the 4-dimensional EFT. Now, let us consider the EFT at the point $p_1 \in \mathcal{M}$ and denote its cutoff by $\Lambda$. In particular all states within this EFT have masses below $\Lambda$. Then, given another point $p_2 \in \mathcal{M}$, the SDC states that there is an infinite tower of massive states appearing in the EFT at $p_2$ with lowest mass of the order of $M_4 \mathrm{e}^{-\alpha T}$ for some $\alpha > 0$ and $T > d(p_1, p_2)$. Obviously, these states can become arbitrarily light if $p_2$ is arbitrarily far away from $p_1$. That means, once the lightest state of the tower becomes lighter then the cutoff $\Lambda$, our original EFT we started with at $p_1$ breaks down because it originally did not include these light states. In other words, we cannot expect an EFT which is valid at one point in moduli space to be also valid at all other points in moduli space.

In order to illustrate this conjecture we can go back to the Einstein-Maxwell theory in five dimensions compactified on a circle. This will give rise to KK states with masses $m_{\mathrm{KK}} = n/\langle R\rangle$ with $n \in \mathbb{Z}$ and $\langle R\rangle$ the vacuum expectation value of the modulus $R$ which corresponds to the radius of the circle. Let us consider the corresponding 4-dimensional EFT with cutoff $\Lambda \ll 1/\langle R\rangle$ such that the massive KK states are not included in the EFT. The 4-dimensional kinetic term of the modulus $R$ reads $(3/2)M_4^2(\partial R)^2/R^2$ where the 4-dimensional Planck mass $M_4$ is related to the 5-dimensional one according to $M_4^2 = 2\pi\langle R\rangle M_5^3$. From this we can easily read off the metric on moduli space to be $ds^2 = (3/2)M_4^2(dR/R)^2$.[1] Now the distance between two points in moduli space specified by $R_1$ and $R_2 > R_1$ is given by $d(R_1, R_2) = \int ds/M_4 = (3/2)\log(R_2/R_1)$. Comparing the lowest KK mass at $R_1$ and $R_2$ we find $m_{\text{KK}}(R_2)/m_{\text{KK}}(R_1) = \exp[-(2/3)d(R_1, R_2)]$. This shows that the KK masses become exponentially light compared to our starting point $R_1$ when we move through the moduli space and our original EFT will eventually become invalid at points far away from $R_1$ in accordance with the SDC. In particular the KK tower of states is infinite.

The essential insight that we can draw from this conjecture is that super-Planckian field ranges, corresponding to large distances in moduli space, are difficult if not impossible to obtain consistently within EFTs. While the original conjecture was formulated mainly with strictly massless moduli in mind it seems to hold also for massive ones [32] or even for the variation of any field [33] (see also [34]). Many explicit examples in string theory [35–40] but also more general studies [41–45] support this point of view. However, there are also examples which are in tension with the SDC [46–49] and may require an improvement of the conjecture.

### 1.2.2 The Weak Gravity Conjecture

Next let us consider a second important swampland constraint: the Weak Gravity Conjecture (WGC) [50]. The simplest version of the WGC can be stated as follows. Consider a U(1) gauge theory that is coupled to gravity. Then there must exist a charged particle with mass $m$ such that $m \lesssim gM_{\text{P}}$ where $g$ is the gauge coupling. Here we have normalized $g$ such that the charge of this particle is one. This condition ensures that the electric force between two such particles is larger than the gravitational attraction, hence the name WGC.

Conceptually, the WGC is closely related to the folk theorem that gravity is incompatible with global continuous symmetries [51–54]. Now, given a local symmetry such as in a U(1) gauge theory, naively nothing prevents us from taking the limit $g \to 0$ and ending up with a global symmetry. Since such a global symmetry seems to be inconsistent with gravity we expect a general constraint on how small $g$ can be chosen. Exactly such a constraint is provided by the WGC which bounds $g$ from below by the mass of charged particles.

Besides the so-called electric form of the WGC, $m \lesssim gM_{\text{P}}$, there is a magnetic counterpart which requires the existence of an unexpected low cutoff $\Lambda$ which obeys $\Lambda \lesssim gM_{\text{P}}$. That means, in an EFT with gauged U(1) symmetry we can come arbitrarily close to a global U(1), $g \to 0$, but only at the expense of a very low cutoff. One can derive this magnetic WGC from the original electric

[1] Note that in this formula we no longer view $R$ as a 4d field but instead as a coordinate of the moduli space. Strictly speaking we hence should write $\langle R\rangle$ but will stick to $R$ for the sake of simplicity.

form by applying it to the monopoles of the U(1) theory. If the mass of the magnetic monopole is $m_{\mathrm{mag}}$, we can apply the electric WGC to it and find $m_{\mathrm{mag}} \lesssim g_{\mathrm{mag}} M_{\mathrm{P}} \sim g_{\mathrm{el}}^{-1} M_{\mathrm{P}}$, where we used that the magnetic coupling constant $g_{\mathrm{mag}}$ and electric coupling constant $g_{\mathrm{el}}$ are related by $g_{\mathrm{mag}} \sim g_{\mathrm{el}}^{-1}$. On the other hand the mass of the magnetic monopole is at least as large as the energy stored in its magnetic field. Formally, this energy is infinite but if we employ a cutoff $\Lambda$ it reads $m_{\mathrm{mag}} \sim g_{\mathrm{el}}^{-2}\Lambda$ which combined with $m_{\mathrm{mag}} \lesssim g_{\mathrm{el}}^{-1} M_{\mathrm{P}}$ gives the magnetic WGC $\Lambda \lesssim g_{\mathrm{el}} M_{\mathrm{P}}$. Alternatively, one can derive this statement by demanding that the minimally charged magnetic monopole must not be a black hole.

Very importantly, the WGC can be generalized to theories containing $p$-dimensional extended objects, i.e. $p$-branes, instead of charged particles. While charged particles couple to the electromagnetic 1-form potential $A_1$ according to $\int_{\mathrm{WL}} A_1$, where WL denotes the particle's worldline, $p$-branes naturally couple to a $(p+1)$-form potential $A_{p+1}$ via $\int_{\mathrm{WS}} A_{p+1}$. WS denotes the worldsheet of the brane which is the $(p+1)$-dimensional surface in spacetime that is swept out by the brane, completely analogous to the worldline of particles. The full action of a theory containing a $p$-brane coupled with unit charge to a $(p+1)$-form potential $A_{p+1}$ reads

$$S_{p\text{-brane}} = -\frac{1}{2g_p^2}\int F_{p+2}\wedge *F_{p+2} - T_p\int_{\mathrm{WS}} *1 - \int_{\mathrm{WS}} A_{p+1}\,, \tag{1.1}$$

where $F_{p+2} = \mathrm{d}A_{p+1}$ is the field strength and $g_p$ is the gauge coupling. Note that the tension $T_p$ has dimension $p+1$ and the gauge coupling has dimension $p+2-d/2$ if the spacetime dimension is $d$. Then the generalized WGC reads

$$T_p \lesssim g_p M_{\mathrm{P}}^{d/2-1}\,, \tag{1.2}$$

where $M_{\mathrm{P}}$ now denotes the $d$-dimensional Planck mass. Similarly to the original WGC for U(1) gauge theories this generalized WGC has a magnetic counterpart which involves the cutoff $\Lambda$ of the corresponding EFT. It explicitly reads

$$\Lambda \lesssim (g_p M_{\mathrm{P}})^{\frac{1}{p+1}}\,. \tag{1.3}$$

In the last few years there has been strong activity in the research on the WGC. It is still missing a rigorous proof, although there have been some attempts towards such a proof [55–61]. Similarly, the precise formulation of the original conjecture for U(1)-theories and its generalizations is not clear. Several versions of the WGC have been discussed and explored in the literature [34, 62–70]. For example, one uncertain point is whether the condition $m \lesssim gM_{\mathrm{P}}$ should hold for one particle, the lightest particle or even for an infinite number of particles. In the latter case we have to explicitly include the charge $Q$ of each particle since we cannot simultaneously normalize all charges to one. Then the WGC requires $m \lesssim gQM_{\mathrm{P}}$, where $m$ is the mass of the particle with charge $Q$. A lot of this recent activity has been triggered by the application of the WGC to natural inflation [71–83] which we will discuss in more detail in the next section.

## 1.3 Natural Inflation and Swampland Constraints on Axions

Early on in the beginning of inflationary model building it has been realized that generic potentials such as polynomials give rise to a phase of inflation if the corresponding scalar inflaton field takes on super-Planckian field values [84]. This has the severe disadvantage that higher-dimensional operators in the EFT become important at such high field values and hence spoil the validity of the theory [28, 85]. In order to overcome this problem symmetries were used to suppress or even completely exclude those higher-dimensional operators. A very successful model of this type is natural inflation [71]. It employs an axion $\phi$ whose perturbative continuous shift symmetry forbids any potential terms and in particular the dangerous higher-dimensional operators. The potential that is necessary for inflation to occur is generated non-perturbatively by instantons of a Yang-Mills sector to which the axion is coupled. It has the well-known form $V(\phi) = V_0 \mathrm{e}^{-S}(1 - \cos(\phi/f)) + \mathcal{O}(\mathrm{e}^{-2S})$. Here $S$ is the instanton action, $f$ is the axion decay constant and inflation requires $f \gtrsim M_{\mathrm{P}}$, i.e. super-Planckian axion decay constant. In addition we need $S \gg 1$ to neglect multi-instanton corrections to the potential at order $\mathcal{O}(\mathrm{e}^{-2S})$.

An interesting experimental prediction of natural inflation is a tensor-to-scalar ratio $r \sim 0.1$ for a scalar spectral index $n_{\mathrm{s}} = 0.9649 \pm 0.0042$ as determined by the PLANCK 2018 data [86]. Primordial gravitational waves associated with such a large value of $r$ could be observed in the BICEP2 experiment [87]. After some excitement about a possible detection of these waves corresponding to $r \sim 0.2$ in 2014 [88], which subsequently has been attributed to dust foreground, no experimental evidence for $r \sim 0.1$ has been found so far. By now the upper bounds on $r$ provided by the PLANCK and BICEP2 collaborations even have become so strong that natural inflation is in substantial tension with experimental data [86, 89].

Even though this is discouraging from a phenomenological point of view a lot of research that has been done on models of natural inflation is valuable beyond its phenomenology. This is especially true for the application of swampland constraints to those models and the associated research on the SDC and WGC in general. In the following we will explain how natural inflation is related to the swampland paradigm and how the corresponding constraints can ultimately be understood as general QG constraints on axions independently of inflation.

When trying to obtain models of natural inflation from string theory it was realized that the necessary super-Planckian axion decay constants are very difficult, if not impossible, to obtain in a controlled setting [90]. This conclusion is supported by the naive application of the WGC to axions. To see this consider the generalized WGC in (1.2). For $d = 4$ and $p = -1$ it applies to a theory of a real scalar $\phi$ with field strength $\mathrm{d}\phi$ and a perturbative shift symmetry: it is nothing else than an axion. $\phi$ naturally couples to $(-1)$-branes which have a point-like worldsheet. Hence $(-1)$-branes are simply events in spacetime and can be viewed as idealized, point-like instantons.

Recall that these two ingredients, an axion coupled to instantons, are exactly what the natural inflation model consists of. The action of natural inflation is given by

$$S_{\mathrm{NI}} = -\frac{f^2}{2}\int \mathrm{d}\phi \wedge *\mathrm{d}\phi - \frac{1}{2g_{\mathrm{YM}}^2}\int \mathrm{tr}(F_2 \wedge *F_2) + \frac{1}{8\pi^2}\int \phi\, \mathrm{tr}(F_2 \wedge F_2)\,, \quad (1.4)$$

where $f$ is the axion decay constant as before, $F_2$ is the field strength of a non-abelian gauge sector, and $g_{\mathrm{YM}}$ its coupling constant. Now we would like to compare (1.4) with (1.1) in order to determine $g_{-1}$ and $T_{-1}$ in terms of $g_{\mathrm{YM}}$ and $f$. This is easily done for the first term by identifying $\phi = A_0$ which implies $g_{-1} = 1/f$.

Comparing the second term is more subtle. The problem is that instantons in (1.4) are localized field configurations of finite size while they are treated idealized as point-like objects in (1.1). However, in the absence of arbitrarily large instantons we can treat the remaining small ones as point-like at sufficiently low energies. This scenario could be realized, for example, by Higgsing the gauge theory at a high scale such that large instantons are suppressed. In this sense (1.1) can be understood as the low energy effective action of (1.4). Now let us compare the second terms in (1.1) and (1.4). For a 1-instanton configuration we denote the latter by $S$. The corresponding term in (1.1) reads $T_{-1}\int_{\mathrm{WS}} *1$. Because the worldsheet of an (idealized) instanton is a point in spacetime this integral simply counts the number of instantons and consequently equals one for a 1-instanton configuration. Thus we conclude $T_{-1} = S$.

For completeness let us also briefly discuss the third terms in (1.1) and (1.4). The integral in (1.1) simply evaluates the axion $A_0 = \phi$ at points in spacetime where an instanton is located. To coincide with the corresponding term in (1.4), the expression $\mathrm{tr}(F_2 \wedge F_2)$ must be proportional to a sum of $\delta$-functions which are centered at the position of the instantons. This is indeed the case in the limit of vanishing instanton sizes.

Using $T_{-1} = S$ and $g_{-1} = 1/f$ in (1.2) yields the constraint $fS \lesssim M_{\mathrm{P}}$. We have seen that $S \gg 1$ is necessary to maintain control over the non-perturbative axion potential. But then $fS \lesssim M_{\mathrm{P}}$ implies $f \ll M_{\mathrm{P}}$, i.e. super-Planckian axion decay constants are forbidden by the WGC in the controlled regime. Thus natural inflation is naively censored by the WGC.

Besides the WGC it is also possible to apply the SDC to natural inflation. To do so we make use of the Lyth-bound [91] which relates the tensor-to-scalar ratio $r$ to the field range $\Delta\phi$ which the inflaton traverses during inflation. It reads

$$\frac{\Delta\phi}{M_{\mathrm{P}}} = \mathcal{O}(1)\sqrt{\frac{r}{0.01}}\,. \quad (1.5)$$

Since data favors $r \sim 0.1$ [86] for natural inflation we conclude that $\Delta\phi \sim M_{\mathrm{P}}$ in this case. But the SDC tells us that Planckian field excursion cannot be consistently described within a single EFT. Hence natural inflation seems to be incompatible with the SDC as well.

All of these findings support the idea that natural inflation is in tension with QG constraints as are formulated within the swampland program. Even though there are proposals to evade these constraints [92–94], none of them is completely free of caveats [36, 39, 66, 73, 74, 76, 79–81, 83]. It is hence an important task to improve the understanding of the SDC and WGC when applied to axions. This is not only motivated by the relevance of axions for natural inflation but also more generally by their importance for different proposals for BSM physics such as the axion solution to the strong CP problem [95–101], dark matter [102–106] (for a review on axion cosmology see [107]) and string theory [28].

Besides finding a rigorous proof there are two main directions along which one can attempt to make progress towards a better understanding of the SDC and WGC for axions. First, there is the example-based approach. Indeed, most of the evidence in favor of the SDC and WGC is drawn from explicit constructions in string theory which all seem to obey the two swampland constraints. Alternatively one can try to find explicit counter examples within string theory. Since string theory is the main motivation for the conjectures this would cast serious doubt on the validity of the SDC and WGC. One then would have to analyze whether an improved version of the conjectures holds or one has to drop them altogether. Second, one can find general arguments for or against the conjectures independently of string theory. This could be done, for example, by showing that EFTs which violate the SDC or WGC suffer from some fundamental pathologies. Evidence found in this way would be complementary to that found in string theory and possibly gives more insight into the underlying principles of swampland constraints. Furthermore the application of the conjectures to different EFTs can serve as an important consistency check and may give hints towards the correct formulation of them.

## 1.4 Contribution of this Thesis

Motivated by the relevance of axions for natural inflation and their frequent appearance in many proposals for BSM physics as well as in string theory this thesis is devoted to the study of the SDC and the WGC applied to axions. In particular we investigate the question whether axions with significantly super-Planckian field range are censored by swampland constraints and whether they exist in string theory. To do so we employ essentially the two approaches described above, i.e. on the one hand we analyze an explicit example in string theory that contains axions and on the other hand we study the magnetic WGC for axions from a low energy perspective. Furthermore we consider axions coupled to Yang-Mills theory which leads us to speculate about constraints on axion potentials.

Our goal in Chapter 2 is to find an axion in the moduli space of a concrete string compactification with as large a field range as possible. Since we are not interested in phenomenological applications we can ignore many complications, such as SUSY-breaking and moduli-stabilization, which one otherwise would have to address to construct realistic models of, for example, natural inflation. We consider a very simple compactification of type IIB string theory on a factorizable 6-torus $(\mathrm{T}^2 \times \mathrm{T}^2 \times \mathrm{T}^2)/\mathrm{Z}_2$ subject to a $\mathrm{Z}_2$-identification. For this compactification all moduli are massless. In particular, several of the moduli are axions which, however, have only exactly Planckian field range.

This situation can be improved by turning on fluxes on the tori which has two main effects. First, a potential for the moduli is generated such that only a complex 2-dimensional subspace of the original moduli space remains flat. In particular this subspace contains two flat axionic direction. Second, the flux on the tori reduces the original $\mathrm{SL}(2,\mathbb{Z})$ symmetry of these and leave us with so-called congruence subgroups of $\mathrm{SL}(2,\mathbb{Z})$. Surprisingly, this reduced symmetry leads to an increase in the size of the moduli space. In fact the field range of one of the axions is enhanced by the a flux number $N$ which is bounded from above by 16 according to a tadpole cancellation condition. Consequently, we successfully constructed an axion with mildly super-Planckian field range. Given this result one would naively also expect the whole moduli space to have increased in size by a factor $N$. This however is not true because the topology and geometry of the moduli space is very non-trivial and, in particular, the large axionic direction is not a geodesic. Instead we find that the size grows only logarithmically with the flux number.

In the second part of our analysis we examine the moduli space from a 4d low energy perspective. To do so we introduce the concept of a restricted moduli space $\mathcal{M}(\Lambda)$ which depends on an energy scale $\Lambda$. $\mathcal{M}(\Lambda)$ is defined as the set of all points in moduli space at which the masses of KK and winding modes, that occur upon dimensional reduction to four dimensions, are larger than $\Lambda$. This means that a 4d EFT with cutoff $\Lambda$ is valid only in the restricted moduli space $\mathcal{M}(\Lambda)$ and breaks down outside due to the appearance of light KK or winding modes. We can estimate the size of the restricted moduli space to be of the order of $\ln(M_\mathrm{P}/\Lambda)$ where $M_\mathrm{P}$ is the 4-dimensional Planck mass. In contrast to the full moduli space the size of the restricted moduli space is not affected by the fluxes.

Motivated by our simple example we formulate the following conjecture about moduli spaces of a general string compactification to four dimensions: Consider two points in the moduli space which have a distance $L$ as determined by an appropriate geodesic connecting these two points. Then there exist points on this geodesic at which the lightest KK or winding mode mass is below or of the order of $M_\mathrm{P}\exp(-\alpha L)$, with $\alpha \sim \mathcal{O}(1)$. Note that this is very similar to the original SDC but differs from it by the fact that our conjecture allows for two widely separated points at which no light KK or winding modes are present but which instead occur somewhere in between on a geodesic connecting the two points.

In Chapter 3 we analyze the magnetic WGC for axions from a low energy perspective. Naively extrapolating (1.3) in four dimensions to axions, i.e. $p = -1$, yields $\Lambda = 0$ which suggests that EFTs with super-Planckian axion decay constants, $f > M_{\mathrm{P}}$, and hence super-Planckian axionic field ranges do not exist. Since this is a strong claim and it is not clear whether (1.3) can be trusted for $p = -1$ we study low energy theories with $f > M_{\mathrm{P}}$ to see whether something is fundamentally wrong in such theories. More specifically, recall that the alternative formulation of the magnetic WGC for U(1) theories was that the magnetic monopoles must not be black holes. The analogues of magnetic monopoles in a theory of axions are strings which are magnetically coupled to the axion. Consequently we focus on such strings and the question whether these develop an event horizon similar to that of a black hole for super-Planckian axion decay constant $f > M_{\mathrm{P}}$.

It turns out that the spacetime of axionic strings has no event horizon but a physical singularity at a finite distance from the string axis. For $f \ll M_{\mathrm{P}}$ this singularity is exponentially far away from the string and may be irrelevant in a cosmological setting but it definitely does not provide a physical solution of an isolated string. One can get rid of this singularity at the expense of having a non-static spacetime as can be shown rigorously in an explicit model that provides an ultraviolet completion of the axionic string [108, 109]. For $f \lesssim M_{\mathrm{P}}$ the spacetime inflates along the string axis only while for $f \gg M_{\mathrm{P}}$ inflation occurs in all directions. This latter situation is known under the name topological inflation. Unfortunately it is unclear whether such highly non-static strings are allowed by the magnetic WGC. We are therefore left with a rather inconclusive situation although it seems unlikely that topological inflation should be allowed by the WGC. This tendency is based on the observation that allowing topological inflation also in U(1) theories would invalidate certain versions of the corresponding magnetic WGC which we trust more than the naive axionic counterpart.

In an attempt to circumvent the problems with topological inflation we next consider a theory that contains two axions with sub-Planckian decay constant and corresponding strings which are at least minimally well-behaved in the sense that they only inflate along the string axis. By appropriately gauging the shift symmetry of these two axions by one and the same 3-form gauge field we end up with a theory that contains a single effective axion with super-Planckian decay constant. The corresponding effective string is built out of the two strings of the original axions which are bound together by domain walls. An estimate of the tension of this composite string, however, suggests that also in this case topological inflation will take place or singularities will be present. Thus our gauging approach failed to construct a well-behaved string with super-Planckian axion decay constant and our original conclusion that the status of the magnetic WGC for axions is rather unclear remains.

Finally, we discuss axion potentials and possible QG constraints on it in Chapter 4. The central idea here is that we expect the perturbative axionic shift symmetry to be broken non-perturbatively in order to satisfy the expected censorship of global symmetries by QG. As is well known, this is indeed the case for an axion that is coupled to a Higgsed Yang-Mills theory. Recall that the corresponding potential has the form $V(\phi) = V_0 \mathrm{e}^{-S}(1 - \cos(\phi/f)) + \mathcal{O}(\mathrm{e}^{-2S})$. The WGC for axions requires $fS \lesssim M_{\mathrm{P}}$ so that the suppression of the potential due to the instanton action cannot be arbitrarily strong. Nevertheless, this does not strictly exclude a non-vanishing axion potential since the prefactor $V_0$ is not constrained at all.

Indeed, if we couple massless fermions to the Yang-Mills theory or let massive ones become massless, $V_0$ vanishes and hence also the axion potential. This can be understood as a consequence of a global symmetry which consists of a shift in the axion and an anomalous U(1) rotation of the massless fermions. A similar symmetry occurs for fermions which have no hard mass term but become massive via Yukawa couplings. In such a theory the axion would be massless as well. We point out that any fermion operator which explicitly breaks anomalous chiral rotations also breaks the shift symmetry that ensures a flat axion potential and hence provides a mechanism by which QG could potentially censor axionic shift symmetries.

So far we have discussed the breaking of axionic shift symmetries by QG on a qualitative level. It would, however, be interesting to know more precisely how strong this breaking of shift symmetries should be. A natural measure for this strength seems to be the amplitude $V_0 \mathrm{e}^{-S}$ of the axion potential which must vanish in the presence of a shift symmetry. Our first, simplest guess, partially based on the WGC for axions, is $V_0 \mathrm{e}^{-S} \gtrsim M_{\mathrm{P}}^4 \exp(-M_{\mathrm{P}}/f)$. This, however, can not be true since $\mathcal{N} = 2$ supergravity has an exactly vanishing potential for axions and, as the low energy EFT of string theory, is certainly compatible with QG. Consequently, we should not require a non-vanishing axion potential under all circumstances.

Our second approach towards a bound on axion potentials makes use of the effective 3-form description of Higgsed Yang-Mills theory which we discuss in some length in the first part of Chapter 4. This effective description is also able to capture the effects of fermions and axions which are coupled to the Yang-Mills theory. Our idea is then that the application of the WGC for 3-forms to the effective description may provide bounds on the axion potential. Unfortunately, it turns out that this method does not lead to any sensible results.

Finally we propose a bound on axion potentials which seems to evade the problems of the two approaches discussed above. Let $\mu$ be the cutoff of the effective low energy theory that exclusively contains the axion, i.e. in particular $\mu = 0$ if there are other massless degrees of freedom. Then the axion potential is bounded from below according to $V_0 \mathrm{e}^{-S} \gtrsim \mu^4 \exp(-M_{\mathrm{P}}/f)$. In theories containing fermion operators which generate an axion potential this translates into a lower bound on the strength of these operators. For example, in the case of a Higgsed Yang-Mills theory with an axion and fermions of mass $m$ we find parametrically $m \gtrsim \exp(-M_{\mathrm{P}}/f)v$ where $v$ is the Higgs scale.

In the last section of Chapter 4 we study the interesting possibility that gravitational instantons generate fermion operators. The underlying mechanism is completely analogous to how gauge instantons generate so-called 't Hooft vertices which are higher-dimensional fermion operators breaking chiral U(1) rotations.

However, the difficulty is to find an appropriate gravitational instanton that can be embedded in an asymptotically flat spacetime and has the right topological properties. We argue that a K3 manifold is a viable candidate for such an instanton. A parametric estimate of the effect of such gravitational instantons reveals that they are not relevant for phenomenology.

# 2 Super-Planckian Axions in String Theory and a Moduli Space Size Conjecture

## 2.1 Introduction

Evidence from string theory (see for example [90]) as well as the SDC and WGC (cf. Sections 1.2, 1.3 and Chapter 3) suggest that axions with super-Planckian field ranges are incompatible with QG and reside in the swampland. In this chapter we challenge this negative conclusion and attempt to construct super-Planckian axion field spaces in a very simple stringy setting which allows for explicit calculations. We are not interested in inflation or any other phenomenological application, which allows us to avoid the problems of realistic string constructions. In addition, we are not prone to the (possibly model-dependent) backreaction effects which underlie the bounds obtained in [32, 36]. This increases the chance that any bounds we find have a generic QG origin.

Thus, our focus are supersymmetric, *flat* axionic directions such that backreaction plays no role. This is close in spirit to the approach taken in [110]. Here we choose to work with type IIB string theory compactified on a toroidal orientifold with supersymmetric 3-form flux. Such a flux generically reduces the dimension of moduli space. It can also introduce a monodromy (with finite but possibly large monodromy group) in the remaining flat directions.[1] To keep the discussion focused on the question at hand, we do not address the problem of stabilizing the remaining moduli. Working out the consequences of our flux choice we find that a certain 2-dimensional subspace of the full moduli space is enlarged by a factor $N$, where $N$ is a flux number. In this way, to the best of our present understanding, a super-Planckian flat axionic direction emerges.

However, one should be careful about an interpretation of this in the sense of a large field space. The key is the geometry of this space. Indeed, the reason for the extended moduli space is the reduced modular invariance of tori with

[1]Monodromies also arise in flux compactifications on Calabi-Yau manifolds and have been discussed in the context of moduli dynamics and tunneling in the string landscape, see e.g. [111–114]. In these works monodromy transformations connect points with different values of the scalar potential or isolated vacua. By contrast, we study monodromy transformations between points on a periodic flat direction, enlarging the periodicity of the latter. Note also that, following the recent literature on inflation, we use the term monodromy for the breaking of a periodicity by flux, not for the large diffeomorphism required to make the original periodicity manifest.

fluxes as compared to tori without flux. The resulting moduli space is given by a fundamental domain of so-called congruence subgroups of $\mathrm{SL}(2,\mathbb{Z})$. Together with the proper metric, this space is a Riemann surface of a certain genus, with locally hyperbolic geometry, with a number of conical singularities and with singular cusps or throats. The natural way to measure distances between two points in this space is via geodesics. However, the long axionic trajectories advertised above are very far from being geodesics. Two points on such an axionic trajectory may have an 'axionic' distance $\sim N$, with $N$ our potentially large flux number. Yet their geodesic distance is only $\sim \ln(N)$. More generally the geodesic distance between any two points is bounded by an expression of order $\ln(M_\mathrm{P}/\Lambda)$, where $\Lambda$ is the cutoff below which the 4-dimensional effective theory is valid.

We try to formalize these findings in terms of two conjectures which are related to but also distinctly different from the well-known SDC and recent variants [30, 31, 33, 34]. Consider the moduli space of a generic 4-dimensional field theory with cutoff $\Lambda$. Then we conjecture that the absolute size of the moduli space, as measured by the appropriately defined diameter, scales as $\ln(M_\mathrm{P}/\Lambda)$. Alternatively, we may focus on the full moduli space of a certain string compactification. Pick two points in this moduli space which are connected by a geodesic with length $L$. Then we claim that there exist points on this geodesic at which the lightest KK or winding mode mass is smaller or of the order of $\exp(-\alpha L)$, with $\alpha \sim \mathcal{O}(1)$.

At first sight all of this might suggest that long and in particular long axionic trajectories are not realizable in 4-dimensional effective field theories with high cutoff. However, recall that we have found a long axionic direction. The fact that this direction was not a geodesic may be irrelevant if one is able to construct an appropriate potential that forces the field onto this long trajectory.[2] Thus, it appears that the question of large-field inflation requires knowledge beyond the WGC and SDC.

## 2.2 A Monodromic Moduli Space via Fluxes

### 2.2.1 KNP vs. Winding Trajectories from Fluxes

We want to construct a long axionic direction in the moduli space of a supersymmetric compactification of type IIB string theory as a long winding trajectory in a compact field space. This is the Kim-Nilles-Peloso (KNP) mechanism [92], but in our case the winding trajectory will arise due to 3-form fluxes rather than the instanton potential employed in [92]. The idea is as follows. Consider a theory with two axions $\varphi_1$ and $\varphi_2$ with small and, for simplicity, equal periodicity given

[2] Recently, a model of inflation has been proposed in which the hyperbolic geometry of field space is essential [115] (see also [116]). It would be interesting to see whether this can be realized in our setting. Such models have also been discussed in [117,118] (see also [119–123]). In particular it has been pointed out therein that compatibility with observation may be achieved without stabilizing all scalar fields except for the inflaton itself.

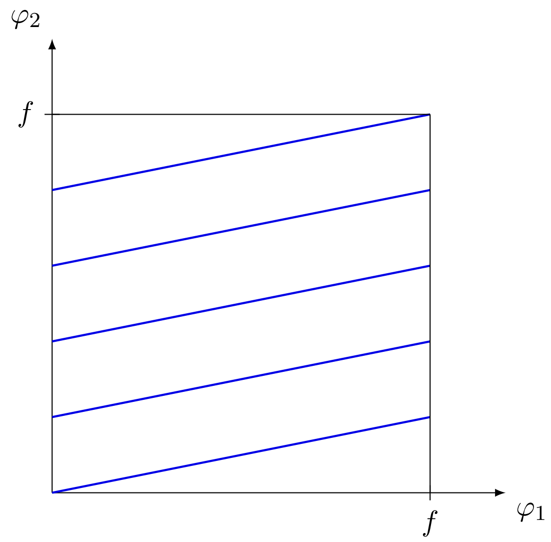


Figure 2.1: Winding flat direction of total length $\sim Nf$ (shown for $N = 5$).

by the axion decay constant $f$. Even though this field space is small one can generate a long trajectory by having a potential for the axions that has a minimum at $\varphi_1 = N\varphi_2$ for a large integer $N$. Now, the remaining flat direction has a periodicity of $\sqrt{N^2+1}f$ which is much larger than the original $f$ for large $N$ (see Fig. 2.1). This is the KNP-mechanism.

We choose to work in a simple setup of toroidal orientifolds. Thus we take as the compact space $X_6 = \mathrm{T}^6/\mathrm{Z}_2 = (\mathrm{T}_1^2 \times \mathrm{T}_2^2 \times \mathrm{T}_3^2)/\mathrm{Z}_2$, i.e. a factorisable 6-torus subject to a $\mathrm{Z}_2$ identification. By turning on 3-form fluxes on the tori we will show how one can generate a superpotential of the form [124–127]

$$W = (M\tau_1 - N\tau_2)(\tau - \tau_3)\,, \tag{2.1}$$

where $\tau = C_0 + \mathrm{i}e^{-\phi}$ is the axio-dilaton, $\tau_i$ with $i = 1, 2, 3$ are the complex structure moduli of the three 2-tori and $M, N$ are integers (flux numbers). For the following analysis it will be useful to label the real and imaginary components of $\tau$ and $\tau_i$ and we hence define

$$\tau = C_0 + \mathrm{i}e^{-\phi} = c + \mathrm{i}s\,, \tag{2.2}$$

$$\tau_i = \mathrm{Re}\,\tau_i + \mathrm{i}\,\mathrm{Im}\,\tau_i = u_i + \mathrm{i}v_i\,. \tag{2.3}$$

Throughout this work we will refer to the real parts $c = \mathrm{Re}\,\tau$ and $u_i = \mathrm{Re}\,\tau_i$ as 'axionic' directions due to their associated shift symmetries.[3]

[3]In the case of $c$ the shift symmetry arises from the $SL(2,\mathbb{Z})$ symmetry of type IIB string theory and persists beyond toroidal orientifold compactifications. The shift symmetries in $u_i$ orig-

Without loss of generality we can take $v_i > 0$. A minimum of the scalar potential is determined by the conditions $D_I W = 0$ and $W = 0$, where $I$ runs over all moduli. This corresponds to the supersymmetric vacuum with $\tau_3 = \tau$ and $M\tau_1 = N\tau_2$. Note that the minimum is not a unique point in field space, as there are several flat directions. First, let us consider only one particular flat direction in the $(u_1, u_2)$ field space, defined by

$$\psi \equiv Mu_1 - Nu_2 = 0 \tag{2.4}$$

and all other moduli fixed. Our main focus is whether this direction can be long enough such that we can traverse a trans-Planckian distance.

Naively, it may seem that there is no bound to this flat direction. If we increase $u_1$ we simply have to increase $u_2$ accordingly to keep $\psi = 0$. Of course, as suggested by Fig. 2.1, we will return to the same geometrical situation after a certain distance. But it is at first sight not obvious whether the flux configuration on the torus has changed.

To study this in detail, recall $u_1$ and $u_2$ are the real parts of $\tau_1$ and $\tau_2$, which are the complex structure moduli of two tori. Further recall that the complex structure moduli sector exhibits a modular symmetry: All tori whose complex structure moduli are related by an $\mathrm{SL}(2,\mathbb{Z})$ transformation are equivalent. Thus, if we wish to limit ourselves to physically inequivalent configurations, we have to limit the range of $\tau_1$ and $\tau_2$ to the fundamental domain of $\mathrm{SL}(2,\mathbb{Z})$. Accordingly, $u_1$ and $u_2$ are constrained to be in the corresponding fundamental domain.

However, the situation becomes more complicated in the presence of 3-form fluxes. Since these are 3-forms on the tori, a modular transformation on them will also induce a transformation of the fluxes. In the following, we show how this leads to a monodromic, i.e. enlarged, moduli space and to a long but finite axionic direction.

### 2.2.2 Brief Interlude Concerning the Action of the Modular Group

Before we explain how to arrive at a superpotential (2.1) and how the moduli space is extended we need to set up some elementary notation concerning $\mathrm{SL}(2,\mathbb{Z})$ and gauge redundancies of tori. Let a torus be defined as the complex plane modded out by some lattice,

$$\mathbb{C}/\mathrm{span}_{\mathbb{Z}}(e_y, e_x)\,. \tag{2.5}$$

Coordinates $y \in [0,1)$ and $x \in [0,1)$ are introduced by

$$z = (y, x) \cdot \begin{pmatrix} e_y \\ e_x \end{pmatrix}\,. \tag{2.6}$$

---

inate from the $SL(2,\mathbb{Z})$ modular symmetries of the compactification tori. For more general compactifications on Calabi-Yau 3-folds, shift symmetries in the complex structure moduli sector are typically broken, but this breaking becomes increasingly weak when approaching large complex structure.

For example, with $e_y = \tau$, $e_x = 1$ we have

$$z = (y, x) \cdot \begin{pmatrix} \tau \\ 1 \end{pmatrix} = x + \tau y \,. \tag{2.7}$$

More generally, the same torus is described by

$$z \;=\; (y, x)\, R^{-1} R \begin{pmatrix} \tau \\ 1 \end{pmatrix} \;=\; e'_x x' + e'_y y' \;\equiv\; e'_x\,(x' + \tau' y')\,, \tag{2.8}$$

with

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}) \;, \qquad \tau' \equiv \frac{e'_y}{e'_x} = \frac{a\tau + b}{c\tau + d} \equiv R(\tau) \tag{2.9}$$

and

$$\begin{pmatrix} y' \\ x' \end{pmatrix} = R^{-1\,\mathrm{T}} \begin{pmatrix} y \\ x \end{pmatrix} . \tag{2.10}$$

For our following analysis it will be important that, by the above logic, the components of any 1-form

$$\omega = \omega_i \mathrm{d}\xi^i \qquad \text{with} \qquad \mathrm{d}\xi^i = \begin{pmatrix} \mathrm{d}y \\ \mathrm{d}x \end{pmatrix} \tag{2.11}$$

transform according to

$$\omega'_i = R_i{}^j \omega_j \,. \tag{2.12}$$

### 2.2.3 Flux Choice

Let us briefly describe how we can arrive at a superpotential of the form (2.1) from flux compactifications in toroidal orientifolds. Here and in the following we will set $(2\pi)^2\alpha' = 1$ where $\sqrt{\alpha'} = l_\mathrm{s}$ is the string length. The superpotential is the Gukov-Vafa-Witten superpotential, which can be written as

$$W = \int_{X_6} \Omega_3 \wedge G_3 \,, \tag{2.13}$$

where

$$\begin{aligned} \Omega_3 &= \mathrm{d}z_1 \wedge \mathrm{d}z_2 \wedge \mathrm{d}z_3 \\ &= (\mathrm{d}x_1 + \tau_1 \mathrm{d}y_1) \wedge (\mathrm{d}x_2 + \tau_2 \mathrm{d}y_2) \wedge (\mathrm{d}x_3 + \tau_3 \mathrm{d}y_3)\,, \\ G_3 &= F_3 - \tau H_3\,, \end{aligned} \tag{2.14}$$

and $(y_i, x_i)$ are the coordinates on the $i$th torus. For completeness, let us also record the Kähler potential

$$\begin{aligned} K &= -\ln\left(-i(\tau - \bar\tau)\right) - 2\ln\mathcal{V} - \ln\left(-i\int_{X_6} \Omega_3 \wedge \overline{\Omega}_3\right) \\ &= -\ln\left(-i(\tau - \bar\tau)\right) - 2\ln\mathcal{V} - \ln\left(i(\tau_1 - \bar\tau_1)(\tau_2 - \bar\tau_2)(\tau_3 - \bar\tau_3)\right) . \end{aligned} \tag{2.15}$$

The superpotential in (2.1) then arises for the following choice for the 3-form fluxes:[4]

$$F_3 = (+M\,\mathrm{d}x_1 \wedge \mathrm{d}y_2 - N\,\mathrm{d}y_1 \wedge \mathrm{d}x_2) \wedge \mathrm{d}x_3\,, \tag{2.16}$$
$$H_3 = (-M\,\mathrm{d}x_1 \wedge \mathrm{d}y_2 + N\,\mathrm{d}y_1 \wedge \mathrm{d}x_2) \wedge \mathrm{d}y_3\,. \tag{2.17}$$

Note that this can also be written more compactly as $F_3 = +\mathcal{A} \wedge \mathrm{d}x_3$ and $H_3 = -\mathcal{A} \wedge \mathrm{d}y_3$, where we introduced the 2-form $\mathcal{A}$ which is only supported on the first two tori:

$$\mathcal{A} = A_{ij}\,\mathrm{d}\xi_1^i \wedge \mathrm{d}\xi_2^j \qquad \text{with} \qquad \xi_1^i = \begin{pmatrix} y_1 \\ x_1 \end{pmatrix} \qquad \text{and} \qquad \xi_2^i = \begin{pmatrix} y_2 \\ x_2 \end{pmatrix}. \tag{2.18}$$

The essential part of the explicit flux information is encoded in the matrix

$$(A_{ij}) = A = \begin{pmatrix} 0 & -N \\ M & 0 \end{pmatrix}. \tag{2.19}$$

This flux choice enforces $M\tau_1 = N\tau_2$ and $\tau_3 = \tau$. We will ignore $\tau$ and $\tau_3$ and focus on the restricted 2-dimensional moduli space resulting from $\tau_1$ and $\tau_2$. It can be parametrized, for example, by $\tau_1$ alone.

There is a constraint on the values of $N$ and $M$ coming from the D3 tadpole cancellation condition. It reads

$$N_{D3} + \frac{1}{2}\int_{X_6} H_3 \wedge F_3 = \frac{1}{4} N_{O3}\,, \tag{2.20}$$

where $N_{\mathrm{D3}}$ is the number of D3-branes and $N_{\mathrm{O3}}$ is the number of O3-planes. For the toroidal orientifold $\mathrm{T}^6/\mathrm{Z}_2$ one finds 64 fixed points corresponding to 64 O3-planes. The flux contribution for our ansatz (2.17) can be calculated as $\int_{X_6} H_3 \wedge F_3 = 2MN$. We thus arrive at the constraint:

$$MN \leq 16\,, \tag{2.21}$$

where the maximal value of 16 is attained for $N_{\mathrm{D3}} = 0$.

### 2.2.4 The Monodromic Moduli Space

Let us now return to the question of the size of moduli spaces in the presence of flux. Given our superpotential (2.1) the minimum at $W = 0$ exhibits two complex flat directions defined by $\tau - \tau_3 = 0$ and $M\tau_1 - N\tau_2 = 0$. Here we will focus on the latter.

As noted before, we can restrict attention to $\tau_1$. Naively, one expects it to take values e.g. in the canonical fundamental domain. We will immediately see that, in the presence of fluxes, this is not any more true. Consider an arbitrary $\tau_1$ and a flux configuration determined by the matrix $A$. Now, while keeping $A$ fixed,

[4] Note that odd flux numbers $M$ and $N$ imply the existence of further 'exotic' O3 planes [124].

move $\tau_1$ in the upper complex half plane to any other $\tau_1'$ that is related to $\tau_1$ by a modular transformation, i.e.

$$\tau_1 = R_1(\tau_1') = \frac{a\tau_1' + b}{c\tau_1' + d} \tag{2.22}$$

for some $R_1 \in \mathrm{SL}(2,\mathbb{Z})$. Then the $(F_3, H_3)$ fluxes also transform non-trivially due to the transformation properties of the matrix $A_{ij}$:

$$A_{ij} \to A'_{ij} = (R_1)_i{}^k A_{kj}\,. \tag{2.23}$$

Therefore, although $\tau_1$ and $\tau_1'$ are related by a modular transformation and the corresponding two tori are identical, the whole physical configuration may be different due to different values of the fluxes given by (2.23). However, it is possible that this non-trivial transformation of the fluxes can be undone by a transformation acting on the second index, associated with a modular transformation of the second torus. For this, one must require that an $\mathrm{SL}(2,\mathbb{Z})$ matrix $R_2$ exists such that

$$A'' = R_1 A R_2^T = A\,. \tag{2.24}$$

The condition for this to be possible is that the matrix $A^{-1}R_1^{-1}A$ is in $\mathrm{SL}(2,\mathbb{Z})$,

$$R_2^{\mathrm{T}} = A^{-1}R_1^{-1}A = \begin{pmatrix} a & cN/M \\ bM/N & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})\,. \tag{2.25}$$

Restricting our attention to the case where $M$ and $N$ have no common divisors, $b$ must be a multiple of $N$ and $c$ a multiple of $M$.

An important consistency check is to verify that, after performing the transformations above, we still satisfy the vacuum condition $M\tau_1' = N\tau_2'$. Indeed, one easily calculates

$$N\tau_2' = N\frac{a\tau_2 + bM/N}{cN\tau_2/M + d} = M\tau_1'\,, \tag{2.26}$$

where we used $M\tau_1 = N\tau_2$.

In the special case of $M = 1$, the only restriction on $R_1$ is that $b$ is a multiple of $N$. This means that the 'smallest' transformation of $\tau_1$, defining the periodicity of its real part, takes the form

$$R_1 = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}\,. \tag{2.27}$$

But this is exactly what we expected: The width of the fundamental domain is not unity, e.g. $\mathrm{Re}\,\tau_1 \in (-1/2, 1/2)$, but has been extended to $N$, such that we can choose e.g. $\mathrm{Re}\,\tau_1 \in (-N/2, N/2)$. Fig. 2.2 shows such an extended fundamental domain for $N = 5$, calculated with the program 'fundomain' by H. Verrill [128]. Since this is the only feature of interest for us we set $M = 1$ throughout the rest of this chapter.

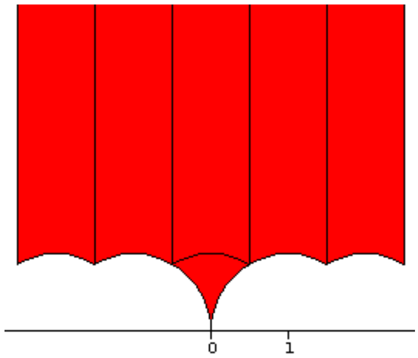


Figure 2.2: A fundamental domain of the congruence subgroup $\Gamma^0(5)$ as a subset of the upper complex half plane is shown. The central strip without the 'triangle' touching the real axis corresponds to the standard fundamental domain of the complex structure modulus of a torus.

Using the Kähler potential (2.15) one can determine the metric in moduli space restricted to $\tau_1$ and $\tau_2$:

$$\mathrm{d}s^2 = \frac{\mathrm{d}\tau_1 \mathrm{d}\overline{\tau_1}}{4(\mathrm{Im}\,\tau_1)^2} + \frac{\mathrm{d}\tau_2 \mathrm{d}\overline{\tau_2}}{4(\mathrm{Im}\,\tau_2)^2}\,. \tag{2.28}$$

Evaluating this in the vacuum $\tau_1 = N\tau_2$ parametrized by $\tau_1$ one finds

$$\mathrm{d}s^2 = \frac{\mathrm{d}\tau_1 \mathrm{d}\overline{\tau_1}}{2(\mathrm{Im}\,\tau_1)^2}\,. \tag{2.29}$$

We are now in a position to calculate the length of the flat direction defined in (2.4). Our result is

$$L = \int_{-N/2}^{N/2} \frac{du_1}{\sqrt{2}\,\mathrm{Im}\,\tau_1} = \frac{N}{\sqrt{2}\,\mathrm{Im}\,\tau_1}\,. \tag{2.30}$$

Note that the value of $N$ is bounded by a tadpole constraint such that $N = 16$ is the largest allowed value. Saturating this bound and setting $\mathrm{Im}\,\tau_1 = 1$ we find $L = 8\sqrt{2}$ for the length of the flat direction. To summarize, it appears that we have succeeded in generating a super-Planckian flat axionic direction.

## 2.3 Topology and Geometry of Fundamental Domains of Congruence Subgroups

The transformations described by Eq. (2.25) constitute so-called congruence subgroups of $\mathrm{SL}(2,\mathbb{Z})$. We have already shown the fundamental domain of such a subgroup for the case $M = 1$ and $N = 5$, denoted by $\Gamma^0(5)$ in Fig. 2.2. We can

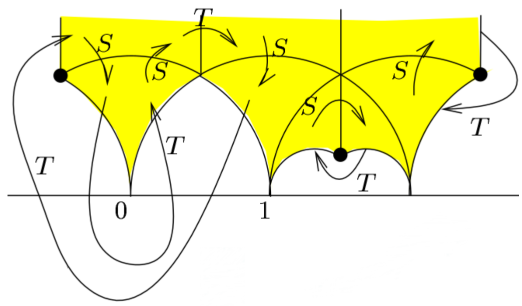


Figure 2.3: The lower part of the fundamental domain of the congruence subgroup $\Gamma^0(7)$ is shown. Appropriate identifications of the boundaries are indicated [128].

explicitly see the enlarged field space for $\mathrm{Im}\tau_1 > 1$ in the direction parallel to the real axis. The vertical boundaries on the left and right of the fundamental domain are identified as is the case for the standard fundamental domain of $\mathrm{SL}(2,\mathbb{Z})$. However, the identifications in the bottom are much more subtle. Fig. 2.3 shows the lower fundamental domain of $\Gamma^0(7)$ with the appropriate identifications indicated [128].

Recall the metric on the moduli space of one torus (see e.g. (2.28)),

$$\mathrm{d}s^2 = \frac{\mathrm{d}u^2 + \mathrm{d}v^2}{4v^2}\,, \tag{2.31}$$

where $u$ is identified with the real and $v$ with the imaginary part of the relevant complex structure modulus. This metric is the natural metric on the space of all tori with fixed volume. The upper complex half plane equipped with this metric is the hyperbolic plane. Fundamental domains of $\mathrm{SL}(2,\mathbb{Z})$ and its congruence subgroups can therefore be viewed as subsets of this plane (with appropriate identifications of boundaries). They can have different topologies (non-trivial genus), cusps and conical singularities [129]. A qualitative picture of such a Riemann surface is shown in Fig. 2.4. The throats in the picture correspond to the cusps in the fundamental domain where it extends to the real axis. Also, the point at infinity in the complex half plane gives rise to such a throat. As one can see in Fig. 2.5 for the congruence subgroup $\Gamma^0(12)$, there may be several of these cusps. The picture also clearly shows the widened fundamental domain, now by a factor 12, compared to the fundamental domain of a torus.

Let us now discuss the potentially long axionic directions corresponding to lines of $\mathrm{Im}\,\tau_1 = \text{const}$. Using the metric (2.31) we see that the length of these lines increases with decreasing $\mathrm{Im}\,\tau_1$. However, the smallest value of $\mathrm{Im}\,\tau_1$ that allows for a straight unbroken line is $\mathrm{Im}\,\tau_1 = 1$. This is a direct consequence of the complicated structure of the fundamental domain at $\mathrm{Im}\,\tau_1 < 1$. We have already

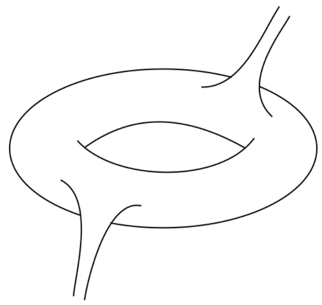

Figure 2.4: A qualitative picture of a fundamental domain of a congruence subgroup as a Riemann surface. The throats correspond to the cusps of the fundamental domain together with the point at infinity.

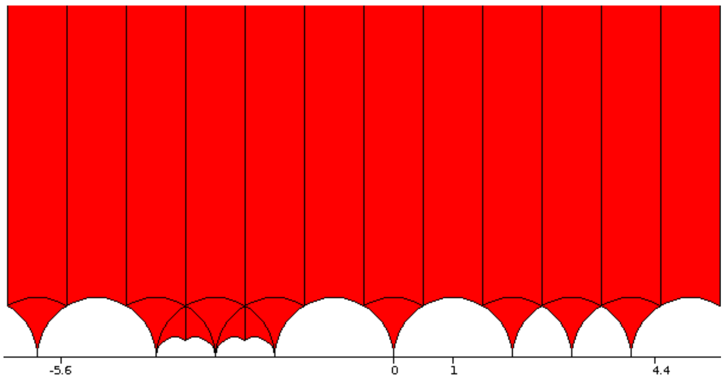


Figure 2.5: A fundamental domain of the congruence subgroup $\Gamma^0(12)$ with several cusps is shown.

calculated the periodicity of this axionic direction to be $N/\sqrt{2}$. In our setting a tadpole condition bounds $N$ by 16 from above which therefore quantifies the maximal length of these axionic directions. We expect that corresponding lengths in more involved compactification on Calabi-Yau manifolds in the large complex structure limit surpass this significantly.

So far this sounds very encouraging. However, as long as there is no potential for $\tau_1$, straight lines defined by $\operatorname{Im}\tau_1 = \text{const.}$ are by no means the most natural paths connecting two points on this line. In fact they are not geodesics with respect to the proper metric (2.31) on moduli space, i.e. there exist shorter paths. It is therefore somewhat arbitrary to declare these non-geodesic paths to be long since one can always generate long paths by means of a detour.

It turns out that geodesics of the hyperbolic plane are given by lines of constant $\operatorname{Re}\tau_1$ and arcs of circles with their center on the real axis (see Fig. 2.6). Let us calculate the length of these geodesics. We start with the straight lines of constant real part and consider only a segment of one of these lines starting at $\mathrm{Im}\,\tau_1 = a$ and ending at $\mathrm{Im}\,\tau_1 = b$. The length is given by

$$L = \int_a^b \frac{dy}{2y} = \frac{1}{2} \ln\left(\frac{b}{a}\right) . \tag{2.32}$$

This is the well-known logarithmic behavior of proper field displacements in moduli space. Now let us calculate the length of an arc of a circle with radius $R$ which starts at a polar angle $\alpha$ and ends at an angle $\beta$. The center of this circle may be located anywhere on the real axis. Parameterizing this path by the polar angle one finds

$$L = \int_\alpha^\beta d\varphi \frac{R}{2R\sin(\varphi)} = \frac{1}{2}\ln\left(\frac{\tan(\beta/2)}{\tan(\alpha/2)}\right) = \frac{1}{2}\ln\left(\frac{1/\sin(\beta) - 1/\tan(\beta)}{1/\sin(\alpha) - 1/\tan(\alpha)}\right) . \tag{2.33}$$

For a symmetric arc with $\beta = \pi - \alpha$ this can be simplified to

$$L = \frac{1}{2}\ln\left(\frac{1+\cos(\alpha)}{1-\cos(\alpha)}\right) . \tag{2.34}$$

Using this formula, we now consider deformations of our long axionic trajectory and determine how short it can become. Indeed, Fig. 2.6 shows the long, closed axionic trajectory as a horizontal line connecting the point $-N/2 + i$ with the (equivalent) point $N/2 + i$. It can be deformed to the arc, also shown in the figure, which again connects this fixed point with itself. For large $N$ and hence small $\alpha$ the result is approximately $L \approx \sqrt{2}\ln(N/2)$, which is clearly much less than our naive expectation in (2.30), which grew linearly with $N$.[5] The upshot is that even if we manage to construct models with large $N$ and hence long axionic directions, we have to be very cautious about the question to which extent these represent large proper distances between points in field space.

One can understand this property pictorially by embedding a section of one of the throats in Euclidean 3-dimensional space (see Fig. 2.7). Note that the axionic direction is the periodic direction around the throat. The shape of the throat is essentially the reason why a simple closed circle around it does not provide the shortest path connecting a point to itself. Instead, we can minimize the length of this circle by pushing it upwards where the circumference of the throat with respect to the embedding space is smaller.

In summary, in spite of the possible $N$-fold widening of one or several throats by the flux, the field space increases only logarithmically with $N$.

## 2.4 Size of the Moduli Space

In the following we want to analyze our model from a 4-dimensional point of view. The idea is to consider the 4d EFT that describes the physics of our model at

[5] Compared to (2.34) this expression for $L$ contains an additional factor $\sqrt{2}$ in order to take the contribution from $\tau_2$ to the length into account, see also (2.28) and (2.29). In the following we will tacitly include this factor in expressions for lengths when appropriate.

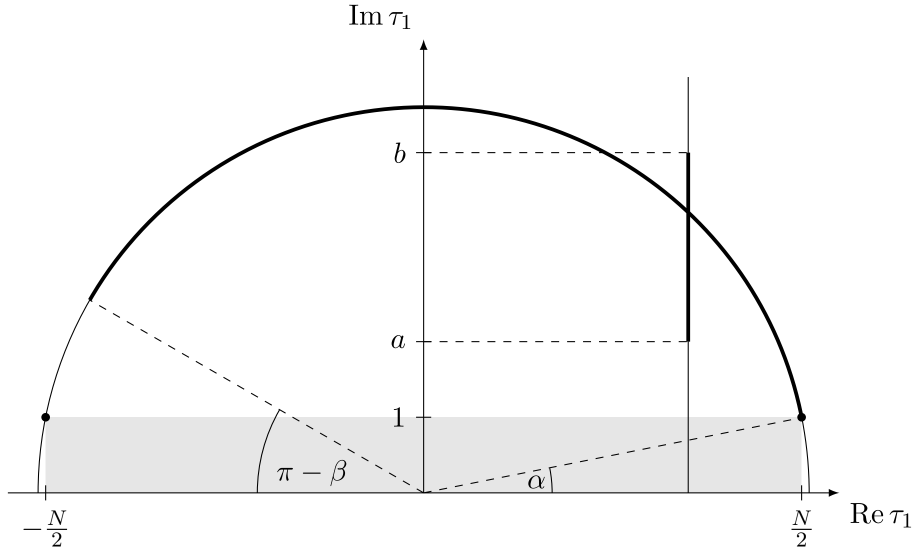


Figure 2.6: The two types of geodesics of the hyperbolic plane are shown: a vertical line and a semi-circle. The segments of these of which the length is calculated in the main text are drawn with thick lines. The shaded region at the bottom corresponds to the region where the fundamental domains of $\Gamma^0(N)$ are in general very complicated (see also Figs. 2.3 and 2.5).

energy scales smaller than a cutoff $\Lambda$, and to determine the regions of moduli space where this theory is valid, i.e. where KK- and winding modes are heavier than the cutoff scale. Once this region has been determined we will introduce a quantitative measure for the size of this region and formulate a conjecture about the dependence of this size on the cutoff.

### 2.4.1 Winding and KK Modes on the Compact Space

Consider the $i$th of our three tori with complex structure modulus $\tau_i$. In (2.5) we have introduced the basis vectors $e_{i,x} = 1$ and $e_{i,y} = \tau_i$ spanning the corresponding lattice in the complex plane. So far, no information concerning the volume is provided. By multiplying $e_{i,x}$ and $e_{i,y}$ by a factor $\sqrt{V_i/\mathrm{Im}\,\tau_i}$ we obtain the basis vectors for a lattice corresponding to a torus with volume $V_i$:

$$\sqrt{\frac{V_i}{\mathrm{Im}\,\tau_i}} \quad \text{and} \quad \sqrt{\frac{V_i}{\mathrm{Im}\,\tau_i}}\tau_i \,. \tag{2.35}$$

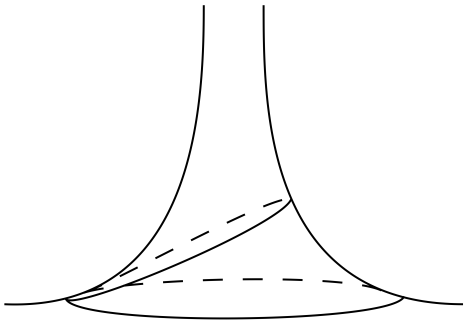

Figure 2.7: The embedding of a throat in 3-dimensional Euclidean space qualitatively shows why a circle around the throat is not the shortest periodic path given a fixed starting point.

These vectors determine the mass $m_{\mathrm{W}}$ of the winding modes on this torus via the formula[6]

$$m_{\mathrm{W}}(n_x, n_y) = \frac{1}{2\pi\alpha'}\sqrt{\frac{V_i}{\operatorname{Im}\tau_i}}|n_x e_{i,x} + n_y e_{i,y}| = 2\pi\sqrt{\frac{V_i}{\operatorname{Im}\tau_i}}|n_x + n_y\tau_i| \tag{2.36}$$

with integers $n_x$ and $n_y$. In the last step we used that in our units $l_{\mathrm{s}} = 2\pi\sqrt{\alpha'} = 1$. Analogously, the dual lattice is spanned by the vectors

$$\frac{\mathrm{i}}{\sqrt{V_i \operatorname{Im}\tau_i}} \quad \text{and} \quad \frac{-\mathrm{i}}{\sqrt{V_i \operatorname{Im}\tau_i}}\tau_i\,, \tag{2.37}$$

and determines the masses $m_{\mathrm{KK}}$ of the KK modes on the torus according to

$$m_{\mathrm{KK}}(n_x, n_y) = 2\pi\frac{1}{\sqrt{V_i \operatorname{Im}\tau_i}}|n_x - n_y\tau_i|\,, \tag{2.38}$$

with, again, integers $n_x$ and $n_y$. Substituting $n_y \to -n_y$ shows that the masses of KK and winding modes differ only by a factor $V_i$.

We achieve equality at the self-dual point $V_i = 1$. This is a convenient choice as it simplifies the analysis regarding the effects of KK and winding modes on the cutoff of the theory. However, for $V_i = 1$ certain 1-cycles in the geometry will necessarily become sub-stringy over large regions of the moduli space of $\tau_i$. In this case, unsuppressed instantons can arise if a string worldsheet or D-brane wraps cycles with sub-stringy volume. They may correct the 4d action, e.g. the Kähler metric. Similarly, light 4d states (particles, strings etc.) can arise from string worldsheets or branes wrapped on small cycles. This may also lead to

[6] The prefactor $(2\pi\alpha')^{-1}$ comes from the Nambu-Goto action $S_{\mathrm{NG}} = (2\pi\alpha')^{-1}\int_{\mathrm{WS}}$.

corrections or affect the value of the cutoff of the effective 4d theory. A complete analysis of the cutoff of the effective theory thus has to take into account KK modes, winding modes as well as instantons and other light states. For a simpler presentation, we will disentangle this as follows. First, in this section we will proceed with the study of the effects of KK and winding modes, working at the self-dual point $V_i = 1$ for simplicity, but ignoring all other corrections and light states. Then we will remove any extra light states and unsuppressed instantons by increasing the volumes $V_i$ such that no sub-stringy cycles remain. As this will also affect the masses of KK and winding modes we will need to modify the analysis of this section, which we will explain in section 2.4.4. It will turn out that this modification is technically straightforward. Having laid out our strategy, we now continue with the analysis for $V_i = 1$.

Now we need to know the mass of the lightest winding mode on the $i$th torus, denoted by $m_{\mathrm{W},i}$, which is equivalent to finding the shortest vector of the lattice spanned by the basis (2.35).[7] This problem is in general not solvable analytically and we will only provide an estimate. First of all, we can apply Minkowski's theorem to this 2-dimensional lattice which will give an upper bound for the length of the shortest lattice vector. According to our choice $V_i = 1$, the area of the parallelogram which is spanned by the basis (2.35) is equal to unity.[8] Then the theorem states that any convex subset of $\mathbb{C}$ that is symmetric with respect to the origin and has a volume larger than four contains a non-zero lattice point. If we choose this subset to be a disk, we can conclude that the shortest lattice vector can not be longer than the radius of this disk. This implies an upper bound of order one for all three tori.

However, there are regions in moduli space in which the true length of the shortest lattice vector is orders of magnitude smaller and we would vastly overestimate the part of moduli space where the low energy EFT is valid. We can improve this situation by analyzing two special regions in which we can find a much better estimate for the length of the shortest lattice vector.

Consider first $\mathrm{Im}\,\tau_i \geq 1$. We have to minimize $(n + m\mathrm{Re}\,\tau_i)^2 + (m\mathrm{Im}\,\tau_i)^2$ with $n, m \in \mathbb{Z}$. For $m \neq 0$, this is larger than unity. For $m = 0$ the minimum is clearly one, realized by the vector $(1, 0)$. The corresponding physical length is $1/\sqrt{\mathrm{Im}\,\tau_i} = m_{\mathrm{W},i}/(2\pi)$.

Second, focus on $|\mathrm{Re}\,\tau_i| \leq \mathrm{Im}\,\tau_i \ll 1$. This always holds at the bottom of the central cusp of the fundamental domain of, for example, $\tau_1$ (see Fig. 2.5). Once again we need to minimize $(n + m\mathrm{Re}\,\tau_i)^2 + (m\mathrm{Im}\,\tau_i)^2$. For $n \neq 0$ the minimum is unity, obtained for $n = 1$ and $m = 0$. If $n = 0$, the shortest lattice vector is simply $\tau_i$, the length of which is smaller than unity. The corresponding physical length is $|\tau_i|/\sqrt{\mathrm{Im}\,\tau_i} \sim \sqrt{\mathrm{Im}\,\tau_i} \ll 1$, giving rise to $m_{\mathrm{W},i} = 2\pi\sqrt{\mathrm{Im}\,\tau_i}$.

In fact, we can extend this result for $i = 1$ to all the other cusps in the funda-

[7] In the following we will only talk about winding modes which in our setting have the same masses as the KK modes. In particular we have $m_{\mathrm{W},i} = m_{\mathrm{KK},i}$.

[8] Note that this volume is independent of the choice of basis.

mental domain of $\tau_1$. Note that, in principle, we can distinguish the cusps due to the flux. However, right now we are only concerned with a pure lattice property, namely the shortest lattice vector, which does not depend on the rest of the physical situation. We can therefore safely ignore the fluxes. This allows us to use the original full modular invariance of the torus to shift all the cusps onto the central one. The result $m_{\mathrm{W},1} = 2\pi\sqrt{\mathrm{Im}\,\tau_1}$ is hence not only valid in the central cusp but also in all the others.

Our complete result for the smallest winding mode mass therefore reads

$$m_{\mathrm{W},i} \sim \begin{cases} 2\pi/\sqrt{\mathrm{Im}\,\tau_i}, & \text{for } \mathrm{Im}\,\tau_i \geq 1 \\ 2\pi\sqrt{\mathrm{Im}\,\tau_i}, & \text{for } \mathrm{Re}\,\tau_i + n \leq \mathrm{Im}\,\tau_i \ll 1 \\ 2\pi, & \text{else} \end{cases} , \tag{2.39}$$

where the integer $n$ is chosen such that $\mathrm{Re}\,\tau_i + n \in (-1/2, 1/2]$.

### 2.4.2 The Restricted Moduli Space

Now we fix the cutoff scale $\Lambda$ with respect to which we want winding modes (and KK modes) to be heavy, i.e. $m_{\mathrm{W},i} > \Lambda$ for all $i$. This condition is only satisfied on a subset of the moduli space which depends on $\Lambda$. We call this subset the restricted moduli space $\mathcal{M}(\Lambda)$ in the following. More precisely, since we take the 4-dimensional point of view, we fix the ratio of the cutoff and the 4-dimensional Planck scale $M_{\mathrm{P}}$,

$$\epsilon \equiv \frac{\Lambda}{M_{\mathrm{P}}} , \tag{2.40}$$

where in our units $M_{\mathrm{P}} = \sqrt{4\pi} g_{\mathrm{s}}^{-1}$ and $g_{\mathrm{s}}$ is the string coupling.[9] In the following we we will restrict ourselves to $g_{\mathrm{s}} < 1$ in order to stay in the perturbative regime. The monodromic moduli space is parametrized by $\{\tau_i\}$ with the vacuum condition imposed and restricted to the appropriate fundamental domains. The next step will now be to determine the region in moduli space that is compatible with the condition $m_{\mathrm{W},i} > \Lambda$ for all $i$, i.e. the restricted moduli space $\mathcal{M}(\Lambda)$.

Let us start by considering $\tau_3$ which is just equal to the axio-dilaton $\tau$ according to the vacuum conditions (2.1). The condition $g_{\mathrm{s}} < 1$ is then equivalent to $\mathrm{Im}\,\tau_3 = \mathrm{Im}\,\tau > 1$. According to (2.39) we need to impose

$$\Lambda = \epsilon M_{\mathrm{P}} = \epsilon\sqrt{4\pi}\mathrm{Im}\,\tau < \frac{2\pi}{\sqrt{\mathrm{Im}\,\tau}} , \tag{2.41}$$

where we used $\tau = \tau_3$ and $\mathrm{Im}\,\tau > 1$. This gives a bound $\mathrm{Im}\,\tau < (\pi^2/\epsilon)^{2/3}$ for the axio-dilaton. Taking into account the appropriate moduli space metric, this is of course consistent with the expected logarithmic growth of moduli space size with $1/\epsilon$. Indeed, we did not try to create long trajectories in the $\tau_3/\tau$-part of moduli space. To simplify our analysis, we will set $\mathrm{Im}\,\tau = \mathrm{Im}\,\tau_3 = 1$ from now on. In

[9] This is due to our choice $V_i = 1$.

this way, we are certain that no light KK or winding modes arise from extreme values of $\tau$ and $\tau_3$.

Next consider $\tau_1$ and $\tau_2$. The vacuum condition for $M = 1$ reads $\tau_1 = N\tau_2$. We choose $\tau_1$ to parametrize the flat directions. Consider first the region defined by $\mathrm{Im}\,\tau_1 \geq 1$. The lightest mode on the first torus in this region has mass $2\pi/\sqrt{\mathrm{Im}\,\tau_1}$ according to (2.39). Requiring $\Lambda < 2\pi/\sqrt{\mathrm{Im}\,\tau_1}$ gives $\mathrm{Im}\,\tau_1 < (2\pi/\Lambda)^2$. The resulting bound on the fundamental domain in the complex $\tau_1$-plane can be visualized as a horizontal line coming down from infinity as we increase $\Lambda$ (see Figs. 2.8 and 2.9).

Now let us focus on the lower part of the moduli space, i.e. $\mathrm{Im}\,\tau_1 < 1$ and in particular on the cusps located near the real axis (see Fig. 2.5). If we go far enough down the cusp we will always satisfy $|\mathrm{Re}\,\tau_1| \leq \mathrm{Im}\,\tau_1$ (possibly after an integer shift along $\mathrm{Re}\,\tau_1$) all the way to the singularity at the real axis. In fact, this condition covers most of the fundamental domain in the regime $\mathrm{Im}\,\tau_1 < 1$ and we will therefore take the resulting constraint on the moduli space to be valid throughout this region. From (2.39) we can read off the lightest winding mass coming from the first torus to be $2\pi\sqrt{\mathrm{Im}\,\tau_1}$ which leads to the bound $\mathrm{Im}\,\tau_1 > (\Lambda/(2\pi))^2$. Similarly to the previously derived bound one can think of this as a horizontal line which now rises from the bottom of the cusps as we increase $\Lambda$. Our final picture of the restricted moduli space is sketched in Fig. 2.8 and Fig. 2.9.

In the previous analysis we have glossed over a subtlety which we want to comment on in the following. So far we have ignored possible bounds coming from the second torus in the last two paragraphs. Now we argue that such bounds do not generically occur throughout the fundamental domain. Ignoring these additional but non-generic constraints will finally lead to an overestimation of the size of $\mathcal{M}(\Lambda)$.

Let us concentrate on the region defined by $|\mathrm{Re}\,\tau_1| \leq \mathrm{Im}\,\tau_1 < N$. Then the vacuum condition $\tau_1 = N\tau_2$ obviously implies $|\mathrm{Re}\,\tau_2| \leq \mathrm{Im}\,\tau_2 < 1$. According to (2.39) we expect the lightest winding mass from the second torus to be $2\pi\sqrt{\mathrm{Im}\,\tau_2} = 2\pi\sqrt{\mathrm{Im}\,\tau_1/N}$. In order to compare this with the corresponding winding masses of the first torus we need to differentiate between two cases.

First focus on $\mathrm{Im}\,\tau_2 < 1/N$, i.e. $\mathrm{Im}\,\tau_1 < 1$. The lightest winding mode on the first torus is then $2\pi\sqrt{\mathrm{Im}\,\tau_1}$. This is heavier than the winding mode on the second torus which therefore provides the strongest bound on the moduli space. Second, consider $1/N \leq \mathrm{Im}\,\tau_2 \leq 1$ or equivalently $1 \leq \mathrm{Im}\,\tau_1 \leq N$. For $\mathrm{Im}\,\tau_1 < \sqrt{N}$ the lightest winding mode on the second torus is in fact lighter than the corresponding mode on the first torus, which has a mass $2\pi\sqrt{\mathrm{Im}\,\tau_1}$. Consequently, the second torus would provide the most important bound on the moduli space in the regime $|\mathrm{Re}\,\tau_1| \leq \mathrm{Im}\,\tau_1 < \sqrt{N}$.

However, the above region covers only the central cusp and a finite part of the upper region of the fundamental domain of $\tau_1$ which does not comprise a substantial part thereof.[10] The corresponding additional bound can hence not be

[10]One might be tempted to extend the validity of this bound to all cusps by using the shift

considered generic and may be safely omitted from our parametric analysis.

### 2.4.3 Estimating the Size of the Restricted Moduli Space

Now we introduce a quantitative measure for the size of the restricted moduli space $\mathcal{M}(\Lambda)$. Since we are interested in distances in field space we may try to use the standard mathematical notion of the diameter. For a Riemannian manifold, in our case $\mathcal{M}(\Lambda)$, it is defined as

$$\mathrm{diam}(\mathcal{M}(\Lambda)) \equiv \sup_{p_1,p_2\in\mathcal{M}(\Lambda)} \inf_{\gamma} L_\gamma(p_1,p_2)\,, \tag{2.42}$$

where the infimum is taken over all curves $\gamma$ that connect the points $p_1$ and $p_2$ and $L_\gamma(p_1,p_2)$ denotes the length of the corresponding path. The quantity $d(p_1,p_2) \equiv \inf_\gamma L_\gamma(p_1,p_2)$ is the usual notion of distance between two points.[11] It is in particular extremal and the corresponding curve must hence be a geodesic. Note that an alternative measure for the size of $\mathcal{M}(\Lambda)$ is its volume which, however, we will not consider in the following.

The technical task now is to estimate the diameter of $\mathcal{M}(\Lambda)$. For the unrestricted moduli space $\mathcal{M}(0)$ it is obvious that points, e.g. in two different throats, can have an arbitrarily large distance, see Fig. 2.4. This is due to the fact that the throats are infinitely long. Now consider $\mathcal{M}(\Lambda)$ with a small $\Lambda$. We will see in a moment that the technical condition is $\Lambda < 4\pi/\sqrt{N}$. In this case, Fig. 2.8 applies. Here we explicitly see that the bounds cut the infinitely long throats. The most widely separated points are still two points in different throats, now pushed up the throat as far as allowed by the bounds.

We have to take two cases into account. Remember that the point at infinity in the $\tau_1$-plane as well as the cusps at the bottom of the fundamental domain correspond to throats. Connecting a point $A_1$ in the upper throat to a point $A_2$ in one of the throats at the bottom yields a potentially long geodesic which is drawn in Fig. 2.8 as a vertical line. The length of this geodesic is, according to (2.32), given by

$$d(A_1,A_2) = 2\sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right)\,. \tag{2.43}$$

The second possibility is to consider two points $B_1$ and $B_2$ which lie in two different cusps, i.e. in two throats at the bottom of the fundamental domain. They are connected by an arc-shaped geodesic as shown in Fig. 2.8. Using (2.34) the length $d(B_1,B_2)$ of this path can be estimated by

$$d(B_1,B_2) = 2\sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right) + \sqrt{2}\ln\left(\frac{N}{2}\right)\,, \tag{2.44}$$

symmetry of the second torus as was done in the last subsection for the first. The following argument shows why this is not possible: Consider a point in one of the cusps other than the central one. Then we have $|\mathrm{Re}\,\tau_1| > \mathrm{Im}\,\tau_1$ and hence also $|\mathrm{Re}\,\tau_2| > \mathrm{Im}\,\tau_2$. Now, in contrast to $\tau_1$, it is not possible to shift $\tau_2$ such that $|\mathrm{Re}\,\tau_2| \leq \mathrm{Im}\,\tau_2$ holds because we already have $|\mathrm{Re}\,\tau_2| \leq 1/2$ (remember that $|\mathrm{Re}\,\tau_1| \leq N/2$).

[11] We will see below that in our physical situation this requires adjustment.

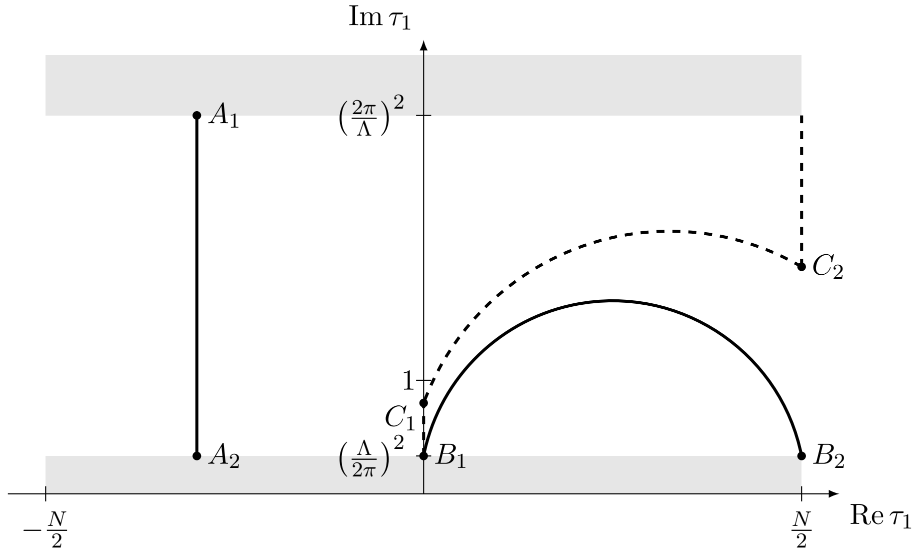


Figure 2.8: The constraints on the moduli space for $\Lambda < 4\pi/\sqrt{N}$ are shown. This picture must be superposed with an appropriate fundamental domain of the congruence subgroup $\Gamma^0(N)$ in order to explicitly see the restricted moduli space. The gray shaded region is excluded by the lower and upper bounds given by $(\Lambda/(2\pi))^2$ and $(2\pi/\Lambda)^2$. In the text we calculate the length of the paths shown.

which is clearly larger than $d(A_1, A_2)$. Hence we conclude that for $\Lambda < 4\pi/\sqrt{N}$ the diameter of the moduli space is bounded by $2\sqrt{2}\ln(2\pi/\Lambda) + \sqrt{2}\ln(N/2)$.

Note that, in principle, the distance between the two points lying in different cusps may actually be smaller than this. It is conceivable that, due to the complicated topology of the central part of $\mathcal{M}(\Lambda)$, a shortcut between the two throats exists which has a length much below $2\sqrt{2}\ln(2\pi/\Lambda) + \sqrt{2}\ln(N/2)$. However, taking (2.43) into account, the diameter of moduli space can not be smaller than $2\sqrt{2}\ln(2\pi/\Lambda)$.

Next consider $\Lambda \geq 4\pi/\sqrt{N}$. This situation is depicted in Fig. (2.9). The formula for the distance between $A_1$ and $A_2$ remains the same as in the previous discussion. However, in the figure one can see that the upper bound cuts part of the arc-shaped geodesic between $B_1$ and $B_2$. It is therefore not a path that determines the distance between its two endpoints any more. Instead, according to our original definition of distance, we must deform it in such a way that it lies completely within $\mathcal{M}(\Lambda)$ and has minimal length. This procedure will, however, lead to an increased distance between the points $B_1$ and $B_2$ because any deformation of this geodesic will increase its length. From a physical point of view this behavior is contrary to our expectation that $\text{diam}(\mathcal{M}(\Lambda))$ is a monotonically decreasing function of $\Lambda$. In the following we present two different meaningful modification

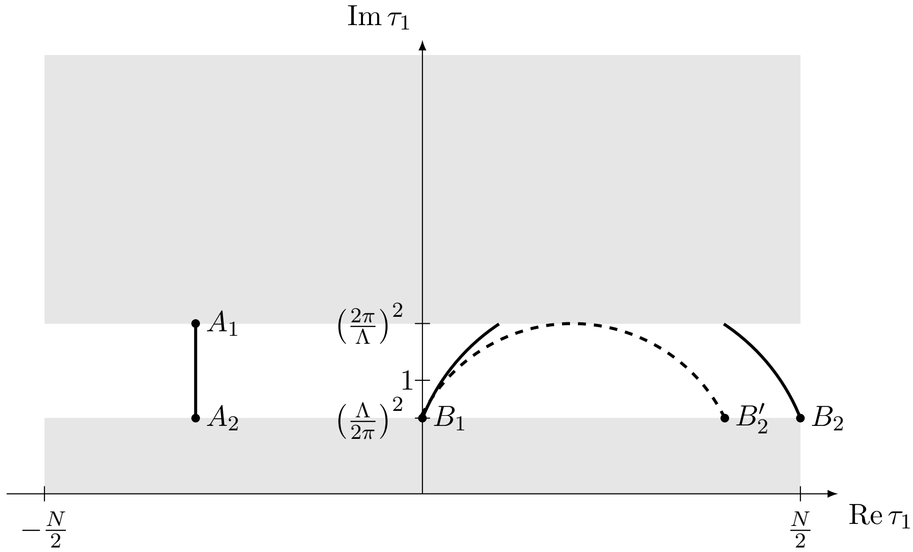


Figure 2.9: The constraints on the moduli space for $\Lambda \geq 4\pi/\sqrt{N}$ are shown. The gray shaded region is excluded by the lower and upper bounds given by $(\Lambda/(2\pi))^2$ and $(2\pi/\Lambda)^2$. The upper bound cuts off a part of the arc-shaped geodesic connecting $B_1$ with $B_2$.

of our definition of distance that are free of this drawback.

Note first that the 4d field theory with cutoff $\Lambda$ breaks down at the boundary of $\mathcal{M}(\Lambda)$. Let us take the four-dimensional point of view and assume that, also outside this boundary, a meaningful 4d physical theory exists. In general, it ceases to be a local field theory and we are unable to make definite statements about the geometry of a corresponding larger moduli space. The most conservative approach is then to assume that all unknown distances are zero, in particular, that all pairs of boundary points have zero distance.

This idea can be made mathematically more rigorous. We know that $\mathcal{M}(\Lambda)$ is a subset of the full moduli space $\mathcal{M}(0)$. However, this may be only one of many manifolds of which $\mathcal{M}(\Lambda)$ could in principle be a subset. Let us denote by $\Omega(\Lambda)$ the set of all manifolds $\mathcal{M}$ such that $\mathcal{M}(\Lambda) \subset \mathcal{M}$ as a metric manifold. One can think of $\Omega(\Lambda)$ as parameterizing our ignorance about the true $\mathcal{M}(0)$ as a four-dimensional observer constrained by $\Lambda$. Our proposal for a new definition of a distance $d^*(p_1, p_2)$ between points $p_1, p_2 \in \mathcal{M}(\Lambda)$ is

$$d^*(p_1, p_2) \equiv \inf_{\mathcal{M} \in \Omega(\Lambda)} d_{\mathcal{M}}(p_1, p_2)\,, \tag{2.45}$$

where $d_{\mathcal{M}}$ is the usual distance on $\mathcal{M}$ and points in $\mathcal{M}(\Lambda)$ may be identified with points in $\mathcal{M}$ via an appropriate injection $i : \mathcal{M}(\Lambda) \to \mathcal{M}$. We expect that points at the boundary of $\mathcal{M}(\Lambda)$ are arbitrarily close in an appropriate $\mathcal{M} \in \Omega(\Lambda)$ which leads to the procedure of effectively compactifying all boundary points of $\mathcal{M}(\Lambda)$ to a single point, as was described in the previous paragraph.

In the foregoing discussion we motivated the definition of $d^*$ by assuming that $\mathcal{M}(\Lambda)$ is part of a larger and more complete moduli space. Now we want to take the more radical point of view that, as 4d observers constrained by $\Lambda$, we are not allowed to venture outside the boundary even in principle. It may then be natural to work with a distance

$$d^{\#}(p_1,p_2) \equiv \begin{cases} d(p_1,p_2), & \text{if } p_1 \text{ and } p_2 \text{ are connected by a geodesic} \\ \text{undefined}, & \text{else} \end{cases}, \tag{2.46}$$

i.e. to assume that points which are not connected by a geodesic that completely lies within $\mathcal{M}(\Lambda)$ do not have a well-defined distance and are treated as completely unrelated. In a sense this definition of distance is much simpler and straightforward than our first proposal. The diameter of a general $\mathcal{M}(\Lambda)$, however, does not necessarily have to be a monotonically decreasing function of $\Lambda$ with this definition of distance, although this problem does not arise in our concrete example.

Now that we have discussed two different modified definitions of distance that are better suited to the problem at hand than the usual definition, we have to re-examine the analysis we have already worked out for $\Lambda < 4\pi/\sqrt{N}$. The main difference between $d$ and $d^*$ is that all boundary points are identified to a single point if we use the latter. In particular, this implies that e.g. a point at the upper and a point at the lower boundary in Fig. 2.8 have zero distance. Therefore points at different boundaries are no longer good candidates for a large distance.

Instead, potentially large distances can be achieved between points $C_1$ and $C_2$ (see Fig. 2.8). These are connected by the dashed arc-shaped geodesic as well as by the two dashed vertical geodesics and the boundary. Altogether these three different paths build a closed curve on which $C_1$ and $C_2$ lie. The maximal distance $d^*(C_1, C_2)$ is achieved if the length of the arc-shaped geodesic equals the sum of the lengths of the two vertical lines and at the same time is maximized. Since the analytic solution of this optimization problem is rather cumbersome, we give a qualitative discussion in three different parametric regimes: $2\pi/\Lambda \gg N/2$, $N/2 \gg 2\pi/\Lambda \gg \sqrt{N}/2$, and $\sqrt{N}/2 \gg 2\pi/\Lambda$. We expect the result to capture the essential behavior of $\text{diam}^*(\mathcal{M}(\Lambda))$.[12]

Let us start in the regime $2\pi/\Lambda \gg N/2$. Then the contribution to the length of the arc-shaped geodesic due to its horizontal extension is completely negligible compared to the vertical direction (cf. (2.44)). Therefore, the length of the closed dashed path is to good accuracy given by $2\sqrt{2}\ln(2\pi/\Lambda)$ where we have only taken the vertical direction into account. At the optimum, $C_1$ and $C_2$ divide the path in two equally long parts such that their distance is

$$d^*(C_1, C_2) = \sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right) \quad \text{for} \quad \frac{2\pi}{\Lambda} \gg \frac{N}{2}\,. \tag{2.47}$$

[12] $\text{diam}^*$ and $\text{diam}^{\#}$ are defined as in (2.42) but using $d^*$ and $d^{\#}$, respectively, as the distance instead of $d$.

As we increase $\Lambda$ the contribution of the horizontal direction to the arc-shaped path becomes more and more important. According to (2.44), it can be estimated by $\sqrt{2}\ln(N/2)$, such that it starts to dominate at $2\pi/\Lambda \sim N/2$. Hence, in the regime $N/2 \gg 2\pi/\Lambda \gg \sqrt{N}/2$, the distance $d^*(C_1, C_2)$ is dominated by the length $\sim \sqrt{2}\ln(N/2)$ of the arc-shaped path. In this regime, the vertical positions of $C_1$ and $C_2$ keep adjusting as $\Lambda$ grows such that the vertical path maintains the same length.

The next qualitative change occurs when $\Lambda$ has increased so much that $\sqrt{N}/2 \sim 2\pi/\Lambda$. Now the arc-shaped geodesic is cut by the upper bound and is hence no longer available in the competition with the vertical path. The vertical positions of $C_1$ and $C_2$ have by now moved to $\operatorname{Im}\tau_1 = 1$, where they will stay from now on. Their distance is determined by the corresponding vertical geodesics connecting them to the lower and upper boundary respectively. Thus, in the new regime $\sqrt{N}/2 \gg 2\pi/\Lambda$, this distance is $2\sqrt{2}\ln(2\pi/\Lambda)$. Combining the three regimes we have

$$\operatorname{diam}^*(\mathcal{M}(\Lambda)) \sim \begin{cases} \sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right) & \text{for } \Lambda \ll 4\pi/N \\ \sqrt{2}\ln\left(\frac{N}{2}\right) & \text{for } 4\pi/N \ll \Lambda \ll 4\pi/\sqrt{N} \\ 2\sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right) & \text{for } \Lambda \gg 4\pi/\sqrt{N} \end{cases} . \tag{2.48}$$

Finally we have to repeat this analysis for $d^\#$. For $\Lambda < 4\pi/\sqrt{N}$ our original analysis remains valid and the diameter of $\mathcal{M}(\Lambda)$ is estimated by $\operatorname{diam}^\#(\mathcal{M}(\Lambda)) = 2\sqrt{2}\ln(2\pi/\Lambda) + \sqrt{2}\ln(N/2)$. Once $\Lambda \geq 4\pi/\sqrt{N}$ the arc-shaped geodesic is cut into two parts and the points $B_1$ and $B_2$ are no longer connected by a geodesic (see Fig. 2.9). Widely separated points that have a well-defined distance are now given by $B_1$ and $B_2'$ which are connected by the path shown in Fig. 2.9. Similarly to our original discussion this path provides an upper bound for the distance of the two points. In particular, the radius of the arc-shaped part is equal to $(2\pi/\Lambda)^2$. With (2.34) we calculate the length of this path to be $2\sqrt{2}\ln(2\pi/\Lambda) + 2\sqrt{2}\ln(\sqrt{2}\pi/\Lambda)$. Hence, the diameter of $\mathcal{M}(\Lambda)$ reads

$$\operatorname{diam}^\#(\mathcal{M}(\Lambda)) \sim \begin{cases} 2\sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right) + \sqrt{2}\ln\left(\frac{N}{2}\right) & \text{for } \Lambda < \frac{4\pi}{\sqrt{N}} \\ 2\sqrt{2}\ln\left(\frac{2\pi}{\Lambda}\right) + 2\sqrt{2}\ln\left(\frac{2\sqrt{2}\pi}{\Lambda}\right) & \text{for } \Lambda \geq \frac{4\pi}{\sqrt{N}} \end{cases} . \tag{2.49}$$

Summarizing, we have found that the diameter of the restricted moduli space $\mathcal{M}(\Lambda)$ is estimated by $\ln(1/\Lambda)$ if we ignore order one pre-factors. Remarkably, this was found independently for two different definitions of distance. This is exactly the logarithmic behavior known from the SDC. However, in our case we have a statement about the absolute size of the restricted moduli space instead of a statement about the relative size of KK and winding mode masses at two different points with a given distance.

Before formulating our conjecture let us return to the problem of sub-stringy cycles. The analysis so far has been performed at the self-dual point with all torus volumes chosen to be $V_i = 1$. As a result we cannot avoid cycles with sub-stringy volumes which in turn gives rise to unsuppressed contributions from both worldsheet and brane instantons. To arrive at a robust result for the diameter of $\mathcal{M}(\Lambda)$ these effects need to be accounted for. This is the subject of the next section.

### 2.4.4 Suppression of Worldsheet Instantons

So far we have neglected the effect of worldsheet and brane instantons on our discussion of the size of moduli space. To ensure that we can safely ignore instanton effects, we need to arrange for the geometry not to possess any cycles with sub-stringy volumes. All cycles have to be super-stringy (which is equivalent to requiring that all winding masses are larger than $2\pi$). Most importantly, this can always be achieved by increasing the torus volumes $V_i$ sufficiently. Here we analyze how this will affect the size of the moduli space.

Let us first consider the third torus. Recall that we have set $\operatorname{Im}\tau_3 = 1$ such that $V_3 = 1$ suffices according to (2.35) to make both cycles of $T_3^2$ have exactly string length. However, for the first torus we have to increase the volume $V_1$ to ensure that both cycles on $T_1^2$ are super-stringy. In particular, we require

$$V_1 = \begin{cases} \operatorname{Im}\tau_1, & \text{for } \operatorname{Im}\tau_1 \geq 1\,, \\ 1/\operatorname{Im}\tau_1, & \text{for } \operatorname{Im}\tau_1 < 1\,. \end{cases} \tag{2.50}$$

Now let us turn to torus $T_2^2$. At the end of Section 2.4.2 we have argued that the winding masses coming from the second torus are generically larger than the ones from $T_1^2$. The argument was made for $V_{1,2} = 1$ but remains true for the more general situation $V_1 = V_2$, as is evident from (2.39). Therefore, by choosing $V_2 = V_1$ with $V_1$ given by (2.50), we find that $m_{\mathrm{W},2} > 2\pi$. This ensures that both cycles on $T_2^2$ are super-stringy, at least generically. With these choices for the volumes $V_i$ no sub-stringy cycles remain and instantons can be safely ignored.

There are two points in the analysis in Sections 2.4.1, 2.4.2 and 2.4.3 that need to be modified because of our different choice of volumes. First of all, as we have increased the winding masses beyond the self dual point we also decreased the masses of KK modes accordingly. Hence the KK modes now give rise to the stronger constraints on the validity of the 4d effective theory. Inserting a factor $1/\sqrt{V_1}$ in (2.39) with $V_1$ as in (2.50) we find for the smallest KK mass

$$m_{\mathrm{KK},1} \sim \begin{cases} 2\pi/\operatorname{Im}\tau_1, & \text{for } \operatorname{Im}\tau_1 \geq 1 \\ 2\pi\operatorname{Im}\tau_1, & \text{for } \operatorname{Re}\tau_1 + n \leq \operatorname{Im}\tau_1 \ll 1 \\ 2\pi, & \text{else} \end{cases}. \tag{2.51}$$

Demanding $m_{\mathrm{KK},1} > \Lambda$ we find that the horizontal lines in Figs. 2.8 and 2.9 are no longer at $(\Lambda/(2\pi))^2$ and $(2\pi/\Lambda)^2$ but at $\Lambda/(2\pi)$ and $2\pi/\Lambda$, respectively. Consequently, all expressions regarding the size of the moduli space have to be modified by substituting $(\Lambda/(2\pi))^2 \to \Lambda/(2\pi)$. Note that this replacement does not change the formulae for the diameter significantly since the cutoff $\Lambda$ always appears within a logarithm.

Furthermore we should take into account that $V_{1,2}$ are no longer constant and therefore contribute to the distance traversed in moduli space as we vary $\tau_1$. Indeed, we have so far discussed distances in a submanifold of moduli space defined by fixing the $V_i$ and $\tau_3$ and only varying $\tau_1 = N\tau_2$. By contrast, we now have to consider a submanifold which is non-trivially embedded in the product of Kähler and complex structure moduli spaces as sketched in Fig. 2.10. The contribution to the metric on this submanifold due to the displacement of Kähler moduli can be calculated from the corresponding Kähler potential. For the Kähler moduli sector this is given by

$$\mathcal{K} = -\ln\left[\frac{1}{8}(T_1+\bar{T}_1)(T_2+\bar{T}_2)(T_3+\bar{T}_3)\right] + \dots\,, \quad \text{with} \quad \mathrm{Re}(T_1) = V_2 V_3\,, \quad \text{etc.} \tag{2.52}$$

Using this and (2.50) we find for the metric of the subset of the full moduli space parametrized by $\tau_1$:

$$ds^2 = 2\,\frac{\mathrm{d}\tau_1 \mathrm{d}\overline{\tau_1}}{(\mathrm{Im}\,\tau_1)^2}\,. \tag{2.53}$$

In our original and simplified analysis the metric (cf. (2.29)) was smaller by a factor four with the corresponding distances smaller by a factor of two. Recall that the replacement $(\Lambda/(2\pi))^2 \to \Lambda/(2\pi)$ introduced a factor 1/2 in those terms in (2.49) which involve a logarithm of $\Lambda$. This factor is canceled by the additional factor two from the new metric, such that the sole net effect is the substitution $\ln N \to 2\ln N$ plus non-logarithmic terms. Thus, the introduction of variable volumes does not change our final formulae (2.49) for the diameter of the moduli space significantly.

### 2.4.5 Statement of the Conjecture

**Moduli Space Size Conjecture 1** *Consider a 4d field theory with cutoff* $\Lambda$. *The diameter of the corresponding moduli space (as defined in section 2.4.3) is then of the order* $\sim \ln(1/\Lambda)$.

This formulation is very natural if one is interested in long flat directions in moduli space in the absence of potentials. For example, if one is interested in EFTs for large-field inflation, this theory must be valid at least at the energy scale of inflation given by $H$. Our conjecture implies then that flat directions have at most lengths of the order $\ln(1/H)$. Note that this statement is true only in the absence of potentials and it therefore does not automatically rule out models of large-field inflation with too large $H$.

Another conjecture which is closer to the original SDC is:

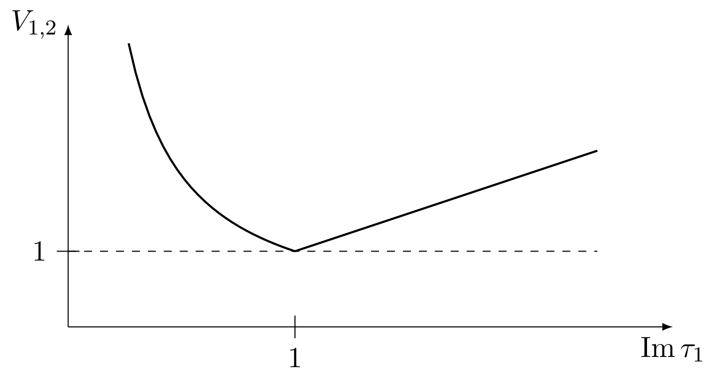


Figure 2.10: The dashed line corresponds to the submanifold of moduli space that is parametrized by $\tau_1$ for constant torus volumes $V_i$. Letting the volumes $V_i$ depend on $\tau_1$ as in (2.50) gives rise to a different submanifold denoted by the solid line. Distances within this submanifold can be calculated by considering the contributions from both the metric on complex structure and Kähler moduli space.

**Moduli Space Size Conjecture 2** *Consider the moduli space of a string theory compactification to four dimensions. Consider two points in this space with a distance L determined by a certain geodesic. Then there exist points on this geodesic at which the lightest KK or winding mode mass is below or of the order* $\exp(-\alpha L)$*, with* $\alpha \sim \mathcal{O}(1)$*.*

A subtle but practically important difference to the SDC is the following: According to our conjecture it is possible to have two points in moduli space which have a large distance and, at the same time, KK and winding modes of the same, high masses. The low-mass or low cutoff situation occurs somewhere in between. This is in particular what happens for points separated in the axionic coordinate (i.e. $\mathrm{Re}\,\tau$ in our explicit model). The lowest cutoff will be experienced at a point along the geodesic connecting the two points, and not at either the beginning or endpoint.

## 2.5 Conclusions

In this chapter we examined the possibility of trans-Planckian field spaces for complex structure moduli in string compactifications employing toroidal orientifolds. The main observation is that by a suitable choice of 3-form fluxes, a certain combination of moduli is lifted, such that the remaining complex flat direction exhibits an enlarged fundamental domain compared to the canonical fundamental domain of a complex structure modulus of a torus. We refer to this as a *monodromic moduli space*.

Mathematically, this moduli space corresponds to the fundamental domain of a congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$. One important observation is that the fundamental domain of such a congruence subgroup is typically widened compared to the canonical fundamental domain of $\mathrm{SL}(2,\mathbb{Z})$. This widening takes the form

$$\mathrm{Re}\,\tau \in [-\tfrac{1}{2}, \tfrac{1}{2}] \quad \longrightarrow \quad \mathrm{Re}\,\tau \in [-\tfrac{N}{2}, \tfrac{N}{2}]\,, \tag{2.54}$$

where $\tau$ is a complex structure modulus and $N$ is an integer set by flux numbers.

We proceeded by examining whether a monodromic moduli space may allow for trans-Planckian field displacements. First we note that 'axionic' trajectories, i.e. trajectories with $\mathrm{Im}\,\tau = \text{const.}$, can become large to the extent that $N$ can. (The tadpole constraint on 3-form fluxes implies $N \leq 16$ in our toy model.) But second we also note that for any two points on such a long (non-geodesic) trajectory much shorter connections exist. They correspond to arcs in the hyperbolic plane and their length scales only as $\ln N$. Moreover, we can restrict our moduli space by demanding that no winding or KK modes appear below a certain cutoff $\Lambda$. It then turns out that an appropriately defined maximal distance between points on an axionic trajectory is not only bounded by $\ln N$ but also by $\ln(1/\Lambda)$. This is reminiscent of the logarithmic limitations of field ranges due to backreaction observed in [32], but here a related phenomenon arises for flat directions.

While we made our observations in a simple string compactification based on a toroidal orientifold, we expect them to hold more widely. To be specific, monodromies also exist in flux compactifications on Calabi-Yau (CY) manifolds, an observation that has been exploited to study moduli dynamics and tunneling between different vacua in the string landscape, see e.g. [111–114]. In our context, the key point is that CY moduli spaces have large complex structure points, analogous to the point at imaginary infinity in the torus fundamental domain. The simplest example is $(\mathrm{T}^2)^3$, where we are dealing with the direct product of three of the familiar throat-like geometries. In general, the geometry near the large complex structure point of a CY is much more complicated, but it always includes 'axionic' directions which characterize short paths around such points. These paths are periodic if one allows for identifications using large diffeomorphisms of the CY. We expect that this periodicity can be enlarged by an appropriate flux choice, analogously to our torus orientifold example. We also expect that the resulting long axionic trajectories will be very far from geodesics, with shortcuts similar to our arcs in the hyperbolic plane. Thus, the qualitative structure of a monodromic moduli space of a CY with 3-form flux should be similar to what we found in this chapter. In the context of inflation, discussions of the moduli space at large complex structure appeared e.g. in [126,130–133]; for recent progress concerning global CY moduli spaces see [134]; for recent work on moduli spaces of CY 4-folds see [135].

The above motivates two conjectures which are related, but distinct from the various forms of the Swampland Conjecture [30,31,33,34]. Given the moduli space of a generic 4d field theory with cutoff $\Lambda$, we conjecture that the absolute size of the moduli space, as measured by the appropriately defined diameter, scales as $\ln(1/\Lambda)$. Alternatively, we may focus on the full moduli space of a certain string compactification. Pick two points in this moduli space which are connected by a minimal geodesic with length $L$. Then we claim that there exist points on this geodesic at which the lightest KK or winding mode mass is smaller or of the order of $\exp(-\alpha L)$, with $\alpha \sim \mathcal{O}(1)$.

One of the key findings of this chapter is that our construction allows for trans-Planckian 'axionic' directions which, however, are not geodesics. In particular, a trajectory along $\mathrm{Re}\,\tau$ for fixed $\mathrm{Im}\,\tau = 1$ is a periodic direction with period $N/\sqrt{2}$. This can be moderately trans-Planckian despite the tadpole constraint on $N$. The upshot is that if it were possible to stabilize $\mathrm{Im}\,\tau$ without completely destroying the structure of the monodromic moduli space, our construction may constitute the first step towards a theory of a trans-Planckian axion.

This is relevant for cosmology where one open question is the compatibility of large-field inflation with QG. It has been suggested that large-field inflation can in principle be embedded in the complex structure moduli sector of string theory compactifications [78, 130–133, 136], as long as there exists a trans-Planckian axionic direction. We suggest that monodromic moduli spaces may be a promising starting point for the construction of such models.

However, there are also obstacles to be overcome: To stabilize $\mathrm{Im}\,\tau$, we require contributions to the potential which may interfere with the proposed simple structure of the monodromic moduli space. Both for this stabilization and to construct a more realistic model of cosmology and particle physics, it is necessary to move beyond simple toroidal orientifolds. While, as noted above, we expect the general structure of the corresponding monodromic moduli spaces of CYs to be similar, the details are far from clear. For example, symmetry structures replacing the modular group and instanton-type (in the mirror dual language) corrections which lift 'axionic' directions non-perturbatively have to be studied.

# 3 The Magnetic Weak Gravity Conjecture for Axions

## 3.1 Introduction

As discussed in Section 1.2 the WGC [50] provides a condition for identifying low EFTs which do not permit a ultraviolet (UV) completion and should hence be assigned to the swampland [30, 31]. In its original form, the WGC is a statement regarding the electric and magnetic particle content of a $U(1)$ gauge theory coupled to gravity. It can be extended to encompass theories with a gauge group consisting of multiple $U(1)$ factors [62] and with charged $p$-branes rather than just particles [50, 63, 76]. A particularly interesting extension of the WGC is that to $(-1)$-branes. The resulting WGC for axions and instantons potentially constrains natural inflation in the field space of one or multiple axions [72–81, 137].

In fact, the WGC can be given in its magnetic and electric form, which are a priori two independent statements. The magnetic WGC arises from requiring the minimally charged magnetic object not to be a black hole or black brane. It can be phrased as a an upper bound for the UV cutoff of the theory. To be specific, for a $(p+1)$-form gauge theory in $d$ dimensions with electrically charged $p$-branes the magnetic WGC requires

$$\Lambda \lesssim \left(g_{\mathrm{e}} M_{\mathrm{P}}^{\frac{d}{2}-1}\right)^{\frac{1}{p+1}}, \tag{3.1}$$

where $g_{\mathrm{e}}$ is the coupling constant of the electric theory. For the case of a theory with particles ($p = 0$) in four dimensions this reduces to the well-known statement $\Lambda \lesssim g_{\mathrm{e}} M_{\mathrm{P}}$.

In this chapter we will focus on the magnetic WGC for axions, i.e. $p = -1$. Naively taking $p = -1$ in (3.1) we find that the exponent diverges. In four dimensions the electric coupling is given by $g_{\mathrm{e}} = 1/f$ where $f$ is the axion decay constant determining the axion period. The r.h. side of (3.1) involves $(M_{\mathrm{P}}/f)$ raised to a diverging power. Clearly, for $f < M_{\mathrm{P}}$ this gives no constraint on the cutoff. By contrast, for $f > M_{\mathrm{P}}$ the inequality implies $\Lambda \to 0$, i.e. the theory simply does not exist. As $f > M_{\mathrm{P}}$ is necessary for large-field inflation, this would imply that the magnetic WGC censors large-field axion inflation.

In what follows, we want to go beyond this simple estimate and develop a more rigorous argument based on the magnetic WGC. As the magnetic WGC is concerned with the existence and the properties of magnetically charged objects, we will study these objects in detail. For an electric theory of instantons coupled to an axion, the corresponding magnetic object is a string coupled to a 2-form field. Thus, in this work, we will study (cosmic) string solutions in a theory with $f > M_{\mathrm{P}}$ explicitly (see also [138] for related work). We take the following statement as the preliminary definition of the magnetic WGC for axions: For the magnetic WGC to be satisfied we require the minimally charged (cosmic) string to exist as a field-theoretic object. The task is hence to examine whether string solutions for $f > M_{\mathrm{P}}$ suffer from any pathologies. Our most promising candidates will be topological inflation and and a composite string à la [67, 139].

We begin by analyzing the static spacetime solution of the exterior of an axionic string [140] (see [53, 54, 141] for discussions related to such strings from a string theory perspective). There are two problems which put the existence of this object in doubt for $f > M_{\mathrm{P}}$.

- There is a singularity at a finite distance from the string core. This singularity even persists for $f < M_{\mathrm{P}}$, but would then be exponentially far away.
- For $f > M_{\mathrm{P}}$ the deficit angle around the string is always negative. This makes it impossible to attach such a spacetime to that of a proper UV completion of the string core since we expect this to be locally flat at the string's center. In contrast, for $f < M_{\mathrm{P}}$ there would have been a finite region with a positive deficit angle as expected in the vicinity of a string.

Hence, the *static* string solution [140] does not seem to exist for $f > M_{\mathrm{P}}$ and thus does not provide us with the magnetic object required for satisfying the WGC.

The problem of the singularity at a finite distance can be resolved by considering a string solution with a dynamical spacetime [108, 109]. Interestingly, this dynamical solution exists for $f > M_{\mathrm{P}}$ up to a maximal value $f_{\mathrm{max}}$ which lies in the range $6 \leq f_{\mathrm{max}}^2/M_{\mathrm{P}}^2 \leq 12$. However, it remains questionable whether the dynamical solution can be interpreted as a string for $f > M_{\mathrm{P}}$. Instead, for $f > M_{\mathrm{P}}$ the Hubble length becomes comparable to the size of the string core leading to an expansion of the defect in all directions. This is known as topological inflation [142, 143]. It then remains to be checked whether this solution can play the role of a bona fide string for an observer outside the inflating core.

However, accepting topological inflation as a UV completion leads to further puzzles. Topological inflation can also arise in 1-form theories where the inflating defect is a magnetic monopole. For example, consider the theory giving rise to such monopoles used by Linde in [142]. This is an SU(2) Yang-Mills theory which is broken down to U(1) by an adjoint scalar. Let $g \ll 1$ be the gauge coupling, $\lambda$ the scalar field self-coupling, and $v$ the symmetry breaking vacuum expectation value. The size of the magnetic monopole is given by the maximum of the two length scales $(gv)^{-1}$ and $(\sqrt{\lambda}v)^{-1}$ where $gv$ is the gauge boson mass and $\sqrt{\lambda}v$ is the scalar mass. The corresponding cutoff scale $\Lambda$ of the monopole is $\min\{gv, \sqrt{\lambda}v\}$. Let us try to make $\Lambda$ as large as possible while keeping the theory under control, i.e. let us make sure that energy densities remain sub-Planckian. We must therefore ensure that $\lambda v^4 \leq M_\mathrm{P}^4$. This optimization problem is solved by $\sqrt{\lambda} = g$ and $v = M_\mathrm{P}/\sqrt{g}$. The resulting cutoff $\Lambda = \sqrt{g} M_\mathrm{P}$ is sub-Planckian but exceeds the WGC bound $g M_\mathrm{P}$. Since $v > M_\mathrm{P}$, the corresponding solution will inflate. If we accept this as the UV completion of the minimally charged monopole required by the WGC, the argument for the low cutoff, $\Lambda \sim g M_\mathrm{P}$, is lost. Indeed, the cutoff explicitly found above is higher. One is then left with the unsatisfactory situation that topological inflation ensures that the magnetic WGC remains satisfied in axion theories, while for 1-form theories it would allow for an explicit violation. This can be avoided by slightly strengthening the requirement underlying the magnetic WGC. Note that besides the inflating monopole solution a minimally charged black hole monopole still exists in our setting. If the correct formulation of the magnetic WGC is that no minimal magnetic monopole should be a black hole, the above situation remains forbidden. One might then be tempted to conclude that topological inflation resides in the swampland.

Thus, having studied both static and dynamical string solutions, our analysis leaves us with the following results:

- If topologically inflating spacetimes are acceptable as the magnetic objects required by the WGC, then $f > M_\mathrm{P}$ cannot be ruled out in the present approach.
- By contrast, if topologically inflating spacetimes are not accepted as magnetically charged objects of axion models, we have good reason to believe that $f > M_\mathrm{P}$ is forbidden. Such a viewpoint can be argued from the fact that the relevant solutions are non-static or from the presence of a horizon. The conclusion concerning large $f$ would then be as negative as in the naive approach mentioned in the beginning.

Last, we turn to axion theories with $f_\mathrm{eff} > M_\mathrm{P}$ which admit a UV completion in terms of a theory of two axions with $f_{1,2} < M_\mathrm{P}$ [92,144,145]. As suggested in [78] using a stringy example, a particularly promising way to realize the required winding trajectory is through 3-form gauging [146] (for applications of this approach in the context of axion monodromy inflation see [147, 148]).[1] As $f_{1,2} < M_\mathrm{P}$, the corresponding magnetic objects exist and avoid the problems mentioned above. Following [67, 139], one can then identify a string solution as bound state of the strings of the original two axions connected by domain walls. This object is then checked explicitly for possible pathologies.

One observation is that the tension of this effective string becomes super-Planckian if $f > M_\mathrm{P}$. The tension can be estimated as $T_\mathrm{eff} \sim f_\mathrm{eff}^2$ and hence one leaves the regime of weak gravity backreaction once $f_\mathrm{eff}$ is sufficiently large. A modestly super-Planckian $f_\mathrm{eff}$ may be possible, but this will depend on the

[1] This corresponds to the familiar Stückelberg mechanism in the 1-form case and can be used to obtain a small effective gauge coupling [67].

precise numerical factors. Our finding is consistent with the previous results for explicit string solutions. Static string solutions for parametrically large $f$ do not seem to exist. Dynamical solutions are possible, but we have to leave their detailed study in this particular case to future work. While they may be similar to topological inflation, they could equally produce a singularity or exhibit a completely different, unexpected behavior.

## 3.2 Naive Estimates

The magnetic WGC follows from the requirement that the minimally charged magnetic object of a U(1) gauge theory is not a black hole. The mass $M$ of a monopole in four dimensions can be estimated by its field energy, leading to $M \sim g_{\rm m}^2\Lambda$, where $g_{\rm m}$ denotes the magnetic coupling constant and $\Lambda$ is a cutoff that determines the radius of the core. The electric coupling constant is related to $g_{\rm m}$ by $g_{\rm e} \sim g_{\rm m}^{-1}$. If the core radius is larger than the Schwarzschild radius, the monopole is not a black hole. The corresponding formula is $\Lambda^{-1} \geq M/M_{\rm P}^2 \sim g_{\rm e}^{-2}\Lambda/M_{\rm P}^2$, implying $\Lambda \lesssim g_{\rm e}M_{\rm P}$.

This easily generalizes to an 'electric' $(p+1)$-form gauge theory with 'magnetic' $(d-p-4)$-branes in $d$ dimensions. The electric coupling constant $g_{\rm e}$ of this system has mass dimension $p+2-d/2$ and the magnetic coupling is $g_{\rm m} \sim g_{\rm e}^{-1}$. The field energy of the magnetic $(d-p-4)$-brane is proportional to $g_{\rm m}^2$ and therefore, analogously to the monopole case, we can estimate the tension by

$$T \sim g_{\rm m}^2\Lambda^{p+1}, \tag{3.2}$$

using dimensional analysis. On the other hand, the tension of a black brane is proportional to the inverse coupling constant $\kappa_d^{-2} = M_{\rm P}^{d-2}$ of gravity. Hence

$$T_{\rm BH} \sim M_{\rm P}^{d-2}R_{\rm S}^{p+1}, \tag{3.3}$$

where $R_{\rm S}$ is the Schwarzschild radius. According to the magnetic WGC we need to impose $\Lambda^{-1} \geq R_{\rm S}$. Expressing $R_{\rm S}$ through $T_{\rm BH}$ and using $T_{\rm BH} \sim T$ gives

$$\Lambda \lesssim (g_{\rm e}^2 M_{\rm P}^{d-2})^{\frac{1}{2(p+1)}}. \tag{3.4}$$

We find it particularly intuitive to rewrite this in terms of the 'characteristic energy scale' or 'strong coupling scale' $\Lambda_{\rm e}$ of the electric gauge theory, defined by $g_{\rm e}^2 = \Lambda_{\rm e}^{2(p+2)-d}$. One finds

$$\frac{\Lambda}{M_{\rm P}} \lesssim \left(\frac{M_{\rm P}}{\Lambda_{\rm e}}\right)^{\frac{d-2(p+2)}{2(p+1)}}. \tag{3.5}$$

If $2(p+2)-d < 0$, the gauge theory is infrared-free and $\Lambda_{\rm e}$ is its intrinsic UV cutoff. For small enough coupling, the energy scale $\Lambda_{\rm e}$ exceeds $M_{\rm P}$ and naively $M_{\rm P}$ would now be the cutoff of the theory. However, if $p+1 > 0$, the magnetic

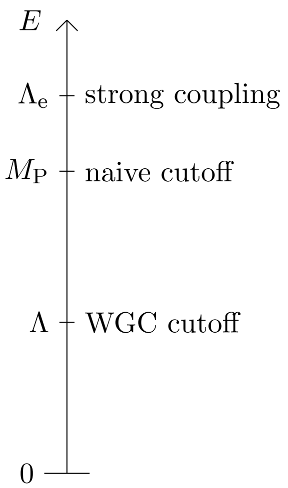


Figure 3.1: Cutoff scales of a weakly coupled gauge theory in the presence of gravity.

WGC (3.5) predicts a cutoff that is smaller than the Planck mass. This situation is sketched in Fig. 3.1. The condition $p+1 > 0$ can be rewritten as $d_{co} > 2$, where $d_{co} \equiv p+3$ is the co-dimension of the magnetic brane.

If $d_{co} = 2$ the picture in Fig. 3.1 applies but in an extreme case: The exponent in (3.5) diverges and, in the weakly coupled case $\Lambda_e > M_P$, the cutoff $\Lambda$ vanishes. Such a situation might be interpreted by saying that the theory does not exist, in other words, the weakly coupled case is forbidden. This occurs for example for the string ($p = -1$) in four dimensions where the magnetic coupling constant is given by the axion decay constant $f = g_m$. Weak coupling corresponds to $f > M_P$ in this case, which should hence be impossible to realize. In the following we want to check this statement by trying to explicitly construct a string with $f > M_P$ in four dimensions.

Before doing so, let us perform a very rough calculation in order to gain some intuition for what one can expect from a more detailed analysis. The field energy density of the string is $\rho \sim f^2/r^2$. This gives

$$T \sim \int_0^{2\pi} d\varphi \int_{\Lambda_{UV}^{-1}}^{\Lambda_{IR}^{-1}} r dr \rho \sim f^2 \ln(\frac{\Lambda_{UV}}{\Lambda_{IR}}), \tag{3.6}$$

where we had to introduce two cutoff scales $\Lambda_{UV}$ and $\Lambda_{IR}$ in order to render the contribution to the string tension finite. Let us now discuss the inclusion of gravity. The static vacuum solution of Einstein's equations with cylindrical symmetry has a deficit angle $\Delta\phi$ in the plane perpendicular to the symmetry axis, i.e. it describes a cone:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}z^2 + \mathrm{d}r^2 + r^2(1 - \frac{\Delta\phi}{2\pi})^2 \mathrm{d}\theta^2. \tag{3.7}$$

This is the exterior spacetime of a string without charge [149] and

$$\Delta\phi = \frac{T}{M_{\mathrm{P}}^2}. \tag{3.8}$$

For $T \gtrsim M_{\mathrm{P}}^2$, i.e. $\Delta\phi \gtrsim 2\pi$, this spacetime breaks down (see [150–152] and in particular [153] for a good review). Therefore, since (3.6) indicates that $T \gtrsim f^2$, one can expect that corresponding axionic string spacetimes do not exist.

In fact, one can say more: Imagine that the field of the charged string is switched off at a distance $R$ from the string. The corresponding total tension of this configuration is estimated by (3.6), where the upper integration limit is now given by $R$. At distances $r > R$ the vacuum solution (3.7) describes this spacetime, with deficit angle given by (3.8). Now, repeating this argument for another radius $R' > R$ one immediately sees that the corresponding deficit angle is larger than that for $R$. We therefore see that the deficit angle of the spacetime of a charged string is not constant but grows with the distance to the string. Thus, one expects a spacetime which is locally conical but eventually breaks down when $\Delta\phi > 2\pi$. Due to the logarithmic behavior, this breakdown happens at exponentially large distance for $f \lesssim M_{\mathrm{P}}$. By contrast, it occurs instantaneously if $f \gtrsim M_{\mathrm{p}}$.

## 3.3 Singular String Spacetimes

The first exact solution of the Einstein equations for the exterior of an axionic string was given by Cohen and Kaplan (CK) [140]:

$$\mathrm{d}s^2 = \frac{u}{u_0}(-\mathrm{d}t^2 + \mathrm{d}z^2) + \gamma^2 \left(\frac{u_0}{u}\right)^{1/2} \exp\left(\frac{u_0^2 - u^2}{u_0}\right)(\mathrm{d}u^2 + \mathrm{d}\theta^2). \tag{3.9}$$

Here $\gamma$ is an integration constant with the dimensions of length, $u_0$ is related to the axion decay constant[2] by $u_0 \equiv 2M_{\mathrm{P}}^2/f^2$, and $0 \leq \theta < 2\pi$. In these coordinates $u = \infty$ corresponds to the singular center of the string. At $u = 0$, which corresponds to a cylindrical surface concentric with the string, one encounters another singularity. Both singularities are physical, as testified by the Kretschmann scalar

$$K = R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = \frac{\exp\left(\frac{2}{u_0}(u^2 - u_0^2)\right)}{4\gamma^4 u^3 u_0^3}(32u^4 - 8u_0u^2 + 3u_0^2), \tag{3.10}$$

[2]We use a slightly different normalization of the axion decay constant, hence the presence of an additional factor 2 compared to the original paper.

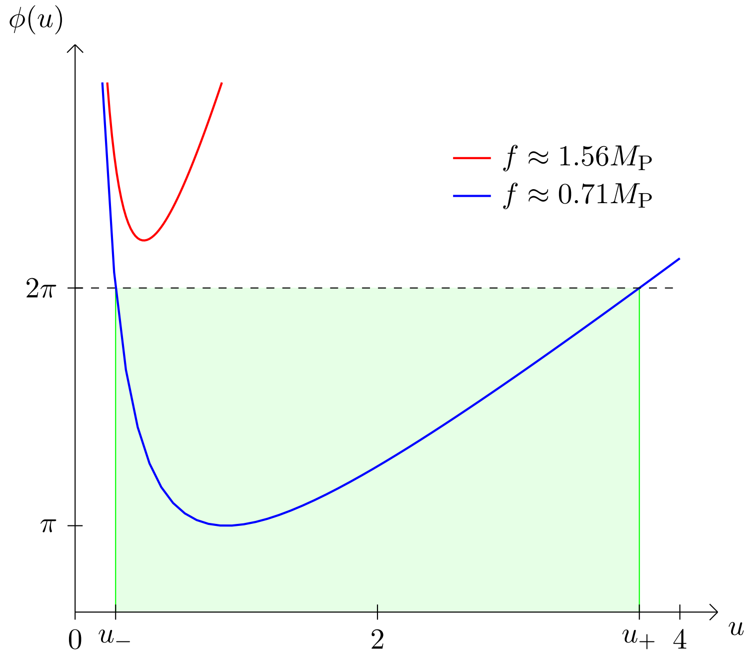


Figure 3.2: The angle $\phi$ defined in (3.12) is shown as a function of the coordinate $u$ for $f \approx 0.71 M_\mathrm{P}$ ($u_0 \approx 4$) and $f \approx 1.56 M_\mathrm{P}$ ($u_0 \approx 0.83$). The shaded green area indicates the coordinate and angle range where one can visualize the space locally as conical, with a positive deficit angle (for $f \approx 0.71 M_\mathrm{P}$). $u = 0$ corresponds to the outer singularity while the string center sits at $u = \infty$. The right intersection point of the blue and black dashed line at $u = u_+$ is roughly the core radius. For $f \approx 1.56 M_\mathrm{P}$ the angle is always larger than $2\pi$.

which diverges at these points. The proper distance between the outer singularity and the center is finite and reads

$$r_{\max} = \gamma \mathrm{e}^{u_0/2} \left(\frac{u_0}{2}\right)^{5/8} \Gamma\left(\frac{3}{8}\right) . \tag{3.11}$$

For small $f$ this distance is exponentially large and might not be physically relevant, e.g. in a cosmological setting. However, for $f \gtrsim M_\mathrm{P}$ one expects problems and we will discuss this momentarily.

In order to gain some intuition for the CK spacetime we note that the region between $u + \mathrm{d}u$ and $u$ at $t, z = \text{const.}$ describes an annulus. Its geometry approximates a piece of a cone and we find it convenient to denote the angle of the lateral surface[3] of this cone by $\phi(u)$. In other words, $2\pi - \phi(u)$ is its deficit angle. A more formal definition is as follows: Consider a closed curve defined by $t, z, u = \text{const.}$

[3] By this we mean the central angle of the corresponding circular segment.

and parametrized by the coordinate $\theta$. We can determine the tangent vector $v$ to this curve at $\theta = 0$ and parallel transport it along the curve until we reach $\theta = 2\pi$, which is of course our starting point. Then calculate the angle between the parallel transported tangent vector $v'$ and $v$. This angle is $\phi(u)$. It turns out that $\phi = dC/dr$, where $C$ is the proper circumference and $r$ the proper radial distance. For the CK spacetime we find

$$\phi(u) = \frac{dC}{dr} = 2\pi \left( \frac{1}{4u} + \frac{u}{u_0} \right). \tag{3.12}$$

For $f < \sqrt{2} M_{\rm P}$ there is an interval of $u$ where $\phi(u)$ is increasing with $u$ and smaller than $2\pi$. Recall that increasing $u$ corresponds to decreasing distance to the string center (see Fig. 3.2). This is exactly the behavior we naively argued for at the end of Section 3.2. However, the angle is not bounded from above and there exist regions where it exceeds the critical value $2\pi$. To be specific, we find $\phi > 2\pi$ for $u < u_-$ and $u > u_+$ where

$$u_\pm = \frac{u_0}{2} \left( 1 \pm \sqrt{1 - \frac{1}{u_0}} \right). \tag{3.13}$$

For $u_0 \gg 1$ one obtains $u_+ \approx u_0$ and $u_- \approx 1/4$. We also see that for $u_0 < 1$ (which corresponds to $f > \sqrt{2} M_P$) the expressions $u_\pm$ become complex and the angle is strictly greater than $2\pi$ throughout the spacetime. We will come back to this point when assessing the existence of the CK solution. In any case, the minimum angle is reached at $u = \sqrt{u_0}/2$ and is given by $\phi_{\rm min} = 2\pi/\sqrt{u_0}$. This is illustrated in Fig. 3.2.

We are now in a position to address the question under which circumstances the CK solution can be trusted and corresponds to a physically acceptable geometry. For example, we should discard solutions where curvature invariants are super-Planckian everywhere. Thus a necessary condition for an acceptable solution is that there exist regions where the Kretschmann scalar $K$ is sub-Planckian.[4] The minimal requirement for this to be possible is that $K$ is sub-Planckian at its minimum. This minimum lies at $u \approx \sqrt{u_0}/2$ and we arrive at the constraint

$$\gamma \gtrsim \frac{1}{M_{\rm P}} \frac{{\rm e}^{-u_0/2}}{u_0^{5/8}}. \tag{3.14}$$

For $f \lesssim M_{\rm P}$ the right-hand-side of this inequality is exponentially suppressed and hence a wide range of $\gamma$ gives rise to a weakly curved spacetime. However, also for $f \gtrsim M_{\rm P}$ the choice of a sufficiently large $\gamma$ will ensure the existence of a weakly curved region of the CK spacetime. Eq. (3.14) also ensures that the radial size (3.11) of the spacetime obeys $r_{\rm max} \gtrsim M_{\rm P}^{-1}$. From the explicit form (3.10) of $K$ we see that we can make the region of small curvature between the inner and outer

[4]Here we use the Kretschmann scalar instead of the scalar curvature $R$ since $K > R^2$ for all $u$. In particular, $R$ does not have a singularity at $u = 0$ while $K$ does.

singularity always as large as desired by choosing $\gamma$ large enough. Pictorially, by increasing $\gamma$ we 'stretch' the spacetime in the radial direction and thereby flatten its geometry.

Another criterion for deciding which solutions to trust employs the angle $\phi$. Although we are primarily interested in $f \gtrsim M_{\rm P}$ let us examine for completeness first $f \leq \sqrt{2}M_{\rm P}$, i.e. $u_0 \geq 1$. For an uncharged string we saw in Section 3.2 that the deficit angle is growing with the tension, i.e. the corresponding angle $\phi$ decreases. By this we argued that the angle of the spacetime of a global string should decrease with the distance to the string center, which turned out to be correct at least far away from the outer singularity. This motivates us to conjecture that this behavior persists in every UV completion of the core. Additionally, in order to avoid a conical singularity at the string center we expect every UV completion to satisfy $\phi(\infty) = 2\pi$. Combining this with our conjecture we can conclude that the angle of any UV completion does not exceed $2\pi$.

Such a UV completion must somehow be matched onto the exterior CK spacetime at the core radius. Our upper bound on the angle in the core implies a lower bound for the core radius: $u_{\rm core} \leq u_+$. Recall that $u_+$ corresponds to the point where the angle would exceed $2\pi$ when approaching the string core, as indicated by the dashed line in Figure 3.2. For $f \leq \sqrt{2}M_P$ we find $u_+ \sim u_0$ and hence $u_{\rm core} \lesssim u_0$. Expressing this in the radial proper distance one finds $r_{\rm core} \gtrsim \gamma$. The core radius can only take this minimal value if the Kretschmann scalar at $u \sim u_0$ is sub-Planckian, which yields a lower bound on $\gamma$: $\gamma \gtrsim f/M_{\rm P}^2$.[5] Altogether, there is no fundamental obstacle to the existence of a UV completion of the string core for $f \leq \sqrt{2}M_P$.

The situation is completely different for $f > \sqrt{2}M_{\rm P}$, i.e. $u_0 < 1$. In this case we have $\phi > 2\pi$ for all $u$. It is then not clear at all how the CK solution could be matched with a UV completion of the core. A possible conclusion is that a UV completion does not exist and hence the whole solution should be discarded. Recall that our main motivation is to gain a better understanding of the magnetic WGC for axions where the magnetically charged objects are given by strings. Our analysis in this section can thus be summarized as follows. For $f > \sqrt{2}M_{\rm P}$ the CK string does not give rise to a trustworthy solution which could correspond to the magnetic object in the WGC for axions. This implies that either there are no string solutions for $f > \sqrt{2}M_{\rm P}$, or string solutions exist but are not of CK type. We will explore the latter possibility in the next section. Note, however, that this conclusion crucially depends on the validity of our conjectured condition on the angle of UV completions. This may well be too strong a requirement and if not true, CK solutions with $f > \sqrt{2}M_{\rm P}$ might after all exist. For example, one may think of a higher dimensional UV completion, where the requirement $\phi(\infty) = 2\pi$ makes no sense.

Independently of this, even for $f \leq \sqrt{2}M_{\rm P}$, one cannot accept this spacetime as physical due to the naked singularity in the exterior. It may however describe

[5] This lower bound is stronger than (3.14) for $f \lesssim M_{\rm P}$.

part of a cosmological spacetime or of a large string loop.

## 3.4 Non-singular String Spacetimes

It turns out that it is possible to find a string spacetime which is not plagued by any physical singularity. Indeed, this can be done by allowing for a time-dependent metric and was first investigated by Gregory [108]. This analysis uses a UV completion of the axion model that is given by a complex scalar field $\Phi$ the phase of which plays the role of the axion:[6]

$$\mathcal{L} = -\frac{1}{2}|\partial\Phi|^2 - \frac{\lambda}{4}(|\Phi|^2 - f^2)^2. \tag{3.15}$$

The vacuum manifold is $S^1$ and gives rise to the string solutions we are interested in. Note that, although the metric depends on time, the field configuration is chosen to be static in Gregory's calculation, i.e. the string has a constant width.

Let us describe the properties of this solution in more detail. Firstly, Gregory is able to show that the metric takes the form

$$\mathrm{d}s^2 = \mathrm{e}^{A(r)}(-\mathrm{d}t^2 + \cosh^2(\sqrt{b_0}t)\mathrm{d}z^2) + \mathrm{d}r^2 + C^2(r)\mathrm{d}\theta^2, \tag{3.16}$$

where $b_0 > 0$ is a constant, and which in fact exhibits no singularities. We see that the spacetime of the string expands along the direction of the string while radial and angular metric components have no time-dependence. Furthermore, Gregory proves that this spacetime has a cosmological event horizon at finite proper distance from the string. This horizon encompasses the string and allows light to exit this interior space while it is not possible for physical objects to enter it from the exterior.

This event horizon moves inwards as $f$ increases. Therefore, the question arises what happens if the horizon enters the string core. Using analytic arguments, Gregory and Santos [109] show that the string solution ceases to exist for $f > f_{\mathrm{max}}$ with $6 \leq f_{\mathrm{max}}^2/M_{\mathrm{P}}^2 \leq 12$. We see two possible solutions that are left in the regime of parametrically large $f$: First, $\Phi = 0$ together with a de Sitter metric is a classical solution. We discard it because of its instability. But second, and this will be the focus of the rest of this section, there is topological inflation.

Topological inflation has been invented independently by Linde and Vilenkin [142, 143] and occurs for topological defects whose width is comparable to the Hubble radius defined by the energy density in their core. Since the fields within the defects are at positive potential one expects the spacetime to be similar to de Sitter there. Hence, the defect expands exponentially in all directions.

Let us apply the above mentioned condition for topological inflation to the string. In order to do this assume that the UV completion of the axion is given

[6] In contrast to Gregory, we have canonically normalized the complex phase of $\Phi$.

by the potential (3.15). The axion is given by $\varphi \equiv \arg(\Phi)$ and the cutoff of the model is set by the mass of the radial scalar $|\Phi|$:

$$\Lambda^2 \sim m_{|\Phi|}^2 \sim \lambda f^2 \,. \tag{3.17}$$

By balancing the gradient and potential energy density in the string core,

$$\lambda f^4 \overset{!}{\sim} (f/R)^2 \,, \tag{3.18}$$

the radius of the core is seen to be

$$R \sim \frac{1}{\sqrt{\lambda}\, f} \sim \frac{1}{m_{|\Phi|}}. \tag{3.19}$$

This also fixes the tension by integrating either gradient or potential energy density over the transverse section of the core:

$$T \sim \lambda f^4 R^2 \sim (f/R)^2 R^2 \sim f^2 \,. \tag{3.20}$$

In fact, we see that the last result does not depend on the precise UV completion since it can be derived more generally: Simply assume that, to realize a UV completion, the $S^1$ field space parametrized by $\varphi$ is embedded in a flat field space (in our case the $\mathbb{R}^2$ parametrized by $\Phi$) with canonical Euclidean metric. It is then clear that the field must cross a distance $\sim f$ in the string core of size $\sim R$. This implies a gradient energy density $\sim (f/R)^2$ and hence $T \sim f^2$, without any reference to the scalar potential and $\lambda$.

Using this we can also state the condition for topological inflation to occur independently of the concrete potential. Let $V_0$ be the potential energy within the string which would be $V_0 = (\lambda/4) f^4$ in the above example. Now, balancing the gradient energy density with the potential energy density leads to the estimate $R \sim f/\sqrt{V_0}$. The Hubble radius corresponding to the potential energy is given by $H_0^{-1} \sim M_\mathrm{P}/\sqrt{V_0}$. Then, finally, the condition $R \gtrsim H_0^{-1}$ implies $f \gtrsim M_\mathrm{P}$ which perfectly agrees with Gregory's result. Note, however, that our estimate has been derived without specifying the potential of the complex scalar explicitly. Only the embedding of the vacuum manifold $S^1$ into the field space $\mathbb{R}^2$ of the UV completion was assumed.

In [154] the model of Gregory was analyzed numerically for a general time-dependent metric and also allowing for time-dependent field configurations. It was found that a topological inflation scenario shows up for $f \gtrsim 0.23 \cdot \sqrt{2} \cdot \sqrt{8\pi} M_\mathrm{P} \approx 1.63 M_\mathrm{P}$.[7] The spacetime structure of the analogue situation for a global monopole has been analyzed in [155]. We expect this analysis to reflect at least the qualitative features of the string case. The main result is the presence of a horizon

[7] The string spacetime structure has also been studied numerically in [138]. The authors have in particular studied models with $f < M_\mathrm{P}$ and cosmic string solutions with multiple windings. They conclude that in such situations the outside observer can see a full trans-Planckian field cycle under certain conditions.

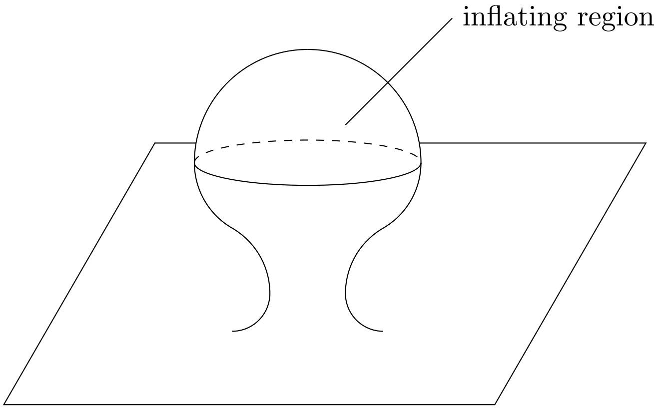


Figure 3.3: Embedding of a 2-dimensional slice of the spacetime for an inflating global monopole in Euclidean 3-space. The upper part of the 'balloon' contains the inflating region.

encompassing the monopole core that is analogous to the horizon in Gregory's string spacetime, i.e. it is not possible to cross the horizon from the exterior while the opposite direction poses no problems. This horizon is the spacelike boundary of the inner inflating region. In fact, any observer that sits within this boundary will be expelled out of this region at some time as the monopole scalar field rolls down its potential and eventually oscillates about its minimum. These outer regions with the oscillating scalar field correspond to a matter dominated spacetime. During the evolution of the monopole the spacetime develops an inflating 'balloon' which is connected by a throat to the exterior spacetime (cf. Fig. 3.3). For an observer it is possible to enter the throat as it contains not only the inflating region but also a matter dominated part.

Let us summarize what we have found so far. For $f \lesssim M_{\mathrm{P}}$ there exist non-singular string spacetimes the metric of which is time-dependent while the field configuration is static. Such solutions are not available for $f \gtrsim M_{\mathrm{P}}$. Instead, topological inflation becomes a possible scenario. However, the field providing the UV completion of the axion then becomes time-dependent.

Naively, one might want the magnetically charged object, i.e. the string, to be static, at least in the sense that the field profile is static. According to this the spacetime found by Gregory is a well behaved UV completion of the string for $f \lesssim M_{\mathrm{P}}$. Whether topological inflation is a viable UV completion of the string is an interesting open question. In any case, from the spacetime structure point of view it seems to be perfectly well behaved and is hence a UV completion of a string with $f \gtrsim M_{\mathrm{P}}$. Note however, that in both topological inflation and Gregory's solution a horizon is present which encompasses the string core and, contrary to the black hole horizon, shields the core from the exterior.

## 3.5 Magnetic string in an Effective Theory with $f > M_{\mathrm{P}}$

In the following we present another candidate UV completion for a magnetic string with $f > M_{\mathrm{P}}$. To be specific, the theory of an axion with $f > M_{\mathrm{P}}$ will be understood as an effective theory, which can be obtained from a more fundamental theory of two (or more) axions with sub-Planckian decay constants. Following [67, 139], we then proceed to construct an effective string with $f > M_{\mathrm{P}}$ out of the strings present in the more fundamental theory and examine it for potential pathologies.

### 3.5.1 Constructing the Effective String with $f > M_{\mathrm{P}}$

We begin by recalling the UV completion of the last section:

$$\mathcal{L} = -\frac{1}{2}|\partial\Phi|^2 - \frac{\lambda}{4}(|\Phi|^2 - f^2)^2 . \tag{3.21}$$

Now consider the sum of two such Lagrangians, where we expand around the vacuum $|\Phi_1| = f_1$, $|\Phi_2| = f_2$. The axionic part can be written as:

$$\mathcal{L} = -\frac{f_1^2}{2}(\partial\varphi_1)^2 - \frac{f_2^2}{2}(\partial\varphi_2)^2 + \cdots . \tag{3.22}$$

For simplicity, we will take $f_1 = f_2 = f$ from now on. Famously, even if $f \ll M_P$, an effective axion with large decay constant can be obtained [92]. The main idea behind [92] is to design a potential on the $\mathrm{T}^2$-field space parametrized by $\{\varphi_1, \varphi_2\}$ which forces the field onto a winding trajectory, e.g.

$$\varphi_1 = N\varphi_2 , \tag{3.23}$$

with $N \gg 1$ (cf. [144] and Fig. 3.4). The desired potential can be viewed as arising from an appropriate combination of instantonic terms.

As suggested in [78], a winding trajectory can alternatively be realized by an appropriate flux choice. This corresponds to making the orthogonal combination of axions massive by gauging à la Dvali [146] (see also [147, 148]). This method can be viewed as a way of avoiding the 0-form WGC in the low energy effective theory, as discussed more explicitly in [67] in the 1-form context. As we will now see in detail, the flux of winding inflation [78] is indeed the axion analogue of the EFT approach of [67] for avoiding the magnetic WGC for vectors. However, we will also see that its success is more ambiguous than in the vector case.

Recall first the more familiar case of gauging of a shift-symmetric scalar by a 1-form[8]

$$\mathcal{L} = -\frac{f^2}{2}|\mathrm{d}\varphi|^2 \qquad \rightarrow \qquad \mathcal{L} = -\frac{f^2}{2}|\mathrm{d}\varphi + A_1|^2 . \tag{3.24}$$

[8] In fact, this has also been used inflationary model building in [156]

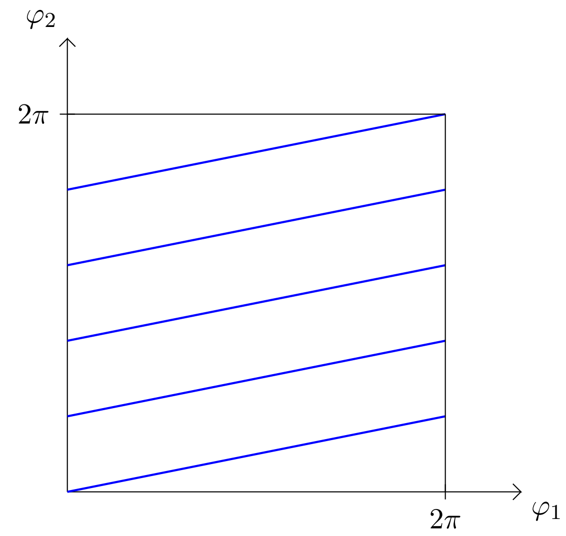


Figure 3.4: Winding effective field space of total length $\sim Nf$ (shown for $N = 5$).

Equivalently, with $F_1 = \mathrm{d}\varphi$, this reads

$$\mathcal{L} = -\frac{f^2}{2}|F_1|^2 \qquad \to \qquad \mathcal{L} = -\frac{f^2}{2}|F_1 + A_1|^2 \,. \tag{3.25}$$

By analogy, we can write down the Lagrangian of a 3-form/(−1)-form theory as

$$\mathcal{L}_{\mathcal{KS}} = -\frac{1}{2g^2}|F_0|^2 \,, \tag{3.26}$$

where $F_0$ is a dimensionless field strength, quantized in units of $2\pi$. The coupling constant $g$ has mass dimension $(-2)$. Then, in analogy to (3.25), we gauge this theory according to

$$\mathcal{L}_{\mathcal{KS}} = -\frac{1}{2g^2}|F_0|^2 \qquad \to \qquad \mathcal{L}_{\mathcal{KS}} = -\frac{1}{2g^2}|F_0 + \varphi|^2 \,. \tag{3.27}$$

In other words, we simply add our 0-form potential to the field strength in the kinetic term of the ungauged model. This is just a simplified version of the idea in [146], which in the context of inflation is also known as the 'Kaloper-Sorbo' approach to axion monodromies [147,148]. Indeed, we simply skipped the detour via the 2-form theory dual to the axion.

In our context, it is interesting to gauge the combination $\varphi_1 - N\varphi_2$, i.e. we consider the model defined by

$$\mathcal{L} = -\frac{f^2}{2}(\partial\varphi_1)^2 - \frac{f^2}{2}(\partial\varphi_2)^2 - \frac{1}{2g^2}|F_0 + \varphi_1 - N\varphi_2|^2 + \cdots \,. \tag{3.28}$$

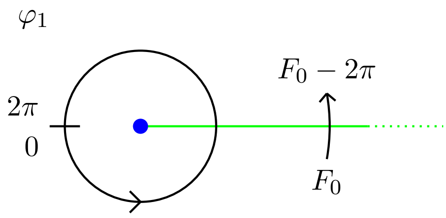


Figure 3.5: The string type 1 is shown with one domain wall attached to it. The jump of $F_0$ across the wall is indicated.

In addition, we include the various charged objects (two types of strings and one type of domain wall) and, if desired, the UV completion of the axions discussed above. For definiteness, we focus on the background with $F_0 = 0$. Clearly, the field space of the effective axion is the submanifold of $\mathrm{T}^2$ specified by the equation $\varphi_2 = \varphi_1/N$ (cf. Fig. 3.4).

Let us make a brief detour to explain what happens at the conceptual level: Originally, we have a field space parameterized as $\{\varphi_1, \varphi_2\}$, with the discrete symmetry group $\mathbb{Z}^2$ (shifts by $2\pi\,\{m, n\}$) being gauged. The $F_0$ theory is unrelated. Then, we associate with any of the $\mathbb{Z}^2$ shifts a particular transformation of $F_0$, in our case

$$F_0 \quad \to \quad F_0 \,-\, 2\pi(m - Nn)\,. \tag{3.29}$$

One could say that we picked a group homomorphism from $\mathbb{Z}^2$ to the proposed $\mathbb{Z}$ symmetry of $F_0$. Now, $|F_0|^2$ is not an invariant Lagrangian any more, but (3.28) provides the appropriately modified, invariant version. This Lagrangian necessarily couples the $F_0$ theory with the $\{\varphi_1, \varphi_2\}$ theory.

To continue, let us introduce the alternative field basis

$$\psi \;=\; N\varphi_1 + \,\varphi_2 \tag{3.30}$$

$$\chi \;=\; \varphi_1 - N\varphi_2\,, \tag{3.31}$$

such that

$$\mathcal{L} = -\frac{f^2}{2(N^2+1)}\left[(\partial\psi)^2 + (\partial\chi)^2\right] - \frac{1}{2g^2}|F_0 + \chi|^2 + \cdots\,. \tag{3.32}$$

We identify $\psi$ as the effective low energy axion which is a periodic field with decay constant $f\sqrt{N^2+1}$. By contrast, $\chi$ is massive and, in addition, closed loops in the field space of $\chi$ are only possible at the expense of passing a domain wall of the $F_0$ theory.

Let us make this last point about domain walls more explicit. For this purpose recall that a string is present whenever the axion follows a closed loop in its field space. Since we have two axions here, $\varphi_1$ and $\varphi_2$, we also have two different species of strings, called string type 1 and 2 from now on. For example, integrating the field strength $d\varphi_1$ along a closed path in space which encompasses a string we must have

$$\oint d\varphi_1 = 2\pi. \tag{3.33}$$

If this loop does not contain an additional string of type 2 the corresponding integral of $\varphi_2$ vanishes. Therefore, we have the general relation

$$\oint d\varphi_i = 2\pi N_i \quad i = 1, 2 \quad , \tag{3.34}$$

where $N_i$ denotes the effective number of strings of type $i$ that lie within the closed integration path. Similarly, a domain wall is defined by the set of points in spacetime at which the value of $F_0$ jumps by an amount that is determined by the unit of quantization of $F_0$, in our case $2\pi$. Combining these facts with the gauging procedure described above we find a relation between strings and domain walls, which we will describe below.

Consider a single string of type 1 embedded in a background of constant $F_0$. Going once around this string in space corresponds to starting with a value $\varphi_1 = 0$ at some point $P$ and finally reaching $\varphi_1 = 2\pi$ when closing the path in space at $P$ again. Since there is no string of type 2 present, $\varphi_2$ takes on the same value when we return to $P$. Continuity demands this to be the same field configuration up to gauge equivalence. In order to identify the final configuration ($\varphi_1 = 2\pi$) with the initial one ($\varphi_1 = 0$) we can perform a gauge transformation $\varphi_1 \to \varphi_1 - 2\pi$. However, according to (3.29), this necessarily implies a shift $F_0 \to F_0 + 2\pi$ which is not the same field configuration we had in the beginning. To make this consistent there has to exist a domain wall attached to the type 1 string such that $F_0 \to F_0 - 2\pi$ when one crosses this wall while circumnavigating the string (cf. Fig. 3.5). Then the above gauge transformation gives us back the initial field configuration. We can repeat the argument for a string of type 2. The result is that a type 2 string must be attached to $N$ domain walls as indicated in Fig. 3.6. This follows straightforwardly from the gauge transformation of $F_0$ (3.29).[9]

We are now in a position to build an effective string out of strings of type 1 and 2 connected by domain walls. This construction is parallel to the procedures in [67,139]. In particular, take one string of type 2 and attach one string type 1 to each free end of the $N$ domain walls that come with the string type 2 (cf. Fig. 3.7). We observe that the orientation of the domain walls automatically fits the shifts in $F_0$ as discussed for the two types of strings separately. The resulting object is a string that is uncharged under the massive field $\chi$ in the sense that $\chi \to \chi$ when going once around it. Similarly we have $\psi \to \psi + 2\pi(N^2 + 1)$. Hence, the effective string just constructed is the string corresponding to the axion $\psi$ of the low energy theory with decay constant $f_{\text{eff}} = f\sqrt{N^2 + 1}$. The massive axion $\chi$ is not visible outside this effective string, i.e. at low energies.

The important point for us is that this object corresponds to a microscopic construction of a string for $f > M_{\text{P}}$. In the following we wish to examine whether

[9] A similar situation with strings ending on monopoles occurs in models of semilocal strings [157].

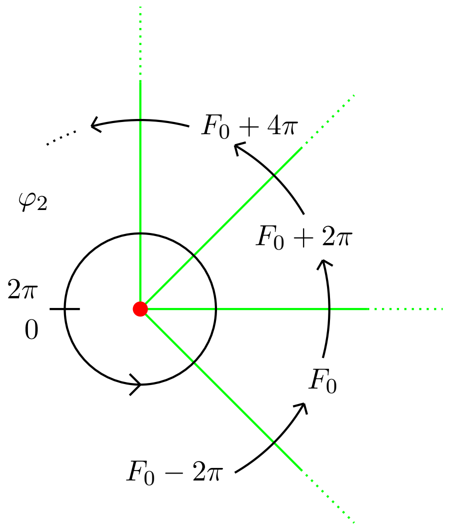


Figure 3.6: The string type 2 is shown with $N$ domain walls attached to it. The jumps of $F_0$ across the walls are indicated.

this effective string can exist as a field-theoretical object. For example, we will only trust this object without gravitational backreaction if the effective string tension is sub-Planckian, i.e. $T_{\mathrm{eff}} \lesssim M_{\mathrm{P}}^2$. At the same time, we have seen in the previous sections that string solutions with $f > M_{\mathrm{P}}$ can at most exist as dynamical solutions giving rise to topological inflation. Hence, we also wish to determine to what extent our effective string is consistent with these previous findings.

### 3.5.2 Estimating the String Tension

Let us therefore estimate the tension of the effective string. All in all, there are three contributions to this:

- The tension of the individual elementary strings contained in the effective string.
- The tension generated by the mutual interaction of the elementary strings.
- The tension due to the domain walls.

Elementary strings have tensions $f^2$ which sum up to a total of $(N+1)f^2 \sim f_{\mathrm{eff}}^2/N$. This can be kept sub-Planckian by choosing $N$ large enough (for fixed $f_{\mathrm{eff}}$) and hence poses no problem for us.

Next consider the mutual interaction of the elementary strings. This contribution is very complicated to determine as it depends on the detailed configuration of the $N + 1$ elementary strings connected by $N$ domain walls. However, one

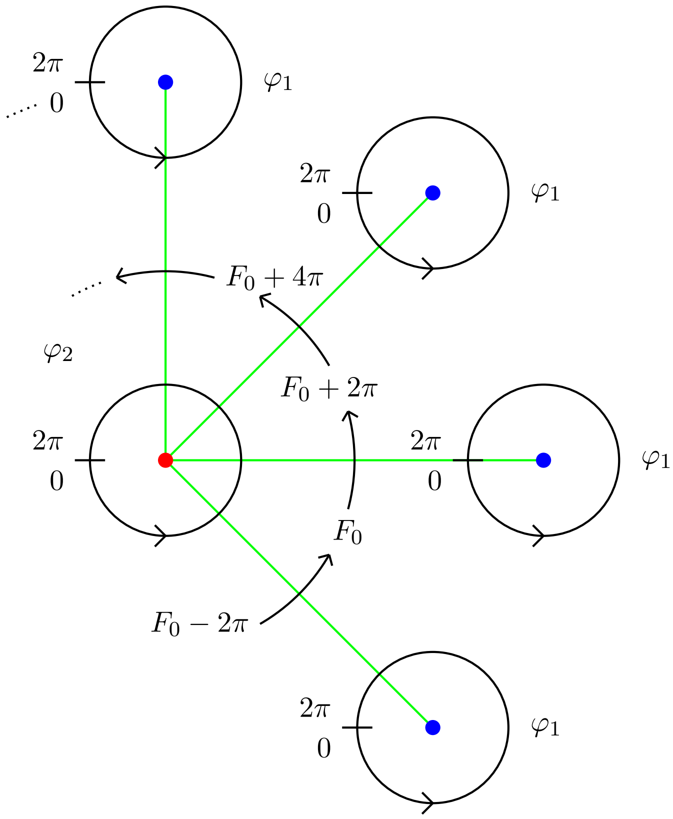


Figure 3.7: The unique combination of the building blocks given by string type 1 and 2 gives rise to an urchin-like structure shown in this figure consisting of one central string of type 2 connected to $N$ strings of type 1 by $N$ domain walls. The orientation of the axions and the jumps of $F_0$ are shown.

can make the following approximation: Let $R$ denote the radius of the effective string. The average density of strings in the cross section of the effective string is given by $\sim N/R^2$. This corresponds to an average distance $\sim R/\sqrt{N}$ between single strings. For large $N$ this is certainly much smaller than the size $R$ of the effective string and it is reasonable to describe the elementary string distribution in terms of a continuous charge density $\varrho$, i.e. the charge per cross section area. In addition, one certainly expects the $N$ strings of type 1 to arrange themselves approximately symmetrically around the single string of type 2. Hence, in the continuum limit, we argue that the charge density depends only on the distance $r$ from the center of the effective string and we can write $\varrho = \varrho(r)$. This density is of course normalized by the charge of the effective string, i.e.

$$\int_0^{2\pi} d\varphi \int_0^R r dr \varrho(r) = N + 1. \tag{3.35}$$

Define

$$Q(r) = \int_0^{2\pi} d\varphi \int_0^r r' dr' \varrho(r') \tag{3.36}$$

to be the string charge contained in a cylinder of radius $r$ centered on the inner type 2 string. The norm of the field strength of such a charge distribution at radius $r$ is given by $Q(r)/r$, while the surrounding charge does not generate a contribution to the field, which is completely analogous to classical Electrodynamics. Exterior to our effective string, i.e. for $r > R$, we have of course $Q(r) = N + 1$. The correct $r$-dependence of $\varrho$ might be calculated by balancing the forces on a elementary string at distance $r$ from the center due to the attached domain wall and the field of all string charge contained in the cylinder of radius $r$. The repulsive force per length between two strings having positive charges $n_1$ and $n_2$ respectively is given by $2\pi n_1 n_2 f^2/r$ with $r$ being the distance between them. The attractive fore per length between a string type 1 and the central string type 2 due to the connecting domain wall is given by the tension $W$ of the domain wall. With this information we can write down the balance equation for the forces,

$$W \sim \frac{Q(r) f^2}{r}, \tag{3.37}$$

which results in $Q(r) \sim Wr/f^2$. The normalization $Q(R) = N + 1$ determines the string radius $R$ as

$$R \sim Nf^2/W. \tag{3.38}$$

Now we can calculate the effective tension due to the mutual interaction of the elementary strings. It is given by integrating the energy density over the cross section of the effective string, i.e.

$$\int_0^{2\pi} d\varphi \int_0^R r dr \frac{f^2}{2} \frac{Q(r)^2}{r^2} \sim N^2 f^2 \sim f_{\mathrm{eff}}^2. \tag{3.39}$$

This shows that the effective tension generically becomes super-Planckian if $f_{\mathrm{eff}}$ is chosen super-Planckian.

Finally, we determine the contribution of the tension of the domain walls. The tension due to one domain wall is $\sim RW$. Collecting the contribution of each wall in the effective string gives a tension $\sim NRW$.

In addition to this inherent domain wall tension, there is another contribution due to the excitation of the massive field when crossing a domain wall. From (3.32) we see that $\chi = -F_0$ in the vacuum. Since a domain wall is accompanied by a jump in $F_0$, it is not possible for $\chi$ to stay at the potential minimum and it is hence excited when crossing a domain wall. This is an additional contribution to the tension of the domain wall and can be estimated as follows. Consider a domain wall at a point far away from any strings that may be attached to it. Then, the translational symmetry tangent to the wall implies that the massive field depends only on the direction orthogonal to the wall, the coordinate of which may be chosen to be $x$. Let the background field be $F_0 = F_-$ for $x < 0$ and $F_0 = F_+$

for $x > 0$ and define $\Delta F_0 = F_+ - F_- = 2\pi$, as appropriate for a single domain wall. Far away from the wall we want the massive field $\chi$ to occupy the vacuum, i.e. we demand $\lim_{x\to-\infty}\chi(x) = -F_-$ and $\lim_{x\to+\infty}\chi(x) = -F_+$. The equations of motion for the Lagrangian (3.32) with these boundary conditions are solved by

$$\chi(x) = \begin{cases} -\pi e^{\alpha x} - F_- & \text{for} \quad x < 0 \\ \pi e^{-\alpha x} - F_+ & \text{for} \quad x \geq 0 \end{cases} , \tag{3.40}$$

where $\alpha = \sqrt{N^2+1}/(gf)$. The typical width of this field profile is given by $\alpha^{-1}$. The field profile is expected to be modified if one approaches the string at the end of the wall to a distance of less than $\alpha^{-1}$. This should be irrelevant to our calculation if the typical wall length $R$ obeys $R \gg \alpha^{-1}$. Since the deviation from the background $F_0$ is always $\mathcal{O}(1)$ the potential energy density of the above field configuration is $\sim g^{-2}$. Integrating this over the field profile with typical width $\alpha^{-1}$ yields a contribution $W_\chi \sim f/(gN)$ to the tension of the domain wall. Thus, the overall contribution to the tension of the effective string reads $\sim (NW + f/g)R \sim (NW + f_{\text{eff}}/(Ng))R$. Using the expression (3.38)[10] for the radius $R$ we find $\sim f_{\text{eff}}^2$ for the total domain wall contribution to the string tension, which is the same as for the contribution from the mutual string interaction.

We are therefore forced to conclude that the existence of strings with super-Planckian charge and sub-Planckian tension is very questionable. Even if we had found that it is possible, we were faced with the problem of the badly behaving spacetime of such a string when including gravity. This was extensively discussed in the previous two sections. There we presented the work of Gregory who argued that abandoning the requirement of a static metric allows for a topological inflation scenario for strings with super-Planckian charge. Hence, Saraswat's method for constructing effective charged objects allowed us to find a possible UV completion of topological inflation in terms of a composite string with super-Planckian charge.

## 3.6 Conclusions

The two standard arguments against super-Planckian axion decay constants, $f \gtrsim M_{\text{P}}$, are the loss of control over instantonic corrections (the electric WGC) and problems with stringy realizations. In this chapter, we have tried to argue for the same conclusion using the magnetic WGC, i.e. the requirement that a dual charged object (a string) exists in the effective theory.

First, we wrote down the general formula for the cutoff resulting from the magnetic WGC and attempted an extrapolation to the naively singular case of axions. This suggests that $f \gtrsim M_{\text{P}}$ should be forbidden.

Second, we examined explicit string solutions with $f \gtrsim M_{\text{P}}$, looking for possible inconsistencies. Static solutions of this type (the CK spacetime of an axionic

[10]Note that now we have to use the full domain wall tension given by $W + W_\chi$ in this formula.

charged string) are singular in both the UV and the infrared. Accepting the infrared problems as a special feature of infinitely extended strings in general relativity, the crucial question is whether the UV problems can be cured by an appropriate UV completion. We argued that smooth, 4d UV completions can not lead to negative deficit angles and showed that this excludes static string spacetimes with $f \gtrsim M_\mathrm{P}$. Super-Planckian axion decay constants would then be ruled out.

We then relaxed the requirement of time-independence in two steps. First, one can allow for a time-dependent metric while the field configuration should still be static, i.e. the string width is fixed. Such solutions, found by Gregory, can be singularity-free in the infrared and have a smooth UV completion. They possess a horizon encompassing the string core. However, they exist only for $f \lesssim M_\mathrm{P}$.

Next, we considered admitting general time-dependent field configurations. In this case, the set-up of topological inflation provides a UV completion of the string, even for $f \gtrsim M_\mathrm{P}$. In fact, topological inflation requires super-Planckian $f$. We argued that the corresponding spacetime has a horizon, similar to that of Gegory's spacetime for $f \lesssim M_\mathrm{P}$. There are two reasons why one might want to reject topological inflation as a string solution in the sense of the magnetic WGC: One is the time-dependence which one could consider unnatural for an object that is supposed to be the analogue of a magnetic monopole. The second is the presence of a horizon (although the latter is very different from the black hole horizon potentially hiding a magnetic monopole). However, consistency would then force us to reject Gregory's solution for $f \lesssim M_\mathrm{P}$ as well, such that no string solution is acceptable at all. In other words: If we demand that string solutions with $f \lesssim M_\mathrm{P}$ exist, we cannot use the horizon as an argument against topological inflation. The interpretation of these observations remains open.

Finally, we tried to apply Saraswat's recent observation [67] that certain low energy effective theories can avoid the collapse of the minimally charged monopole to a black hole, even though they violate the WGC. We considered theories with two sub-Planckian axions and two types of strings. At low energies such models can have one effective super-Planckian axion coupled to one effective string. However, in contrast to the gauge theory case, the tension of this string remains super-Planckian. Although we do not provide an exact solution of this system including gravity, the previously discussed properties of string spacetimes suggest very strongly that time dependence will be an unavoidable feature. But it may be, that this time-dependence is restricted to the metric while corresponding field configurations remain static, as is the case in Gregory's solution. It would be important to understand this in detail.

The decisive question whether the magnetic WGC rules out super-Planckian $f$ depends on the precise definition of what a minimally charged object is allowed to be. In this context, it may be crucial to understand the gravitational dynamics of the composite effective string introduced above.

# 4 Effective 3-form Description of Instantons and Axion Potentials

## 4.1 Introduction

In this chapter, we investigate the effective description of gauge or gravitational instantons via 3-form gauge theories and in particular study the effect of fermions on this effective theory. Furthermore, we discuss the role of fermionic operators in the context of axionic shift symmetries. Such global shift symmetries are expected to be in the swampland and we propose a lower bound on axion masses to quantify how strong these symmetries should be broken.

Gauge theories of 3-form potentials have been discussed since a long time [158]. It is also well-known that the Chern-Simons 3-form of a non-abelian gauge theory transforms under the non-abelian gauge transformation exactly like a fundamental 3-form gauge potential [159]. A similar argument can be made for gravity. Taking this seriously, one can use the 3-form gauge theory as an effective description of Yang-Mills (YM) theory at low energies. This has been discussed in [160] in the context of chiral perturbation theory of QCD and more generally in [161–163].

Our first, simple, technical point in this chapter is to apply this description to the case of a Higgsed YM theory at energies below the symmetry breaking scale. In this regime, the instanton gas is dilute and a quantitatively controlled analysis is possible. As a result, the effective 3-form gauge theory description can be rigorously established as long as the source term $\theta$ (used for probing the theory through the coupling $\theta\,\mathrm{tr}(F\tilde{F})$) is small. We will rely on this controlled model of a dilute instanton gas and its 3-form description in what follows.

An aspect which we are particularly interested in is the 3-form description of YM theories with instantons and massless fermions [161–163]. Especially the idea that this may imply fermion condensates independently of confinement and that small fermion masses may be generated through gravitational instantons are intriguing.

A crucial assumption for this line of reasoning is that the effective 3-form description of YM theory with massless fermions includes a massless pseudoscalar. Indeed, in QCD this is the familiar $\eta'$ meson. This scalar may be dualized into a 2-form which then gauges the effective 3-form and makes it massive [146]. As a result, one obtains a very reasonable effective description for how fermions remove instanton effects. However, as we point out, a different option arises in our case of a Higgsed YM system: Light fermions suppress the gauge coupling of the effective 3-form, completely decoupling it in the massless limit. This interpretation is supported, in our calculable setting, by the fact that no evidence of the massless scalar can be found. Thus, the effective description of how fermions remove instanton effects may change depending on the diluteness of the gas.

We go on to study Higgsed YM theories which are coupled to an axion. Gauge instantons then generically induce an axion-potential $\propto e^{-S}$ where $S$ denotes the instanton action. According to the WGC for axions, $S$ is bounded from above as $S \lesssim M_P/f$ with the axion decay constant $f$ [50]. However, a priori the WGC does not make a claim about the overall size of the axion potential. Thus, the potential could be small because $e^{-S}$ comes with a small prefactor [72, 164–166].

Such a situation arises in the presence of light fermions with mass $m$ because the prefactor scales as $(m/v)^{N_f}$, with $v$ the Higgs scale and $N_f$ the number of flavors. Obviously, for $m \to 0$ the potential vanishes identically and a global symmetry involving a shift in the axion and anomalous U(1) rotations of fermions emerges. This is inconsistent with QG expectations [51–54]. A similar phenomenon can be observed in a model in which the fermions remain heavy at all times. This can be achieved by providing a mass for them via Yukawa couplings in addition to the hard mass terms. Again, once the hard masses are taken to zero, the axion potential vanishes due to the emergence of a global symmetry. This shows that the problematic feature of the theory is not massless fermions but really the presence of a global symmetry. Note also that in this second model one can render the axion massless without changing the infrared (IR) degrees of freedom.

One could restore consistency with QG by simply claiming that flat axions reside in the swampland and thereby exclude such models. However, there is a counterexample in string theory. $\mathcal{N} = 2$ supergravity indeed contains a flat axion and obviously is not in the swampland. This motivates us to conjecture that the dangerous global symmetries described above are ultimately broken by appropriate fermion operators such as explicit mass terms or even higher-dimensional operators.

Ultimately we are interested to find a general lower bound on axion potentials which quantifies to which extent an approximate global axionic shift symmetry is compatible with QG. To do so we try to derive such a bound by applying the WGC to the effective 3-form description of YM theory coupled to an axion. The magnetic version of the WGC for 3-forms bounds the cutoff $\mu$ of the 3-form theory according to $\mu \lesssim (\Lambda^2 M_P)^{1/3}$ where $\Lambda^2$ denotes the 3-form gauge coupling. As long as the effective 3-form theory is valid only for small $\theta$, its cutoff is given by $\Lambda$ and no constraint arises from the WGC.

We argue for the existence of an improved 3-form description which is valid beyond $\theta \ll 1$. This enhanced range of validity allows for a cutoff which can be larger than $\Lambda$ and is given by the mass scale of the lightest massive degrees of freedom in the UV, for example the mass of light fermions or the Higgs sale. In contrast to the original 3-form description, this new 3-form theory is severely constrained by the WGC. For example, it would parametrically constrain the Higgs scale $v$ of Higgsed YM models with gauge coupling $g$ according to $v \lesssim M_P \exp(-1/g^2)$. This would put weakly coupled YM theories which are Higgsed at a high scale into the swampland. We feel that this is too strong a statement to be taken seriously and conclude that the WGC for 3-forms is at best only valid for the canonical quadratic 3-form theory.

Since our WGC-approach failed we propose a bound on the generic axion potential $V(\phi) = -V_0 e^{-S} \cos\phi$ based just on the requirement of simplicity and consistency with examples we have considered so far, in particular $\mathcal{N} = 2$ supergravity, and the WGC for axions. Our proposal is

$$V_0 e^{-S} \gtrsim \mu^4 \exp(-M_P/f)\,, \tag{4.1}$$

where $f$ denotes the axion decay constant and $\mu$ is the cutoff of the low energy theory that describes exclusively the axion. From this one can easily derive a bound on the axion mass $m_\phi$:

$$m_\phi \gtrsim \frac{\mu}{f} \exp\left(-\frac{M_P}{f}\right) \mu\,. \tag{4.2}$$

We have already mentioned that we expect global axionic shift symmetries to be broken by appropriate fermion operators which then contribute to the axion potential. If the corresponding fermions are not massless, we can use our proposed bound on axion potentials to quantify the strength of such fermion operators. We do so for the mass terms of fermions in Higgsed YM theory with and without additional mass contributions from Yukawa couplings. In both cases we obtain for the fermion mass $m$ parametrically

$$m \gtrsim \exp\left(-\frac{M_P}{f}\right) v\,, \tag{4.3}$$

with the caveat that in the case without Yukawa couplings this holds only for more than four fermion flavors.

Finally, we argue for the possibility of gravity-induced fermion interactions via gravitational instantons. To appreciate our argument, recall that the axial U(1) symmetry of gauged massless fermions is anomalously broken according to $\partial_\mu J_5^\mu \propto \mathrm{tr}(F\tilde{F})$, where $J_5^\mu$ is the axial U(1) current. In the presence of gauge instantons, the spacetime integral of the topological density $\mathrm{tr}(F\tilde{F})$ is non-trivial. This implies the existence of an effective fermion interaction that explicitly breaks the axial U(1). This instanton-induced fermion interaction, also called 't Hooft interaction, can be determined explicitly in the dilute gas approximation [167,168].

The same logic can be applied to pure gravity. There we have $\partial_\mu J_5^\mu \propto \mathrm{tr}(R\tilde{R})$ with the right hand side being a gravitational topological term. To the best of our knowledge, K3 is the only compact manifold non-trivially contributing to this term [169, 170]. We describe how a K3 manifold can be glued into flat $\mathbb{R}^4$, such that it can be considered a local fluctuation of it. This is analogous to the localized field strength of a gauge instanton and also implies the gravitational analogues of 't Hooft fermion interactions. Using the dilute gas approximation we naively estimate the strength of these interactions and find that the associated energy scale is, in the case of the SM, larger than $10^{16}$ GeV. We therefore conclude that they are phenomenologically irrelevant.

## 4.2 The Physics of Massless and Massive 3-forms

In this section we collect some results about 3-form gauge theories [158–161] (see Appendix 4.A for details and derivations). In particular, we argue for the equivalence with the dual (−1)-form description. We also show how the force between domain walls can be used to follow the transition between Coulomb and Higgs phase. This quantifies features discussed in [146].

The free theory is defined by the Euclidean action

$$S_{\mathrm{E}}[A_3, \theta] = \int_{M_4} \left( \frac{1}{2\Lambda^4} F_4 \wedge *F_4 - \mathrm{i}\theta F_4 \right) , \tag{4.4}$$

where $F_4 = \mathrm{d}A_3$ is the field strength, $\Lambda^2$ is the gauge coupling, and $M_4$ is the 4d Riemannian manifold on which the theory lives. If $M_4$ is compact, the $F_4$-flux on it is quantized and one can dualize the partition function based on (4.4) in terms of a sum over discrete values of $F_0 = *F_4/\Lambda^2$. Explicitly,

$$Z[\theta] = \int \mathcal{D}A_3 \exp(-S_E[A_3, \theta]) = C \sum_n \exp\left( -\frac{\Lambda^4}{2} \int_{M_4} (\theta + 2\pi n)^2 * 1 \right) , \tag{4.5}$$

where $C$ is a normalization constant. This partition function is invariant under the shift $\theta \to \theta + 2\pi$ and we can hence view $\theta$ as a periodic variable which takes values in the range $[-\pi, \pi)$. Using this we conclude that for constant $\theta$ the term with $n = 0$ corresponds to the lowest energy state. In the limit $M_4 \to \mathbb{R}^4$ this is the only one that contributes:

$$Z[\theta] \propto \exp\left( -\frac{\Lambda^4}{2} \int \theta^2 * 1 \right) . \tag{4.6}$$

We also note that, for compact space and noncompact (Euclidean) time, i.e. for $M_4 = \mathbb{R} \times M_3$, the theory clearly represents a non-trivial quantum mechanical system: The fundamental degree of freedom can be characterized by $\int_{M_3} A_3$. This system corresponds to a quantum particle on a circle. From now on we will, however, take $M_4 = \mathbb{R}^4$ for simplicity.

Let us now introduce Cartesian coordinates $(x, y, z, t)$ and choose the source $\theta$ such that it represents two parallel domain walls localized at $x = a$ and $x = b > a$:

$$\theta(x) = \begin{cases} \theta_1 & \text{for } x \le a \\ \theta_2 & \text{for } a < x < b \\ \theta_1 & \text{for } b \le x \end{cases} . \tag{4.7}$$

They are subject to a force per unit area,

$$f^{(a)} = -f^{(b)} = \frac{\Lambda^4}{2} (\theta_2^2 - \theta_1^2) , \tag{4.8}$$

which is independent of the distance $b-a$. This is expected since one has no propagating degrees of freedom. Instead, the force is due to a constant background field strength $F_4$ which is different between the walls and outside.

Next we consider the coupling of a dynamical scalar $\phi$ with mass $m$ to the 3-form:

$$S_{\mathrm{E}}[A_3,\phi,\theta] = \int_{M_4} \left( \frac{1}{2\Lambda^4} F_4 \wedge *F_4 - \mathrm{i}(\theta+\phi)F_4 + \frac{f^2}{2}\mathrm{d}\phi \wedge *\mathrm{d}\phi + \frac{1}{2}m^2 f^2 \phi^2 * 1 \right) . \tag{4.9}$$

Here $f$ determines the normalization of $\phi$. If $m=0$, one may take the scalar to be periodic such that $f$ becomes its axion decay constant. The action above is then dual to the situation where a 2-form is gauged by a 3-form as discussed in [146].[1] For us, $m$ is a convenient parameter to switch this gauging on and off.[2] Integrating out $A_3$ gives

$$Z[\theta] \propto \int \mathcal{D}\phi \exp\left[ - \int_{M_4} \left( \frac{\Lambda^4}{2}(\phi+\theta)^2 * 1 + \frac{f^2}{2}\mathrm{d}\phi \wedge *\mathrm{d}\phi + \frac{1}{2}m^2 f^2 \phi^2 * 1 \right) \right] \tag{4.10}$$

which, upon carrying out the $\phi$ integration, simplifies to

$$Z[\theta] \propto \exp\left( -\frac{\Lambda^4}{2} \int_{M_4} \theta(x) \frac{\Box - m^2}{\Box - M^2} \theta(x) * 1 \right) \tag{4.11}$$

with $M^2 = m^2 + \Lambda^4/f^2$. Also the force per area is altered:

$$f^{(a)} = -f^{(b)} = \frac{\Lambda^4}{2} \left[ \frac{m^2}{M^2}(\theta_2^2 - \theta_1^2) + \left(1 - \frac{m^2}{M^2}\right)(\theta_2-\theta_1)^2 \mathrm{e}^{-M(b-a)} \right] . \tag{4.12}$$

The additional term exponentially decays with the distance of the two domain walls and indicates the presence of a propagating degree of freedom with mass $M$. The effect of the constant background field strength is also still present, but it is now suppressed by $m^2/M^2$. At $m=0$, the 3-form theory is Higgsed and this long-distance effect disappears.

## 4.3 3-form Gauge Theory as Effective Field Theory of Instantons

In the late 70s it has been noted that YM theory always contains a 3-form $C_3$ that inherits a corresponding gauge transformation, $C_3 \to C_3 + \mathrm{d}\Omega_2$, from the original non-abelian gauge symmetry. It therefore can be considered a proper

[1] This is the gauge-field-theoretic description [147, 148] of axion monodromy inflation [94, 171], recently revived in the context of $F$-term axion monodromy [130, 172, 173].

[2] Note that a non-zero $m$ may even be made consistent with a fundamentally axionic nature of $\phi$: All one needs is to interpret the effect of the gauging by a further 3-form (which has been integrated out) as an effective monodromy or mass parameter.

gauge 3-form [159]. More recently it was argued that this 3-form may provide an alternative description of the Peccei-Quinn solution of the strong CP problem in terms of a 3-form which gauges the axionic shift symmetry [146]. This logic and its implications have been developed further in [161].

In this section we want to analyze the relation between YM theory and 3-form gauge theory more systematically. To do so we focus on the calculable case of a weakly coupled Higgsed YM theory such that we can employ the dilute instanton gas approximation in our computations. We find that the instanton induced correction to the vacuum energy can be effectively described by a pure 3-form gauge theory at small $\theta$-angle and below the Higgs scale. Adding light fermions to the YM theory destroys this correspondence only at subleading order in the exponential of the instanton action, so that the 3-form description remains valid below the fermion masses.

In [161] a similar theory has been considered but with the non-abelian gauge theory unbroken and not necessarily weakly coupled. It was argued that there may be a bosonic bound state of fermions in the limit of vanishing fermion mass. However, in our case we cannot find evidence for such bound states. Instead we observe that 3-form gauge theory is an EFT of a theory that contains only fermions which are subject to a special type of interaction, so-called 't Hooft interactions.

### 4.3.1 Pure Yang-Mills Theory

Let us first consider pure YM theory with Euclidean action

$$S_{\mathrm{E}}[A_1,\theta] = \int \left( \frac{1}{2g^2}\mathrm{tr}(F_2 \wedge *F_2) - \frac{\mathrm{i}\theta}{8\pi^2}\mathrm{tr}(F_2 \wedge F_2) \right) , \tag{4.13}$$

where $A_1$ is the Lie-algebra-valued gauge potential and $F_2 = \mathrm{d}A_1$. As is well known, $\mathrm{tr}(F_2 \wedge F_2)/(8\pi^2)$ is a total derivative, i.e. it can be written as the exterior derivative of a 3-form $C_3$. This is the proper 3-form gauge potential mentioned in the introduction to this section [159]. $g$ denotes the gauge coupling and $\theta$ is again an arbitrary external source. For $\theta = \text{const.}$ we may identify it with the usual $\theta$-parameter of YM theory. In order to be able to deal with this theory computationally we assume the gauge symmetry to be broken spontaneously at a scale $v$ and take the running gauge coupling to be small at this scale: $g(v) \ll 1$. All instantons larger than $v^{-1}$ are then cut off and the dilute instanton gas approximation is valid. In this case one can integrate out the gauge field and obtains the partition function

$$Z[\theta] \propto \exp\left( 2Kv^4 \mathrm{e}^{-S} \int d^4x \cos\theta \right) , \tag{4.14}$$

where $S = 8\pi^2/g(v)^2$ denotes the instanton action and $K \propto S^4$ (see Appendix 4.B for details). For small $\theta$ this reduces to

$$Z[\theta] \propto \exp\left( -Kv^4 \mathrm{e}^{-S} \int d^4x \, \theta^2 \right) , \tag{4.15}$$

which is exactly the same as (4.6) for the pure 3-form gauge theory if we set

$$\Lambda^4 = 2Kv^4\mathrm{e}^{-S}\,. \tag{4.16}$$

By comparing the actions (4.4) and (4.13) we see that $\theta$ generates the same correlation functions for $F_4$ in the 3-form gauge theory as for $\mathrm{tr}(F_2 \wedge F_2)/(8\pi^2)$ in the gauge theory. Also the forces on domain walls will obviously be the same. Hence we have established the pure 3-form gauge theory as an EFT of Higgsed YM theory at energies below the symmetry breaking scale $v$.

### 4.3.2 Yang-Mills Theory with Fermions

#### Comparison at the 1-instanton Level

Next we add fermions with mass $m$ to the Higgsed YM theory. For simplicity we add only one fermion field in the fundamental representation of the gauge group which is captured by the action[3]

$$S_{\mathrm{E}}[A_1,\psi,\theta] = \int \left(\frac{1}{2g^2}\mathrm{tr}(F_2\wedge *F_2) - \frac{\mathrm{i}\theta}{8\pi^2}\mathrm{tr}(F_2\wedge F_2) + \overline{\psi}(\hat{\gamma}_\mu\hat{D}_\mu + m)\psi * 1\right)\,, \tag{4.17}$$

where $\hat{D}_\mu$ is the Euclidean covariant derivative and $\hat{\gamma}_\mu$ are matrices in spinor space satisfying the Euclidean version of the Clifford algebra, $\{\hat{\gamma}_\mu,\hat{\gamma}_\nu\} = 2\delta_{\mu\nu}$.

As in Subsection 4.3.1 we would like to integrate out all dynamical fields to obtain the corresponding effective action describing the vacuum of the theory. If $m \gg v$, we can first integrate out the fermions in an instanton background which will give an effective action for the gauge field that has additional terms suppressed by powers of $v/m$ [174]. Ignoring these small corrections we are left with a pure gauge theory, i.e. for $m \gg v$ we can ignore the fermions at scales below $v$ and the analysis presented in Subsection 4.3.1 applies.

However, for light fermions, i.e. $m \lesssim v$, we have to take them into account properly. To do so, we integrate out the gauge field first and find the effective action for the fermions in a background of a dilute instanton gas. The corresponding calculation has been done by 't Hooft [167, 168] and leads to the following partition function:

$$Z[\theta] = \int \mathcal{D}\overline{\psi}\mathcal{D}\psi \exp\Big[-\int d^4x\,\big(\overline{\psi}(\hat{\gamma}_\mu\partial_\mu + m)\psi + \kappa v\mathrm{e}^{-S}(\overline{\psi}P_{\mathrm{L}}\psi\mathrm{e}^{\mathrm{i}\theta} + \overline{\psi}P_{\mathrm{R}}\psi\mathrm{e}^{-\mathrm{i}\theta})\big)\Big]\,. \tag{4.18}$$

$\kappa$ is some constant and $P_{\mathrm{L/R}}$ is the left- and right-handed projection operator, respectively. Note that the instanton induced 2-fermion interaction corresponds to the well-known 't Hooft determinant for one flavor and is suppressed by the

[3] For the sake of simplicity we have not included the Higgs sector in the action which is, nevertheless, always implicitly assumed to be present.

instanton action via $e^{-S}$. In the process of integrating out the fermions these interactions will give rise to loop corrections to the effective action which are suppressed by powers of $e^{-S}$ and correspond to many-instanton effects. Therefore it is possible to view the effective action as a power series in this suppression factor.

For now let us ignore these loop corrections and calculate the partition function to leading order in $e^{-S}$, i.e. at the 1-instanton level, which can be done exactly [167]. The result is

$$Z[\theta] \propto \exp\left(2K'v^4\frac{m}{v}e^{-S}\int d^4x\cos\theta\right) \tag{4.19}$$

and reduces to

$$Z[\theta] \propto \exp\left(-K'v^4\frac{m}{v}e^{-S}\int d^4x\,\theta^2\right) \tag{4.20}$$

for small $\theta$. Fortunately, this exactly coincides with (4.15) up to a suppression factor $m/v$ and therefore we can once again apply the logic of Subsection 4.3.1 to conclude that, at leading order in $e^{-S}$, Higgsed YM theory with a light fermion is at energies below the fermion mass $m$ effectively described by a 3-form gauge theory (4.4) with

$$\Lambda^4 = 2K'v^4\frac{m}{v}e^{-S}\,. \tag{4.21}$$

For the sake of completeness let us let us also give the corresponding result for $N_{\mathrm{f}}$ fermion flavors with mass $m$ (cf. Appendix 4.B):

$$\Lambda^4 = 2K'v^4\left(\frac{m}{v}\right)^{N_{\mathrm{f}}}e^{-S}\,. \tag{4.22}$$

This shows that in the limit of massless fermions, $m \to 0$, the 3-form gauge coupling $\Lambda^2$ vanishes or, in other words, the 3-form becomes non-dynamical. At the same time, the cutoff of the effective 3-form theory goes to zero of course. This consistently reproduces the fact that the $\theta$-parameter of YM theory becomes unphysical and instantons are suppressed in the presence of massless fermions.

Let us give an intermediate summary and make an observation which we find interesting: We are considering a YM theory that is Higgsed at a scale $v$ and contains light fermions of mass $m$ below $v$. We may assume $m \ll v$, such that an EFT at scale $\mu$ with $m \ll \mu \ll v$ can be defined. In this EFT, the massive gauge bosons have been integrated out such that we are dealing with a purely fermionic theory. In addition to the kinetic and mass term, these fermions are subject to the famous, instanton-induced 't Hooft interaction. Next, we may also integrate out the fermions (at the 1-instanton level) to obtain the EFT relevant at scales below $m$. We argued that this is a massless 3-form gauge theory. The only assumptions were small $\theta$ and that higher-order corrections in $e^{-S}$ do not modify (4.20) significantly. The interesting implication of this is that the low-energy limit of a fermion theory with 't Hooft interactions is provided by a 3-form theory. Note that this 3-form has a priori nothing to do with the 3-form $C_3 \propto \mathrm{tr}(\mathrm{d}A \wedge A + A \wedge A \wedge A)$ present in the original YM theory.

### Multi-instanton Effects

Let us now have a look at the next-to-leading order corrections due to the instanton induced fermion interaction in order to check whether they spoil the validity of the effective 3-form description already below the naive cutoff scale $m$. In particular we calculate the corrections to the force between two domain walls in Appendix 4.B. We find

$$f^{(a)} = -f^{(b)} = \cdots + 2\kappa^2 v^4 \left(\frac{m}{v}\mathrm{e}^{-S}\right)^2 \left(-\int \frac{d^3p}{(2\pi)^3}\frac{1}{\omega_p^3}(\theta_2^2 - \theta_1^2) + \int \frac{d^3p}{(2\pi)^3}\frac{1}{\omega_p m^2}\mathrm{e}^{-2\omega_p(b-a)}(\theta_2-\theta_1)^2\right) + \mathcal{O}\left(\frac{m}{v}\mathrm{e}^{-S}\right)^3 , \quad (4.23)$$

where the dots ... denote the leading order contribution (4.8) with (4.21). The first term in the brackets is logarithmically divergent. It corrects the leading order result and hence contributes to the gauge coupling $\Lambda^2$ of the effective 3-form theory. Physically we expect this to renormalize the fermion mass $m$ such that (4.21) remains true when used with the appropriate renormalized mass. The second term, however, is finite and contains a non-trivial dependence on the distance between the domain walls.

We can rewrite this second term in terms of the modified Bessel function $K_1$ of the second kind:

$$\int \frac{d^3p}{(2\pi)^3}\frac{1}{\omega_p m^2}\mathrm{e}^{-2\omega_p(b-a)} = \frac{1}{2\pi^2}\frac{K_1(2m(b-a))}{2m(b-a)} . \quad (4.24)$$

For very small distances, $m(b-a) \ll 1$, as well as very large distances, $m(b-a) \gg 1$, this can be approximated as

$$\frac{1}{2\pi^2}\frac{K_1(2m(b-a))}{2m(b-a)} = \begin{cases} \frac{1}{2\pi^2}\frac{1}{(2m(b-a))^2} + \mathcal{O}(1) & \text{for } m(b-a) \ll 1 \\ \left[\frac{1}{(4\pi m(b-a))^{3/2}} + \mathcal{O}\left(\frac{1}{(m(b-a))^2}\right)\right]\mathrm{e}^{-2m(b-a)} & \text{for } m(b-a) \gg 1 \end{cases} . \quad (4.25)$$

Note that in the limit $m \to 0$ the force becomes exactly proportional to $1/(b-a)^2$.[4] Let us now consider the regime $m(b-a) \gg 1$ in which we naively expect the 3-form description and the 1-instanton approximation to be valid. Indeed, in this regime (4.25) is exponentially suppressed compared to the leading term (4.8) with (4.21). They become comparable only in the other regime $m(b-a) \ll 1$ at

$$b - a \sim \frac{S^2}{\sqrt{mv\mathrm{e}^{-S}}} .^5 \quad (4.26)$$

[4] In this limit one must not forget about the factor $m^2$ in front of the bracket in (4.23).
[5] Here we used the expression (4.100) for $K'$.

This is self-consistent because in the weak coupling regime $S \gg 1$ and therefore, using (4.26),

$$m(b-a) \sim S^2 \sqrt{\frac{m}{v}\mathrm{e}^{-S}} \ll 1\,. \tag{4.27}$$

Consequently, the 3-form description is always valid up to the fermion mass $m$.

In the presence of an emergent massive bosonic degree of freedom we would expect a purely exponential contribution to the force while a massless bosonic degree of freedom would lead to a constant term. Neither of this is the case for the force in (4.24). Hence, we conclude that there is no sign for the presence of some emergent bosonic degree of freedom and the subleading corrections to the force are really due to the exchange of multiple fermions according to the 't Hooft interaction in (4.18).

There is, however, a possible loophole to our conclusion that no bosonic degree of freedom is present. In the general case of $N_\mathrm{f}$ flavors the 't Hooft interaction is a $2N_\mathrm{f}$-fermion interaction and we therefore have a Nambu-Jona-Lasinio type effective theory for the fermions [175, 176]. For such theories it has been shown that non-perturbatively generated masses for the fermions and bosonic bound states of fermions are present at large enough coupling $g$ [177, 178]. In particular this is what is thought to be happening in QCD and leads to chiral symmetry breaking. However, we have constrained ourselves to the small coupling regime, $g \ll 1$, so that this possibility is not relevant for us.

### 4.3.3 Yang-Mills Theory with Fermions and Yukawa Couplings

So far we have seen that Higgsed YM theory with light fermions, i.e. lighter than the Higgs scale, can effectively be described by a 3-form theory,

$$\mathcal{L} = \frac{1}{2\Lambda^4}(*F_4)^2 + \theta(*F_4)\,, \tag{4.28}$$

at scales below the fermion mass $m$ and with $\Lambda^4 \sim (m/v)^{N_\mathrm{f}} v^4$. This means in particular that this description completely breaks down in the limit $m \to 0$.

Now we consider a Higgsed SU(2) gauge theory with gauge coupling $g$ and which is Higgsed by the vacuum expectation value of a scalar doublet $H$ with $\langle H^\mathrm{T} \rangle = v(0,1)$. Furthermore, we add Weyl fermions in the following representations: two SU(2) doublets $Q_{1,2}$ and four singlets $\chi_i$ with $i \in \{1,2,3,4\}$. We generate masses via Yukawa couplings

$$\mathcal{L}_\mathrm{Y} = y_1 H^\dagger Q_1 \chi_1 + y_2 \epsilon_{\alpha\beta} H^\alpha Q_1^\beta \chi_2 + y_3 H^\dagger Q_2 \chi_3 + y_4 \epsilon_{\alpha\beta} H^\alpha Q_2^\beta \chi_4\,, \tag{4.29}$$

where $\alpha, \beta$ are SU(2) indices. For simplicity we choose the Yukawa coupling constants $y_i = 1$ such that the fermions obtain masses of the order of the Higgs scale $v$. Finally we add an explicit mass term

$$\mathcal{L}_M = M \epsilon_{\alpha\beta} Q_1^\alpha Q_2^\beta\,. \tag{4.30}$$

It is crucial to note that whatever value $M$ has, the fermion masses will always be at least $v$ due to the Yukawa couplings. Consider a U(1) transformation according to which $Q_{1,2} \to \mathrm{e}^{\mathrm{i}\alpha} Q_{1,2}$ and $\chi_i \to \mathrm{e}^{-\mathrm{i}\alpha}\chi_i$. This U(1) is anomalous with respect to SU(2) and explicitly broken by the mass term $\mathcal{L}_M$. Thus the SU(2) $\theta$-parameter is physical and below the Higgs scale the theory is effectively described by a 3-form theory like (4.28). The cutoff of this effective theory is given by $m$.

Let us consider how the effective 3-form description is affected by the parameter $M$. For $M = 0$ the anomalous U(1) can be used to rotate away the SU(2) $\theta$-parameter. It is hence unphysical. However, in the effective 3-form description $\theta$ is still physical which seems to be a contradiction. Furthermore, for all $M > 0$ there is no reason for the 3-form description to break down. In particular its cutoff $v$ is independent of the value of $M$. So how can the two points $M = v$ and $M = 0$ be smoothly connected to each other in the effective 3-form description?

A reasonable and simple answer is that the 3-form decouples in the limit $M \to 0$. By this we mean that $\Lambda^2 = \Lambda^2(M)$ such that $\Lambda^2(0) = 0$. In this way the 3-form is unable to generate a potential for the $\theta$-parameter and makes it effectively unphysical. Hence consistency with the UV theory is restored. We expect that the coupling constant $\Lambda^2$ of the 3-form theory must be proportional to some positive power of $M$. Its role is analogous to that of the fermion masses in our original model without Yukawa couplings. In the following we will assume that this analogy can be taken literally and $\Lambda^2(M)$ is given by (4.21) with $m = M$. Note also that the decoupling of the 3-form is due to the change of a parameter of the UV theory and takes place without changing the degrees of freedom in the IR.

## 4.4 Swampland Constraints on Axions and Fermions

### 4.4.1 Global Symmetries and Fermion Operators

Axions have by definition a perturbative global shift symmetry. If this symmetry were exact also at the non-perturbative level, it would violate the QG censorship of global symmetries. We therefore expect this symmetry to be broken by non-perturbative effects in a consistent theory.

Indeed, this is realized if the axion couples to a (Higgsed) YM theory with $N_\mathrm{f}$ massive gauged Dirac fermions. In the following we will refer to such a model as the light-fermion-scenario in contrast to the heavy-fermion-scenario explained in Subsection 4.3.3. This terminology is supposed to stress the fact that the fermions in the model presented in Subsection 4.3.3 remain always massive due to the Yukawa couplings. The Lagrangian of the light-fermion-scenario reads

$$\mathcal{L} = -\frac{1}{4g^2}\mathrm{tr}(F_{\mu\nu}F^{\mu\nu}) + \sum_{i=1}^{N_\mathrm{f}} \overline{\psi}_i(\mathrm{i}\not{D} - m_i)\psi_i - \frac{1}{2}f^2(\partial\phi)^2 - \frac{\mathrm{i}\phi}{16\pi^2}\mathrm{tr}(F_{\mu\nu}\tilde{F}^{\mu\nu})\,,^6 \quad (4.31)$$

[6] We have suppressed the Higgs sector here.

where $\phi$ denotes the axion and $f$ its decay constant. Now the shift symmetry is non-perturbatively broken by instantons which induce an effective potential for the axion at low energies (below the Higgs scale).

If at least one of the fermions becomes massless the Lagrangian becomes invariant under the transformation $\psi_i \to e^{i\alpha\gamma_5}\psi_i$ and $\phi \to \phi - 2\alpha$, where $\psi_i$ is the massless fermion. Since this symmetry contains a shift in the axion and is exact at the quantum level, we conclude that the axion potential must vanish in the presence of a massless fermion.[7] This can be viewed as a simple symmetry argument for a technical result of the instanton calculus. However, as we will discuss later on, we expect such a theory to be constrained due to the QG censorship of global symmetries [51–54].

Let us now consider the heavy-fermion-scenario of Subsection 4.3.3. It consists of a Higgsed SU(2) YM theory with two Weyl fermion doublets $Q_1, Q_2$ and four Weyl singlets $\chi_1, \chi_2, \chi_3, \chi_4$. These eight Weyl fermions are given a mass via the Higgs mechanism such that we end up with four massive Dirac fermions. Furthermore, we add an explicit mass term $M\epsilon_{\alpha\beta}Q_1^\alpha Q_2^\beta$. If we couple an axion to this theory, instantons will generate an effective axion potential. In the limit $M \to 0$ the theory becomes invariant under an exact global symmetry with the transformation law $Q_{1,2} \to e^{i\alpha}Q_{1,2}$, $\chi_i \to e^{-i\alpha}\chi_i$ and $\phi \to \phi - 2\alpha$. Similarly to the light-fermion-scenario we can conclude from this that the axion potential must vanish in the limit $M \to 0$. The existence of this global symmetry is again in conflict with QG.

At first sight the two examples given above may look very similar. Here we would like to point out an important difference. In both theories the axion potential becomes zero in the limit of a vanishing mass parameter. While in the light-fermion-scenario this results in the presence of a truly massless fermion, in the heavy-fermion-scenario all fermions remain massive in the limit $M \to 0$ due to the Yukawa couplings. In particular, in the former case the axion potential vanishes only at the expense of changing the IR degrees of freedom by introducing massless fermions while in the latter case those degrees of freedom are fixed for all values of $M$.

We have seen that both examples have a global symmetry which eventually allows for the presence of an exactly massless axion. One may argue that those two theories are simply incompatible with QG and hence reside in the swampland. However, a similar situation arises in $\mathcal{N} = 2$ supergravity which contains an axion with an exactly flat potential [179]. Nevertheless, we expect higher fermion interaction terms to break this symmetry explicitly and thereby make the theory consistent with QG again. Similarly, such additional fermion interactions could be used to make our two scenarios consistent with QG, too.

Motivated by this we conjecture that such additional fermion interactions are

[7] Strictly speaking we do not have an exact global symmetry here because of the chiral gravitational anomaly. However, this effect is severely suppressed, as we will discuss in the next section, and can be eliminated by adding an appropriate number of ungauged massless fermions.

mandatory for fermions without a hard mass term in the presence of axions in order to prevent the existence of a global symmetry. This is a non-trivial statement and we find it interesting how the exclusion of a global axionic shift symmetry imposes constraints on fermion interactions. One could ask whether a minimal strength of these interactions can be inferred on general grounds and in the next subsection we attempt to do so by conjecturing a constraint on axion potentials.

### 4.4.2 (Too Strong) a Constraint on Axions from the WGC for 3-forms

Let us now bring back in the 3-form description of our two models. So far we have discussed the effective 3-form description of Higgsed YM theory without axions in Section 4.3. However, according to (4.10) the axion is easily accommodated by replacing the source $\theta$ by the axion field $\phi$ and adding a corresponding kinetic term for it. Then we immediately see that the axion mass $m_\phi^2$ is given by $m_\phi^2 = \Lambda^4/f^2$. That means, whenever a global symmetry of the full UV theory forbids an effective axion potential, the effective 3-form description must decouple as $\Lambda^2 = 0$ (cf. Subsection 4.3.3) is required for a vanishing potential. This is exactly what we observe in the case of a Higgsed YM theory in the limit of massless fermions and what we still expect to happen once Yukawa couplings have been introduced as in the heavy-fermion-scenario. More generally, the 3-form gauge coupling $\Lambda^2$ should always be proportional to a symmetry-breaking parameter of the UV theory.

The idea is now to apply the WGC to the effective 3-form description of YM theory and thereby derive a constraint on the 3-form gauge coupling and hence also on the axion potential. Let us start by stating the two versions of the WGC for 3-forms. The electric WGC for 3-forms requires the existence of domain walls which naturally couple to the 3-form and whose tension $T$ is bounded from above according to

$$T \lesssim \Lambda^2 M_\text{P}\,, \tag{4.32}$$

where $\Lambda^2$ is the 3-form gauge coupling [50]. As long as the cutoff of the 3-form theory is below $(\Lambda^2 M_\text{P})^{1/3}$ this bound has no consequences because the theory breaks down before the presence of such heavy domain walls is required by the WGC. On the other hand, if the cutoff is larger than $(\Lambda^2 M_\text{P})^{1/3}$ and domain walls are not part of the theory, such a theory is forbidden by the WGC. The magnetic version of the WGC for 3-forms simply bounds the cutoff $\mu$ from above according to

$$\mu \lesssim (\Lambda^2 M_\text{P})^{1/3}\,, \tag{4.33}$$

which is exactly the condition for the electric WGC to be satisfied without the presence of a light domain wall.

In order to apply the WGC to our effective 3-form theory of Higgsed YM theories we need to discuss the cutoff of this effective theory in some detail. In Section 4.3 we have found that the effective 3-form description of Higgsed YM theory is naively valid up to the Higgs scale $v$ while the presence of light fermions with masses $m \lesssim v$ reduce this cutoff down to $m$. If these fermions acquire their masses via order one Yukawa couplings, the cutoff stays at $v$. However, we have ignored a caveat in the corresponding argument which we would like to point out now. Recall from Section 4.2 that the energy density of the 3-form theory is given by $\epsilon = \Lambda^4\theta^2/2$. Furthermore, the effective 3-form description is only valid for $\theta \ll 1$ and thus breaks down at energy densities corresponding to $\theta \sim 1$, i.e. $\epsilon \sim \Lambda^4$. Therefore the actual cutoff $\mu$ of the EFT is $\mu \sim \Lambda$. Using this new cutoff we find $(\Lambda^2 M_\mathrm{P})^{1/3} \gtrsim \Lambda \sim \mu$ which means that the theory always perfectly satisfies both the magnetic and electric WGC.

The last paragraph showed that the effective 3-form description of instantons breaks down at a scale set by the 3-form gauge coupling $\Lambda^2$. In particular this cutoff can be much lower than naively expected. Consider for example the heavy-fermion-scenario. In this theory the 3-form gauge coupling $\Lambda$ can be made arbitrarily small by choosing the parameter $M$ appropriately. In particular we can choose it such that we have a cutoff $\mu \sim \Lambda \ll v$, where $v$ is the Higgs scale as usual. On the other hand, there are no new degrees of freedom in the theory below the Higgs scale. Hence, there should be an effective theory that is valid in the energy range between $\mu \sim \Lambda$ and $v$ and contains the same degrees of freedom as the original effective 3-form theory.

In order to get rid of the constraint $\theta \ll 1$ we would like to find a 3-form theory with action $S[F_4, \Lambda^2]$ such that it reproduces the full partition function (4.14), i.e.

$$\int \mathcal{D}A_3 \exp\left(-S[F_4,\Lambda^2] - \mathrm{i}\int \theta F_4\right) \propto \exp\left(\Lambda^4 \int d^4x \cos\theta\right) , \tag{4.34}$$

with $\Lambda^2$ satisfying (4.16) or (4.21), depending on whether we include fermions or not. Although we cannot determine the explicit form of $S[F_4, \Lambda^2]$, the above formula implicitly defines it and thereby also defines the effective 3-form theory we are looking for. In general $S[F_4, \Lambda^2]$ can be a very complicated functional. For example, from the instanton calculus we expect $\exp(-S[F_4, \Lambda^2])$ to have non-trivial support only at $F_4$-configurations which are $\delta$-function-localized at certain points, each contributing one unit to $\int F_4 \in \mathbb{Z}$.

Now we can use this improved effective 3-form theory and apply the WGC to it. According to (4.33) the 3-form gauge coupling must obey $\Lambda^2 \gtrsim (\mu/M_\mathrm{P})\mu^2$. This turns out to be an extremely strong statement. To see this consider for example a simple Higgsed YM theory with Higgs scale given by $v$. As usual the corresponding effective 3-form theory has a gauge coupling $\Lambda^4 = v^4 \mathrm{e}^{-S}$ and cutoff $\mu = v$. (4.33) then implies

$$v \lesssim \exp\left(-\frac{4\pi^2}{g^2}\right) M_\mathrm{P} \tag{4.35}$$

where $g$ is the gauge coupling constant of the YM theory evaluated at the scale $v$ and we have used the relation $S = 8\pi^2/g^2$. This would imply that weakly coupled YM theories can only be spontaneously broken at exponentially low scales. Even though this is a valid result we think that it is too strong as there is naively no good reason why such a scenario should not be realizable in string theory.

Therefore we discard (4.33) and conclude that the application of the WGC to the effective 3-form description of Higgsed YM theory leads to peculiar results. On the other hand, as we have discussed above, the WGC applied to the 3-form theory with the conservative estimate $\mu \sim \Lambda$ for the cutoff $\mu$ is satisfied. From this we conclude that, if the WGC for 3-forms has any regime of validity at all, it can only be applied to canonical 3-form theories with the standard action given by (4.4).

### 4.4.3 A Conjecture on Axion Potentials and Implications for Fermions

Given our failed attempt to use the WGC for 3-forms to constrain axion potentials we now instead try to find a reasonable conjecture for a bound on the axion potential in the following. In general we expect the non-perturbative axion potential to be of the form

$$V(\phi) = -V_0 \mathrm{e}^{-S} \cos\phi + \mathcal{O}\left(\mathrm{e}^{-2S}\right) . \tag{4.36}$$

We would like to find a lower bound on the amplitude of this potential. A first step into this direction is the WGC for axions which constrains the action $S$ according to $S \lesssim M_\mathrm{P}/f$, i.e. $V_0\mathrm{e}^{-S} \gtrsim V_0\mathrm{e}^{-M_\mathrm{P}/f}$ [50]. However, as long as $V_0$ is completely free this does not provide a hard bound on the potential. Very naively one could conjecture that $V_0\mathrm{e}^{-S} \gtrsim M_\mathrm{P}^4\mathrm{e}^{-M_\mathrm{P}/f}$ but this, again, is too strong since the axion potential in $\mathcal{N} = 2$ supergravity vanishes exactly and therefore provides a counterexample in the landscape.

How can we reconcile a vanishing axion potential in SUSY moduli space and, at the same time, a lower bound on it? A possible answer is that the bound depends on the cutoff of the effective axion theory such that it vanishes for zero cutoff. We therefore propose the following bound on axion potentials as a conjecture:

$$V_0\mathrm{e}^{-S} \gtrsim \mu^4 \exp\left(-\frac{M_\mathrm{P}}{f}\right) , \tag{4.37}$$

where $\mu$ is the cutoff of the low energy theory that exclusively contains the axion. The corresponding bound on the axion mass is

$$m_\phi \gtrsim \frac{\mu}{f} \exp\left(-\frac{M_\mathrm{P}}{2f}\right) \mu\, . \tag{4.38}$$

Note that this lower bound obeys $(\mu/f)\mathrm{e}^{-M_\mathrm{P}/(2f)}\mu < (\mu/M_\mathrm{P})\mu$ and is therefore smaller than the cutoff $\mu$ for all $f$. Hence our constraint on the axion mass is always consistent with the effective axion theory itself.

Now let us discuss the implications of this bound for the two scenarios discussed so far. Let us start by considering the light-fermion-scenario with $N_\mathrm{f}$ fermions of mass $m$ and an axion as defined by (4.31). The cutoff of the EFT of the axion is given by $\mu = m$. In order to satisfy (4.37) the fermion mass has to be bounded from below according to

$$m \gtrsim \Theta(N_\mathrm{f}-4)\exp\left(-\frac{1}{N_\mathrm{f}-4}\left[\frac{M_\mathrm{P}}{f}-S\right]\right)v \gtrsim \Theta(N_\mathrm{f}-4)\exp\left(-\frac{M_\mathrm{P}}{f}\right)v \quad (4.39)$$

where we have used (4.22) to determine $V_0$ in (4.37) and $\Theta$ denotes the Heaviside step function. Consequently the fermion mass is only restricted for $N_\mathrm{f} > 4$. For $N_\mathrm{f} \leq 4$ an exactly flat axion potential is possible at the expense of massless fermions. In this case the degrees of freedom in the IR change and there exists no low energy theory that contains only the axion. As already discussed in Subsection 4.4.1, such a theory has a global symmetry consisting of a shift in the axion and an anomalous U(1) rotation of fermions. We therefore expect additional fermion interactions to be present that break this symmetry. Unfortunately we are not able to constrain these fermion operators quantitatively. The constraint (4.38) on the axion mass reads

$$m_\phi \gtrsim \frac{m}{f}\exp\left(-\frac{M_\mathrm{P}}{2f}\right)m\,. \quad (4.40)$$

Next consider the heavy-fermion-scenario. To be more general consider the $N_\mathrm{f}$-fold duplicates version of the model we have discussed so far such that we have $N_\mathrm{f}$ explicit mass terms with mass $M$. With $\mu = v$ (4.37) implies

$$M \gtrsim \exp\left(-\frac{1}{N_\mathrm{f}}\left[\frac{M_\mathrm{P}}{f}-S\right]\right)v \gtrsim \exp\left(-\frac{M_\mathrm{P}}{f}\right)v \quad (4.41)$$

which is very similar but stronger than what we have found in the last paragraph. Finally (4.38) reads in this case

$$m_\phi \gtrsim \frac{v}{f}\exp\left(-\frac{M_\mathrm{P}}{2f}\right)v\,. \quad (4.42)$$

This procedure could also be used to constrain other fermion operator that breaks a global symmetry which protects the axion potential.

## 4.5 Gravitational Instantons and Fermion Interactions

In the last section we have argued that QG may break axionic shift symmetries via certain fermion operators which explicitly break an anomalous chiral U(1) symmetry. Certain types of gravitational instantons may generate such fermion operators as we will discuss in the following.

Consider YM theory with $N_\mathrm{f}$ massless Dirac fermions $\psi_i$, $1 \leq i \leq N_\mathrm{f}$, in the fundamental representation of the gauge symmetry and field strength $F_{\mu\nu}$. This theory has an axial $\mathrm{U}(1)_\mathrm{A}$ symmetry on the classical level which is anomalously broken by instantons at the quantum level [174]. As a result, the corresponding current $J_5^\mu$ of the symmetry is not conserved:

$$\partial_\mu J_5^\mu = -\frac{N_\mathrm{f}}{8\pi^2}\mathrm{tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) \quad (4.43)$$

with $\tilde{F}_{\mu\nu} = (1/2)\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$ and

$$J_5^\mu = \sum_{i=1}^{N} \overline{\psi}_i \gamma^\mu \gamma_5 \psi_i \,. \tag{4.44}$$

By recalling that instantons are topologically non-trivial field configurations, obeying

$$\frac{1}{16\pi^2}\int d^4x \mathrm{tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) = n \in \mathbb{Z}\,, \tag{4.45}$$

we can conclude that

$$\begin{aligned} -2nN_\mathrm{f} = -\frac{N_\mathrm{f}}{8\pi^2}\int d^4x \mathrm{tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) &= \int d^4x \partial_\mu J_5^\mu \\ &= \int_{-\infty}^{+\infty} dt \frac{\partial}{\partial t}\int d^3x J_5^0 \\ &= Q_5(t=+\infty) - Q_5(t=-\infty)\,. \end{aligned} \tag{4.46}$$

This shows that the axial charge $Q_5$ must change by $-2N_\mathrm{f}$ along a single instanton event. Since the $Q_5$ counts the number of right-handed minus the number of left-handed fermions, an instanton must convert $N_\mathrm{f}$ right-handed fermions into $N_\mathrm{f}$ left-handed fermions.[8] Hence we conclude that instantons induce a $2N_\mathrm{f}$-fermion interaction that explicitly breaks the axial $\mathrm{U}(1)_\mathrm{A}$ symmetry. These are the well-known 't Hooft interactions and they can be explicitly calculated. In particular, it turns out that the vectorial $\mathrm{U}(1)_\mathrm{V}$ and the chiral flavor symmetry $\mathrm{SU}(N)_\mathrm{V} \times \mathrm{SU}(N)_\mathrm{A}$ are left unbroken by the interaction. Simply by using this symmetry breaking pattern we can construct the flavor structure of the 't Hooft vertex. To do so we define the $\mathrm{U}(1)_\mathrm{V}$-invariant matrix $\Psi_{ij} = \overline{\psi}_{\mathrm{R}i}\psi_{\mathrm{L}j}$ with $\psi_{\mathrm{L/R}} = P_{\mathrm{L/R}}\psi$. With this quantity we can construct exactly one interaction that is invariant under the chiral flavor symmetry, namely $\det\Psi$. We can not fix the color index structure by this line of reasoning but it will not be relevant for us anyway.

Having discussed the YM case let us now turn to gravity. There we also have a chiral anomaly [180, 181] given by

$$\partial_\mu J_5^\mu = -\frac{N_\mathrm{f}}{192\pi^2} R_{\mu\nu\rho\sigma}\tilde{R}^{\mu\nu\rho\sigma}\,, \tag{4.47}$$

where $R_{\mu\nu\rho\sigma}$ is the Riemann tensor and $\tilde{R}_{\mu\nu\rho\sigma} = (1/2)\epsilon_{\mu\nu\alpha\beta}R^{\alpha\beta}{}_{\rho\sigma}$. Note that now $N_\mathrm{f}$ counts the total number of massless Dirac fermion species. Furthermore, the K3 manifold is the only compact manifold for which the right hand side of (4.47) is non-zero [169, 170], namely

$$\frac{1}{48\pi^2}\int_{\mathrm{K3}} d^4x R_{\mu\nu\rho\sigma}\tilde{R}^{\mu\nu\rho\sigma} = 16\,. \tag{4.48}$$

[8] To conclude this we have to use the fact that the sum of right- and left-handed fermions is conserved.

As in the case of gauge instantons this result depends exclusively on the topology of K3 and can in fact be derived by the Atiyah-Singer index theorem [182]. The topological nature of this relation will be important for us in the following.

Although K3 is not a solution of the Einstein equations, it may nevertheless contribute to the path integral as a quantum fluctuation of spacetime [169, 183]. Because K3 is exactly flat there is hope that such a contribution is not completely suppressed by its action. However, in order to treat K3 as a fluctuation we need to find a way to glue it into flat spacetime, $\mathbb{R}^4$. In fact, this can be done quiet easily by cutting small 4-dimensional balls out of K3 and $\mathbb{R}^4$ and gluing these together via a wormhole-like throat. Sections of this throat are topologically $S^3$. The resulting manifold is of course not a K3 anymore and has changed its topology. Therefore one may be worried whether the topological relation (4.48) still holds for the new manifold. In the following we argue why there is no problem.

Let us start with K3 that is glued onto an $S^4$ instead of $\mathbb{R}^4$ as described above. This manifold is topologically still a K3 and has no boundary. Hence (4.48) remains true. In the next step we split the anomaly integral into three parts, corresponding to the $S^4$, the wormhole and the original K3 contribution:

$$\frac{1}{48\pi^2}\int d^4x R_{\mu\nu\rho\sigma}\tilde{R}^{\mu\nu\rho\sigma} = A_{S^4} + A_{\mathrm{WH}} + A_{\mathrm{K3}}\,. \tag{4.49}$$

But the metrics on $S^4$ and on the wormhole have $R_{\mu\nu\rho\sigma}\tilde{R}^{\mu\nu\rho\sigma} = 0$ locally and hence also $A_{S^4} = 0 = A_{\mathrm{WH}}$. Now delete a point from the $S^4$ which gives $\mathbb{R}^4$ but certainly does not change the integral $A_{S^4}$. Hence we can conclude that a K3 glued into a flat region indeed gives a contribution to the anomaly integral. Finally, this allows us to conclude, similarly to the gauge instanton case, that a K3 fluctuation is, according to (4.47) and (4.48), accompanied by a change of the axial charge $\Delta Q_5 = -4N_{\mathrm{f}}$. Hence K3 must induce an effective $4N_{\mathrm{f}}$-fermion interaction which, again for symmetry reasons, is of the form $(\det\Psi)^2$. Note that this is the highest dimensional fermion operator that is possible in a theory with $N_{\mathrm{f}}$ Dirac fermions.

Now we would like to estimate the amplitude for a K3 fluctuation by evaluating the contributions of K3 to the path integral of Euclidean QG. To do so we need to integrate over all metrics of an asymptotically flat spacetime that contains one K3 fluctuation. The asymptotic $\mathbb{R}^4$ is of course flat and does not contribute to the action. K3 is Ricci flat as well and hence we only have to deal with the wormhole metric. There is a simple way to estimate the action of the wormhole. Let $\rho^{-1}$ be the typical curvature scale of it. Then, by dimensional analysis, its action must be of the order $S_{\mathrm{WH}} \sim \rho^2 M_{\mathrm{P}}^2$.

Next we determine the general form of the pre-exponential factor of the K3 amplitude. Note that the Atiyah-Singer index theorem tells us that the Dirac operator has at least two normalizable zero-modes in a K3 background. In the following we assume that there are exactly two zero modes. Hence, the K3 amplitude vanishes in the presence of massless fermions, perfectly analogous to the case of gauge instantons. However, if the fermions have a small mass $m$, they will contribute a factor $m^{2N_\mathrm{f}}$ to the amplitude. Dimensional analysis then fixes the amplitude for a K3 fluctuation of size $\rho$ to occur in a spacetime volume $d^4x$ as

$$d^4x\, d\rho\, m^{2N_\mathrm{f}} \rho^{2N_\mathrm{f}-5} f(\rho M_\mathrm{P}) \mathrm{e}^{-\rho^2 M_\mathrm{P}^2} . \tag{4.50}$$

Here the function $f$ parametrizes our ignorance about the true $\rho$-dependence of the integration measure. In particular we expect $f$ to be polynomial so that large fluctuations are always exponentially suppressed.

In order to calculate the K3 density we assume $f = 1$ and integrate (4.50) over $\rho$ to obtain

$$n_{\mathrm{K3}} \sim \int_{1/\Lambda}^{\infty} d\rho\, m^{2N_\mathrm{f}} \rho^{2N_\mathrm{f}-5} \mathrm{e}^{-\rho^2 M_\mathrm{P}^2} = \frac{1}{2} M_\mathrm{P}^4 \left(\frac{m}{M_\mathrm{P}}\right)^{2N_\mathrm{f}} \Gamma\left(N_\mathrm{f}-2, \frac{M_\mathrm{P}^2}{\Lambda^2}\right) , \tag{4.51}$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1}\mathrm{e}^{-t} dt$ is the upper incomplete $\Gamma$-function and $\Lambda$ cuts off small K3 fluctuations which we do not trust anymore. We leave the precise value of $\Lambda$ unspecified but it is certainly below the QG scale $M_\mathrm{p}$. Let us ask whether we can treat all K3 fluctuation with size $\gtrsim \Lambda^{-1}$ as a dilute gas, i.e. we want to check the relation $n_{\mathrm{K3}} \ll \rho_\mathrm{c}^{-4}$ where $\rho_\mathrm{c}$ is some characteristic size of K3 fluctuations. Defining $\rho_\mathrm{c}$ to maximize the integrand in (4.51) one obtains $\rho_\mathrm{c}^{-1} \sim M_\mathrm{P}/\sqrt{N_\mathrm{f}}$ for $N_\mathrm{f} \gg 1$. All fluctuations larger than this are exponentially suppressed. Using $\Gamma(s,x) \le \Gamma(s) = (s-1)! \le (s-1)^{s-1}$ for positive integer $s$ we find the following sufficient condition for a dilute gas:

$$\sqrt{N_\mathrm{f}} \frac{m}{M_\mathrm{P}} \ll 1 . \tag{4.52}$$

For the Standard Model with $N_\mathrm{f} < 100$ and $m < 1$TeV this is safely satisfied.

Having established the dilute gas approximation for K3 fluctuations we are now ready to estimate the strength of the induced fermion interaction $(\det \Psi)^2$ by simply repeating the calculation of the effective action of gauge instantons as described in Appendix 4.B. Since in this calculation the fermions are not integrated out no factor of the fermion mass can arise in front of the effective interaction. Therefore, we can use dimensional analysis to find

$$\mathcal{L}_{\mathrm{int}} \sim \int_{1/\Lambda}^{\infty} d\rho\, \rho^{6N_\mathrm{f}-5} \mathrm{e}^{-\rho^2 M_\mathrm{P}^2} (\det \Psi)^2 = \frac{1}{2} M_\mathrm{P}^4 \Gamma\left(3N_\mathrm{f}-2, \frac{M_\mathrm{P}^2}{\Lambda^2}\right) \frac{(\det \Psi)^2}{M_\mathrm{P}^{6N_\mathrm{f}}} .^9 \tag{4.53}$$

The typical scale $\mu$ of this interaction is

$$\mu = \frac{M_\mathrm{P}}{\Gamma(3N_\mathrm{f}-2, M_\mathrm{P}^2/\Lambda^2)^{1/(6N_\mathrm{f}-4)}} . \tag{4.54}$$

For $N_\mathrm{f} \gg 1$ we can estimate $\mu$ by $\mu \gtrsim M_\mathrm{P}/\sqrt{N_\mathrm{f}}$ which gives $\mu \gtrsim 10^{17}$ GeV for the Standard Model. For $\Lambda = M_\mathrm{P}$ we find $\mu \approx 10^{17}$ GeV. Even if this

[9] In principle we should have added an unknown function $g(\rho M_\mathrm{P})$ to the integrand here, similarly to the $f$ in (4.50).

interaction indeed exists, it is extremely suppressed and in foreseeable future of no phenomenological relevance. Furthermore, it would be interesting to determine whether these interactions may contribute to effective potentials of axions. If this the axion couples to gravity via the topological term (4.49), this does not seem to be the case for symmetry reasons. However, a detailed discussion is needed to answer this question properly.

## 4.6 Conclusions

In the first part of this chapter we have studied the effective 3-form description of instantons. To do so we coupled both theories to an external source $\theta$ and calculated the respective partition function and forces on domain walls which we modeled by a spatially varying $\theta$. While this calculation can be done exactly in the 3-form theory, for the gauge instantons one needs to employ the dilute gas approximation. This restricts the range of applicability to weakly coupled Higgsed YM theories. We found that the partition functions and forces agree for small values of the external source $\theta$ and an appropriately chosen 3-form gauge coupling constant $\Lambda^2$. This shows that 3-form theories indeed are EFTs of Higgsed YM theories for small $\theta$. We expect this correspondence to hold also for gravitational instantons.

With the same method we analyzed the effect of gauged fermions on the effective 3-form description. It turned out that they simply alter the expression for the gauge coupling $\Lambda^2$ of the effective 3-form theory by a factor proportional to their mass. This implies that massless fermions decouple the effective 3-form theory which is consistent with the fact that they completely suppress isolated gauge instantons. Recently, it has been argued that the effective 3-form description of YM theory with massless fermions could potentially contain a massless bosonic degree of freedom [161,163]. While this is the case in a confining theory like QCD with an $\eta'$ pseudoscalar, we are not able to find evidence for this in the case of a Higgsed YM theory.

A careful analysis of the cutoff $\mu$ of the effective 3-form theory revealed that the restriction to $\theta \ll 1$ implies it to be given by the 3-form gauge coupling $\Lambda^2$ according to $\mu \sim \Lambda$. In fact this can be much lower than the naive cutoff which, for example, is given by the lowest fermion mass or the Higgs scale. Nevertheless, we argued that there must exist an effective 3-form theory which is valid beyond this low cutoff $\Lambda$ and in particular does not break down at $\theta \sim 1$. Unfortunately, we have not been able to explicitly determine this improved 3-form theory but it would be important to do so in order to improve our understanding of the effective 3-form description of instantons and extend its range of validity.

After having discussed the effective 3-form description of instantons, we considered axionic shift symmetries in the second part of this chapter. Ultimately, we expect such global symmetries to be broken due to quantum gravitational effects. Intuitively, this should manifest itself in terms of a non-vanishing potential and mass for the axion. It is an interesting question whether there is a quantitative bound on how small axion masses can be.

First of all, it is easy to come up with a seemingly innocuous example for a vanishing axion potential. We just have to couple an axion to Higgsed YM theory with massless fermions as discussed in the first part of this chapter. In this theory the axion is exactly massless. The same is true if the fermions obtain a mass via Yukawa couplings but lack a hard mass term. In both cases we identified a global symmetry, that involves a shift in the axion and anomalous U(1) rotations of the fermions, as the reason for the vanishing axion potential. This motivated us to conjecture that such theories should contain certain possibly higher-dimensional fermion operators which break this symmetry and generate a mass for the axion. In this case a lower bound on axion masses could be used to constrain such operators.

However, we found that a strict censorship of massless axions can not be true because $\mathcal{N} = 2$ supergravity is a counterexample in the string landscape. A less restrictive constraint on axions which is consistent with this observation is the following: Let $\tilde{m}_\phi(\mu)$ be the lower bound on the axion mass as a function of the cutoff $\mu$ of the low energy effective theory of the axion. In particular we require this effective theory to exclusively contain the axion. Hence, if there are any other massless degrees of freedom in the theory, the cutoff $\mu$ is zero. Then, in order to restore consistency with $\mathcal{N} = 2$ supergravity, we require $\tilde{m}_\phi(\mu = 0) = 0$. It remains to determine the actual form $\tilde{m}_\phi(\mu)$.

In a first attempt to do so we applied the WGC to the effective 3-form description of Higgsed YM theory with an axion and fermions. Unfortunately, this approach turned out to be unsuccessful. Using the standard 3-form theory, as discussed in the first part of this chapter, we are not able to derive any constraint since the theory trivially satisfies the WGC. The reason for this is the low cutoff of the theory which is given by the 3-form coupling constant $\Lambda$. This restriction can be dropped if we assume that a modified 3-form theory exists whose cutoff is not $\Lambda$ but given by the naive cutoff scale which, in our examples, are light fermion masses or the Higgs scale. However, using this theory we found that the resulting constraint on axion masses and the associated YM theory are too strong to be taken seriously.

Based on simplicity, the examples we have considered so far, and the WGC for axions we finally proposed the following bound on axion masses $m_\phi$:

$$m_\phi \gtrsim \frac{\mu}{f} \exp\left(-\frac{M_{\mathrm{P}}}{2f}\right) \mu \,. \tag{4.55}$$

Here $f$ is the axion decay constant and $\mu$ the cutoff of the low energy axion theory. Since in YM theory instantons and fermions determine the non-perturbatively generated axion mass, we were able to use this bound to obtain a constraint on fermion masses $m$ which parametrically takes the form

$$m \gtrsim \exp\left(-\frac{M_{\mathrm{P}}}{f}\right) v \,, \tag{4.56}$$

where $v$ denotes the Higgs scale of the YM theory.

There are at least two promising directions to make progress with this conjecture in the future. First of all, it is important to test it in a variety of stringy constructions. In particular, it would be interesting to understand how exactly the axion in $\mathcal{N}=2$ supergravity remains massless and determine which role SUSY plays in this context. Second, one can take our bound for granted and explore possible phenomenological implications. Especially constraints on certain fermion operators which break axionic shift symmetries could be studied in different models.

Finally, in the last section we briefly discussed the possibility of fermion operators which are generated by gravitational instantons. These operators are the analogues of the so-called 't Hooft interactions in YM theory. We argued that K3 instantons seem to be the only gravitational instantons capable of inducing those interactions. A rough estimate of the strength of these operators showed that they are severely suppressed and not relevant for phenomenology. It would be interesting to see whether K3 instantons with fluxes have a similar effect and how axions are affected by such fermion operators.

## 4.A 3-form Gauge Theory

### 4.A.1 Pure 3-form Gauge Theory

The free theory of a 3-form gauge potential $A_3$ is defined by the Euclidean action

$$S_{\mathrm{E}}[A_3,\theta] = \int_{M_4} \frac{1}{2\Lambda^4} F_4 \wedge *F_4 - \mathrm{i} \int_{M_4} \theta F_4 \,, \tag{4.57}$$

where $F_4 = \mathrm{d}A_3$ is the field strength associated to $A_3$ and $\theta$ is an external source. $\Lambda^2$ corresponds to the coupling constant and $M_4$ is the 4-dimensional Riemannian manifold on which the gauge theory lives. In the following we take $M_4 = S^1 \times M_3$ with $M_3$ having finite volume and no boundary unless otherwise stated. The corresponding (thermal) partition function is

$$Z[\theta] = \int \mathcal{D}A_3 \mathrm{e}^{-S_{\mathrm{E}}[A_3,\theta]} \,. \tag{4.58}$$

We have normalized $F_4$ such that the Dirac quantization condition reads $\int_{M_4} F_4 = n \in \mathbb{Z}$. Making use of this quantization condition we can rewrite the partition function as

$$Z[\theta] = \int \mathcal{D}(*F_4) \sum_n \delta\left(\int_{M_4} F_4 - n\right) \mathrm{e}^{-S_{\mathrm{E}}[F_4,\theta]} \,, \tag{4.59}$$

where now we view $F_4$ as an independent integration variable and treat the action as a functional of $F_4$. After rewriting $\delta(\int_{M_4} F_4 - n) = \int d\chi/(2\pi) \exp(\mathrm{i}\chi(\int_{M_4} F_4 - n))$, performing the $F_4$-integral and using the identity $\sum_n \exp(\mathrm{i}\chi n) = 2\pi \sum_n \delta(\chi-$

$2\pi n$) we find for the partition function

$$Z[\theta] = C \sum_n \exp\left(-\frac{\Lambda^4}{2}\int_{M_4} (\theta + 2\pi n)^2 * 1\right) , \tag{4.60}$$

with $C$ being a possibly infinite constant.

In order to understand this theory physically let us consider the partition function $Z[\theta]$ in the limit of constant $\theta$. Then

$$Z[\theta] = C \sum_n \exp\left(-\frac{\Lambda^4 \beta V}{2}(\theta + 2\pi n)^2\right) , \tag{4.61}$$

where $\beta$ is the circumference of $S^1$ and $V$ denotes the volume of $M_3$. This is the partition function of a theory with infinitely many orthogonal energy eigenstates labeled by all integers $n$ and with energy given by $E_n = \Lambda^4 V/2(\theta + 2\pi n)^2$. From the form of $Z[\theta]$ it is clear that the theory is invariant under the shift $\theta \to \theta + 2\pi$. Hence it is sufficient to consider only $\theta \in [-\pi, \pi)$.[10] For this choice the vacuum energy is given by $E_0$. For $V \to \infty$ only the vacuum state remains while all other energy eigenstates disappear due to their exponential suppression relative to the vacuum. In the following we will only keep the vacuum state as we are primarily interested in the limit $V \to \infty$.

The partition function for infinite volume and arbitrary $\theta(x)$ reads

$$Z[\theta] = C \exp\left(-\frac{\Lambda^4}{2}\int_{M_4} \theta^2 * 1\right) . \tag{4.62}$$

From this we easily read off the energy density

$$\epsilon = \frac{\Lambda^4}{2}\theta^2 \tag{4.63}$$

and calculate the vacuum expectation value of $*F_4$:

$$\langle *F_4 \rangle = \frac{1}{\mathrm{i}Z[\theta]}\frac{\delta Z[\theta]}{\delta\theta} = \mathrm{i}\Lambda^4\theta \, . \tag{4.64}$$

We also find for the correlator of $*F_4$

$$\langle *F_4(x) * F_4(0) \rangle = \Lambda^4 \delta(x) \, . \tag{4.65}$$

The appearance of the $\delta$-function in the correlator and the fact that the vacuum expectation value exactly follows the external source shows that the field strength is a purely local object that does not propagate any degree of freedom through spacetime.

[10] Note that for $\theta = -\pi$ there are two degenerate energy eigenstates. We will ignore this subtlety in the following.

So far we have seen that the 3-form theory in the large volume limit $V \to \infty$ has only one energy eigenstate, the vacuum, and lacks any propagating degrees of freedom. Therefore, one may be tempted to conclude that the theory does not contain any dynamics. This is not true as we will explain now. As is well known, the gauge potential $A_3$ naturally couples to the worldsheet (WS) of a domain wall via $\int_{\text{WS}} A_3$. Alternatively we may write this as $\int_{M_4} A_3 \wedge J_1$ where $J_1$ is the conserved current of the domain wall. After integration by parts we see that the source term $\int_{M_4} \theta F_4$ is exactly of this form with $J_1 = \mathrm{d}\theta$. In the following we will choose an appropriate $\theta(x)$ that describes two domain walls and calculate the force that acts on them. It turns out that this force is not zero and therefore the theory is not trivial.

For simplicity we will choose $M_3$ to be $\mathbb{R}^3$ with coordinates $(x, y, z)$. We would like to describe two domain walls defined by $x = a$ and $x = b$. Assuming $a < b$ this is realized by the choice

$$\theta(x) = \begin{cases} \theta_1 & \text{for } x \leq a \\ \theta_2 & \text{for } a < x < b \\ \theta_1 & \text{for } b \leq x \end{cases} \tag{4.66}$$

with $\theta_1, \theta_2 \in [-\pi, \pi)$. The force on the domain wall is simply the negative derivative of the energy associated to this configuration with respect to the position of the domain wall. However, this force is going to be infinite due to the domain walls being infinitely extended. Hence, the proper quantity to determine is the force per area. To do so we consider a cylinder with base area $A$ and place it such that its base is parallel to the domain walls and it is centered at $x = a$, i.e. at the first domain wall. Now we calculate the change $\Delta E$ in the energy residing in the cylinder due to a small change $\Delta a > 0$ in the position of the first domain wall. Using (4.63) we find $\Delta E = A \Delta a (\Lambda^2/2)(\theta_1^2 - \theta_2^2)$. Since this expression is linear in $A$ and $\Delta a$ we can immediately read off the force per area acting on the domain walls at $a$ and $b$, respectively,

$$f^{(a)} = -f^{(b)} = \frac{\Lambda^4}{2}(\theta_2^2 - \theta_1^2)\,. \tag{4.67}$$

Note that the forces do not depend on the distance between the domain walls. In fact, they do not depend on spatial coordinates at all and are constant throughout space. This shows that there is always a force acting on them, even if we would push one domain wall out to infinity. Therefore the force is not due to the interaction between the walls but only due to the interaction of the walls with the field strength $F_4$. The situation is somewhat analogous to an electric charge in a constant electric field although this analogy is not perfect as the background field strength $\langle * F_4 \rangle$ is not spatially constant.

### 4.A.2 3-form Gauge Theory Coupled to a Scalar Field

Now we extend the 3-form theory by introducing a scalar field $\phi$ with mass $m$ that couples to $F_4$ according to the new action

$$S_{\mathrm{E}}[A_3,\phi,\theta] = \int_{M_4} \frac{1}{2\Lambda^4} F_4 \wedge *F_4 + \int_{M_4} \frac{f^2}{2} \mathrm{d}\phi \wedge *\mathrm{d}\phi + \int_{M_4} \frac{1}{2} m^2 f^2 \phi^2 * 1 - \mathrm{i} \int_{M_4} (\theta + \phi) F_4 \,, \quad (4.68)$$

where $f$ determines the normalization of $\phi$. This theory is special for $m = 0$ since in that case it is dual to a 2-form theory that is gauged by $A_3$. The dual action reads

$$\tilde{S}_{\mathrm{E}}[A_3, B_2] = \int_{M_4} \frac{1}{2\Lambda^4} F_4 \wedge *F_4 + \frac{1}{2f^2} \int_{M_4} (\mathrm{d}B_2 - A_3) \wedge *(\mathrm{d}B_2 - A_3) \quad (4.69)$$

which can be easily checked by dualization under the path integral. This action is invariant under the simultaneous transformations $B_2 \to B_2 + \Omega_2$ and $A_3 \to A_3 + \mathrm{d}\Omega_2$ for an arbitrary 2-form $\Omega_2$. It realizes the Stückelberg mechanism for $A_3$, i.e. the gauge symmetry is spontaneously broken by the vacuum $B_2 = \text{const.}$ such that only a massive $A_3$ is left. In the following we want to argue that on the $\phi$-side of the duality this symmetry breaking can be possibly understood as an effect of the (quantum) dynamics of the massless $\phi$.

Let us make this argument for a more familiar example. Consider the following action:

$$\tilde{S}_{\mathrm{E}}[A_1, \varphi] = \int \frac{1}{2e^2} F_2 \wedge *F_2 + \int \frac{v^2}{2} (\mathrm{d}\varphi - A_1) \wedge *(\mathrm{d}\varphi - A_1) \,. \quad (4.70)$$

This action realizes the Stückelberg mechanism for a 1-form gauge potential $A_1$. If we embedded this theory in a Higgs theory, $v$ would be the vacuum expectation value of the Higgs field. Hence we expect the gauge symmetry to be restored in the vacuum for $v = 0$ which indeed is the case as is clear by inspection of the action.

Next let us have a look at the dual action which reads

$$S_{\mathrm{E}}[A_1, B_2] = \int \frac{1}{2e^2} F_2 \wedge *F_2 + \int \frac{1}{2v^2} \mathrm{d}B_2 \wedge *\mathrm{d}B_2 - \mathrm{i} \int B_2 \wedge F_2 \,. \quad (4.71)$$

Since $B_2$ does not transform under the gauge symmetry of $A_1$ its classical vacuum configuration $B_2 = \text{const.}$ does not break it. Let us inspect the case $v = 0$ for which the spontaneous symmetry breaking is turned off. In this case the dynamics of the field $B_2$ is frozen and it effectively acts as a source for $F_2$. This observation suggests that the dynamics of $B_2$ is ultimately responsible for the spontaneous symmetry breaking. Note also that the duality of (4.70) and (4.71) breaks down when $B_2$ is massive. Hence this property of $B_2$ seems to be crucial for the dynamics behind the spontaneous symmetry breaking. All of these observations carry over to the theories defined by (4.68) and (4.69).

Let us go back to the generic case with arbitrary $m$ and calculate the partition function of the theory. The $A_3$-integration can be carried out as before which leads to

$$Z[\theta] = C \int \mathcal{D}\phi \exp\left( - \int_{M_4} \frac{\Lambda^4}{2} (\phi + \theta)^2 * 1 \right. \\ \left. - \int_{M_4} \frac{f^2}{2} \mathrm{d}\phi \wedge * \mathrm{d}\phi - \int_{M_4} \frac{1}{2} m^2 f^2 \phi^2 * 1 \right) . \quad (4.72)$$

This path integral is Gaussian in $\phi$ and we can therefore simply use the classical equation of motion,

$$\Box \phi = M^2 \phi + (M^2 - m^2)\theta \quad (4.73)$$

with $M^2 = m^2 + \Lambda^4/f^2$, to find the formal result

$$Z[\theta] = C' \exp\left( -\frac{\Lambda^4}{2} \int_{M_4} \theta(x) \frac{\Box - m^2}{\Box - M^2} \theta(x) * 1 \right) . \quad (4.74)$$

For $f \to \infty$, i.e. $M \to m$, this reduces, up to constant factors, to (4.62) as it should be. The vacuum expectation value and correlator of $*F_4$ are now calculated to be

$$\langle *F_4 \rangle = \mathrm{i}\Lambda^4 \frac{\Box - m^2}{\Box - M^2} \theta \quad (4.75)$$

and

$$\langle *F_4(x) * F_4(0) \rangle = \Lambda^4 \frac{\Box - m^2}{\Box - M^2} \delta(x) . \quad (4.76)$$

From the pole structure of the correlator we infer the presence of a massive degree of freedom with mass $M$ which continues to exist even for $m = 0$. In fact this is not surprising as we have seen that for $m = 0$ a Stückelberg mechanism is at work in the dual description.

Instead of using the formal expression (4.74) to determine the force on domain walls we explicitly use a solution to the equation of motion with $\theta$ as defined in (4.66). This solution can be written as

$$\phi(x) = \left(1 - \frac{m^2}{M^2}\right) \times \begin{cases} \frac{\theta_1 - \theta_2}{2} (\mathrm{e}^{M(x-a)} - \mathrm{e}^{M(x-b)}) - \theta_1 & \text{for } x \leq a \\ -\frac{\theta_1 - \theta_2}{2} (\mathrm{e}^{M(x-b)} + \mathrm{e}^{-M(x-a)}) - \theta_2 & \text{for } a < x < b \\ \frac{\theta_1 - \theta_2}{2} (\mathrm{e}^{-M(x-b)} - \mathrm{e}^{-M(x-a)}) - \theta_1 & \text{for } b \leq x \end{cases} . \quad (4.77)$$

Matching the solutions in the different regimes to each other at the boundary and demanding $\phi$ to be constant at $x \to \pm\infty$ fixes all six integration constants uniquely. Upon using the equation of motion (4.73) the action in (4.74) can be rewritten as

$$\int_{M_4} \frac{\Lambda^4}{2} \theta(x)(\phi(x) + \theta(x)) * 1 . \quad (4.78)$$

The integrand of this action is the energy density in the presence of the two domain walls. In order to appreciate its structure it is helpful to explicitly calculate it:

$$\epsilon(x) = \frac{\Lambda^4}{2} \times \begin{cases} \frac{m^2}{M^2}\theta_1^2 + \left(1 - \frac{m^2}{M^2}\right)\theta_1 \frac{\theta_1-\theta_2}{2}(\mathrm{e}^{M(x-a)} - \mathrm{e}^{M(x-b)}) & \text{for } x \leq a \\ \frac{m^2}{M^2}\theta_2^2 - \left(1 - \frac{m^2}{M^2}\right)\theta_2 \frac{\theta_1-\theta_2}{2}(\mathrm{e}^{M(x-b)} + \mathrm{e}^{-M(x-a)}) & \text{for } a < x < b \\ \frac{m^2}{M^2}\theta_1^2 + \left(1 - \frac{m^2}{M^2}\right)\theta_1 \frac{\theta_1-\theta_2}{2}(\mathrm{e}^{-M(x-b)} - \mathrm{e}^{-M(x-a)}) & \text{for } b \leq x \end{cases} . \tag{4.79}$$

We clearly see that the first term equals the energy density (4.63) of the pure 3-form theory corrected by a factor $m^2/M^2$. The effect of this part of the energy density on the force per area is hence exactly as we have calculated in (4.67) but with the additional factor $m^2/M^2$. Now consider the second term in (4.79). We would like to repeat the computation of the change in energy within a given cylinder as we have done in Subsection 4.A.1. However, this time the energy density changes at arbitrarily large distances from the domain walls if we move them around. Hence, we have to use an infinitely extended cylinder. The total energy within such a cylinder with base area $A$, ignoring the first term in (4.79) we have discussed already, is

$$E = A\frac{\Lambda^4}{2}\left(1 - \frac{m^2}{M^2}\right)\frac{1}{M}(\theta_1 - \theta_2)^2(1 - \mathrm{e}^{-M(b-a)})\,. \tag{4.80}$$

Taking the negative derivative with respect to $a$, dividing by $A$ and combing with the contribution from the first term in (4.79) gives for the total force density

$$f^{(a)} = -f^{(b)} = \frac{\Lambda^4}{2}\left[\frac{m^2}{M^2}(\theta_2^2 - \theta_1^2) + \left(1 - \frac{m^2}{M^2}\right)(\theta_2 - \theta_1)^2\mathrm{e}^{-M(b-a)}\right] \tag{4.81}$$

Let us compare this result with (4.67). We have again a constant contribution in (4.12) which is suppressed by the factor $(m/M)^2$ compared to (4.67) and a new second term that exponentially falls off with the distance between the domain walls. Note that this exponential fall-off is exactly what we could have anticipated from the presence of a massive degree of freedom with mass $M$. While the first contribution to the force is due to the interaction of the domain walls with the background field strength, the second exponential term represents an interaction between the two domain walls due to a massive scalar field. In the limit $f \to \infty$, i.e. in the decoupling limit of $\phi$, (4.81) reduces to (4.67) as it should be. For $m = 0$ the constant part of the force disappears while the second essentially remains unaffected. This can be intuitively understood by observing from (4.75) that the background field strength vanishes for $m = 0$ and constant $\theta$. Hence there is no field strength the domain walls can interact with anymore and the corresponding force becomes zero. On the other hand, as already explained above, even though $m = 0$ there is a massive scalar present which is why the second contribution to the force remains.

## 4.B Instantons in Yang-Mills Theory

In this appendix we collect some well known results about gauge instantons. A good reference is for example [174]. An SU($N$) gauge theory is described by the Euclidean action

$$S_{\mathrm{E}}[A_1,\theta] = \int \frac{1}{2g^2}\mathrm{tr}(F_2\wedge *F_2) - \frac{\mathrm{i}\theta}{8\pi^2}\mathrm{tr}(F_2\wedge F_2)\,, \tag{4.82}$$

where $A_1$ is the Lie-algebra-valued gauge potential and $F_2 = \mathrm{d}A_1$. $g$ denotes the gauge coupling and $\theta$ can in principle be an external source that depends on space. Here we assume the topology of space to be simply $\mathbb{R}^4$. An instanton corresponds to a topologically non-trivial field configuration $A_1^{\mathrm{I}}$ which minimizes the action and has the properties

$$S_{\mathrm{E}}[A_1^{\mathrm{I}},\theta] = \frac{8\pi^2}{g^2} - \mathrm{i}\theta\,. \tag{4.83}$$

The instanton configuration $A_1^{\mathrm{I}}$ has $4N$ moduli, four for the instanton location, one for its size and the rest for the orientation in group space. The contribution of an instanton with a given size $\rho$ and location $x$ to the partition function reads

$$\frac{d^4x d\rho}{\rho^5}C(N)f(g(\rho),N)\mathrm{e}^{\mathrm{i}\theta}\,, \tag{4.84}$$

where

$$f(g(\rho),N) = \left(\frac{8\pi^2}{g^2(\rho)}\right)^{2N}\exp\left(-\frac{8\pi^2}{g^2(\rho)}\right) \tag{4.85}$$

and

$$\frac{8\pi^2}{g^2(\rho)} = \frac{8\pi^2}{g^2(\rho_0)} - \frac{11}{3}N\ln\left(\frac{\rho}{\rho_0}\right) + \mathcal{O}\left[g^2(\rho_0)\ln\left(\frac{\rho}{\rho_0}\right)\right] \tag{4.86}$$

takes into account the running of the coupling with the instanton size $\rho$. $\rho_0$ is an arbitrary reference scale. Furthermore, we have

$$C(N) = \frac{C_1}{(N-1)!(N-2)!}\mathrm{e}^{-C_2N} \tag{4.87}$$

with $C_1$ and $C_2$ being order one numerical constants. Besides the instanton there is also an anti-instanton configuration $A_1^{\mathrm{A}}$ with

$$S_{\mathrm{E}}[A_1^{\mathrm{A}},\theta] = \frac{8\pi^2}{g^2} + \mathrm{i}\theta \tag{4.88}$$

and the corresponding contribution to the partition function is

$$\frac{d^4x d\rho}{\rho^5}C(N)f(g(\rho),N)\mathrm{e}^{-\mathrm{i}\theta}\,. \tag{4.89}$$

In the following we will use the abbreviation $S = 8\pi^2/g^2(\rho_0)$.

Next we would like to determine the full contribution of instantons to the partition function. This can be done in the dilute gas approximation in which all instantons are considered point-like. Such an approximation is only valid if the density of instantons in space is small compared to their maximal size, i.e. if there is no overlap between them. However, in principle we have to integrate (4.84) over all $\rho$ and hence take into account instantons of all sizes. In fact the contribution of large instantons, which are problematic for the dilute gas approximation, diverges. Indeed, inserting (4.86) into (4.84) reveals that the integrand of the $\rho$-integration is given by $\rho^{11N/3-5}$. The exponent is positive for any $N \geq 2$ which renders the integral IR divergent. Hence, the dilute gas approximation is not applicable in a pure non-abelian gauge theory.

Fortunately, this problem can be avoided by introducing a scalar field that breaks the gauge symmetry spontaneously with its vacuum expectation value $v$ and gives a mass to the gauge field. In this case the contribution to the partition function involves an additional factor $\sim \mathrm{e}^{-(\rho v)^2}$ so that large instantons are exponentially suppressed and the $\rho$-integral becomes finite. Now, performing the $\rho$-integration in (4.84) with the exponential suppression factor, dividing by $d^4x$, and ignoring the phase $\mathrm{e}^{\mathrm{i}\theta}$ gives for the instanton density at leading order

$$n = Kv^4\mathrm{e}^{-S}\,, \tag{4.90}$$

where we have chosen $\rho_0 = 1/v$ and

$$K \sim C(N) S^{2N} \Gamma\left(\frac{11}{6}N - 2\right) \tag{4.91}$$

The size of the instantons is now effectively cut off at $\rho \sim 1/v$ and therefore, as long as $K\mathrm{e}^{-S} \ll 1$[11], the dilute gas approximation is valid. In particular, the cutoff of the resulting effective theory is given by $v$ as we will treat everything (in particular instantons) smaller than $1/v$ as point-like.

Now we are in a position to sum the contribution of all possible ways to place instantons and anti-instantons in spacetime and find

$$\begin{aligned} Z[\theta] &= \sum_{n,\overline{n}=0}^{\infty} \frac{1}{n!\overline{n}!} \prod_{k=1}^{n} \left( \int d^4x_k K v^4 e^{-S} e^{\mathrm{i}\theta} \right) \prod_{\overline{k}=1}^{\overline{n}} \left( \int d^4x_{\overline{k}} K v^4 e^{-S} e^{-\mathrm{i}\theta} \right) \\ &= \exp(2K\mathrm{e}^{-S} \int d^4x v^4 \cos\theta))\,. \end{aligned} \tag{4.92}$$

In the next step we consider YM theory with $N_\mathrm{f}$ Dirac fermions of mass $m$ in the fundamental representation of $\mathrm{SU}(N)$. The corresponding action is[12]

$$S_\mathrm{E}[A_1, \psi, \theta] = \int \frac{1}{2g^2} \mathrm{tr}(F_2 \wedge * F_2) - \frac{\mathrm{i}\theta}{8\pi^2} \mathrm{tr}(F_2 \wedge F_2) + \sum_{i=1}^{N_\mathrm{f}} \overline{\psi}_i (\hat{\gamma}_\mu \hat{D}_\mu + m) \psi_i * 1\,, \tag{4.93}$$

[11] This can always be achieved by choosing the gauge coupling small at the symmetry breaking scale.

[12] For the sake of simplicity we have not included the Higgs sector in the action which is, nevertheless, always implicitly assumed to be present.

where $\hat{D}_\mu$ is the Euclidean covariant derivative and $\hat{\gamma}_\mu$ are matrices in spinor space satisfying the Euclidean version of the Clifford algebra, $\{\hat{\gamma}_\mu, \hat{\gamma}_\nu\} = 2\delta_{\mu\nu}$.

Once again we would like to obtain an effective action that is valid below the scale $v$. If $m \gg v$, we can first integrate out the fermions in an instanton background which will give an effective action for the gauge field that has additional terms suppressed by powers of $v/m$. Ignoring these small corrections we are left with a pure gauge theory, i.e. for $v \ll m$ we can simply ignore the fermions at scales below $v$ and the analysis presented at the beginning of this section applies.

For fermions which are light, i.e. $m \lesssim v$, we can no longer simply integrate them out but have to take them into account properly. For later use let us define the operator-valued matrix $\Psi_{ij}(x) = \overline{\psi}_i(x) P_{\mathrm{L}} \psi_j(x)$ where $P_{\mathrm{L,R}}$ denotes the left-handed and right-handed chirality projector, respectively. We would like to integrate out the gauge field and find the effective action for the fermions in a background of a dilute instanton gas. The corresponding calculation has been done by 't Hooft [167] (see also [168]). In particular he showed that the partition function corresponding to the action (4.93) in an instanton background gives rise to the same fermion propagators as the partition function

$$\int \mathcal{D}\psi \mathcal{D}\overline{\psi} \exp\left( - \int d^4x \sum_{i=1}^{N_{\mathrm{f}}} \overline{\psi}_i (\hat{\gamma}_\mu \partial_\mu + m) \psi_i \right) \\ \times \frac{d^4z d\rho}{\rho^5} C'(N, N_{\mathrm{f}}) f(g(\rho), N) \mathrm{e}^{\mathrm{i}\theta} \rho^{3N_{\mathrm{f}}} \det\Psi \,, \quad (4.94)$$

where $z$ and $\rho$ denote position and size of the instanton, $C'$ depends on $N$ and $N_{\mathrm{f}}$ and, most importantly, the running coupling now includes a contribution due to the fermions:

$$\frac{8\pi^2}{g^2(\rho)} = \frac{8\pi^2}{g^2(\rho_0)} + \left( \frac{2}{3} N_{\mathrm{f}} - \frac{11}{3} N \right) \ln\left( \frac{\rho}{\rho_0} \right) + \mathcal{O}\left[ g^2(\rho_0) \ln\left( \frac{\rho}{\rho_0} \right) \right] . \quad (4.95)$$

Note that the color-structure of the operator $\det\Psi$ is in general non-trivial but has been suppressed by our simplified notation. In the following the details of this will not be relevant for us but they can be found for example in [167, 184].

The contribution of an anti-instanton at the same position and of the same size is the complex conjugate of (4.94). Now we can perform the $\rho$-integration in (4.94) and sum the instanton and anti-instanton contribution as in (4.92) to get the following partition function for the fermions

$$Z[\theta] = \int \mathcal{D}\overline{\psi} \mathcal{D}\psi \exp\left( - \int d^4x \left[ \sum_{i=1}^{N_{\mathrm{f}}} \overline{\psi}_i (\hat{\gamma}_\mu \partial_\mu + m) \psi_i \right. \right. \quad (4.96) \\ \left. \left. + \kappa v^{4-3N_{\mathrm{f}}} \mathrm{e}^{-S} (\det\Psi \mathrm{e}^{\mathrm{i}\theta} + (\det\Psi)^\dagger \mathrm{e}^{-\mathrm{i}\theta}) \right] \right) ,$$

with

$$\kappa \sim C'(N, N_{\mathrm{f}}) S^{2N} \Gamma\left( \frac{11}{6} N + \frac{7}{6} N_{\mathrm{f}} - 2 \right) . \quad (4.97)$$

From this we see explicitly that integrating out the gauge fields yields effective fermion interactions, also called 't Hooft interactions, which in the case of one flavor reduce to a simple mass term that explicitly reads

$$\mathcal{L}_{\text{mass}} = \kappa v \mathrm{e}^{-S}(\overline{\psi} P_{\mathrm{L}} \psi \mathrm{e}^{\mathrm{i}\theta} + \overline{\psi} P_{\mathrm{R}} \psi \mathrm{e}^{-\mathrm{i}\theta}) \,. \tag{4.98}$$

Next we would like to determine the vacuum expectation value of $\langle \mathrm{tr}(F \wedge F) \rangle$. To do so we need to integrate out the fermions in (4.96) to obtain an explicit expression for the partition function $Z$ as a functional of $\theta$. The result can be organized in a series expansion in the small quantity $\mathrm{e}^{-S}$ where terms of order $(\mathrm{e}^{-S})^n$ correspond to $n$-instanton contributions. If one is only interested in the leading term of this expansion, one can skip the derivation of the effective action (4.96) and instead directly integrate out gauge fields and fermions in (4.93) in one step.

This calculation has also been done by 't Hooft [167] and the result differs from that without fermions only by an additional factor $(\rho m)^{N_{\mathrm{f}}}$ in the instanton contribution (4.84) and the proper running coupling as stated in (4.95). Furthermore, the constant $C(N)$ in (4.84) is changed by a factor $\sim \mathrm{e}^{N_{\mathrm{f}}}$ and hence also depends on $N_{\mathrm{f}}$ now. Note that the negative contribution $-(2/3)N_{\mathrm{f}}$ from the running coupling to the exponent of $\rho$ is over-compensated by the factor $(\rho m)^{N_{\mathrm{f}}}$ so that the $\rho$-integration remains UV finite for all values of $N_{\mathrm{f}}$.

After repeating the familiar steps of summing all instanton contributions we find

$$Z[\theta] = \exp\left(2VK'\mathrm{e}^{-S}\left(\frac{m}{v}\right)^{N_{\mathrm{f}}} v^4 \cos\theta\right) \tag{4.99}$$

with

$$K' \sim C(N, N_{\mathrm{f}}) S^{2N} \Gamma\left(\frac{11}{6}N + \frac{1}{6}N_{\mathrm{f}} - 2\right) . \tag{4.100}$$

Remember that the exponent of this formula is only exact up to order $\mathrm{e}^{-S}$ as we have ignored multi-instanton contributions. Compared to the theory without or with heavy fermions, (4.92), we essentially get a suppression factor $(m/v)^{N_{\mathrm{f}}}$. With this formula it is easy to calculate

$$\langle \mathrm{tr}(F_2 \wedge F_2) \rangle = \frac{1}{\mathrm{i}Z[\theta]} \frac{1}{V} \frac{\partial Z[\theta]}{\partial \theta} = 16\pi^2 \mathrm{i} K' \mathrm{e}^{-S} v^4 \left(\frac{m}{v}\right)^{N_{\mathrm{f}}} \theta \tag{4.101}$$

for small $\theta$ and formally infinite volume $V$ of the 4-dimensional Euclidean space. We observe that the vacuum expectation value vanishes for $m = 0$. This is of course the well-known result that massless fermions screen the topological susceptibility of non-abelian gauge theories. Ultimately, this is due to the fact that massless fermions render $\theta$ unphysical.

Having discussed massless fermions in some detail, let us go back to the effective theory of fermions with cutoff $v$ and partition function (4.96). Similarly to the discussion of the 3-form gauge theory in Appendix 4.A we now want to introduce domain walls via the external source $\theta$ and calculate the forces acting on them.

We have seen that instantons induce $2N_\mathrm{f}$-fermion interactions. This implies that fermions can be exchanged between two distinct instantons and hence induce interactions between them. This effect should contribute to the force between domain walls. In order to understand this better we can calculate the vacuum energy $E_0$ in the presence of two domain walls, as defined by (4.66), in perturbation theory. For simplicity we consider the case $N_\mathrm{f} = 1$ for which the instanton-induced fermion interaction is just a correction to the mass term. We organize our calculation as an expansion in two small parameters. First we assume $\theta$ to be small and keep only terms up to quadratic order in the interaction term in (4.96) which gives the following interaction Lagrangian

$$\mathcal{L}_\mathrm{int} = \tilde{v}\overline{\psi}\left(1 - \mathrm{i}\gamma_5\theta - \frac{1}{2}\theta^2\right)\psi\,, \tag{4.102}$$

where we have introduced $\tilde{v} = \kappa v \mathrm{e}^{-S}$. Our final result, the force on the domain walls, will be given up to quadratic order in $\theta$ as well. Second, there is a factor $\mathrm{e}^{-S}$ in front of the instanton induced interaction term in (4.102) which is a measure for the interaction strength. This is a small quantity and therefore we use it as our second expansion parameter. Recall that each instanton comes with this factor and hence we can view terms of order $(\mathrm{e}^{-S})^n$ as an $n$-instanton effect. We will include contributions up to order $(\mathrm{e}^{-S})^2$.

There are two vacuum diagrams that contribute to the ground energy at second order in our expansion parameter $\mathrm{e}^{-S}$. The sum of these is

$$\begin{aligned} E_0 = &-2\tilde{v}L^3\left(2\int\frac{d^4p}{(2\pi)^4}\frac{m}{p^2+m^2} + \tilde{v}\int\frac{d^3p}{(2\pi)^3}\frac{p^2}{\omega_p^3}\right) \\ &+ 2\tilde{v}L^2\left(\int\frac{d^4p}{(2\pi)^4}\frac{m}{p^2+m^2} - \tilde{v}\int\frac{d^3p}{(2\pi)^3}\frac{m^2}{\omega_p^3}\right)\left((L-(b-a))\theta_1^2 + (b-a)\theta_2^2\right) \\ &- \tilde{v}^2L^2\int\frac{d^3p}{(2\pi)^3}\frac{1}{\omega_p^2}\left(\mathrm{e}^{-2\omega_p(b-a)} - 1\right)(\theta_2-\theta_1)^2\,. \end{aligned} \tag{4.103}$$

Here, as usual, $\omega_p = \sqrt{m^2+p^2}$ and $L^3$ is the formally infinite spatial volume. Except for the integral in the last line all momentum integrals are divergent but let us ignore this issue for a moment and proceed to calculate the force acting on the domain walls. Differentiating (4.103) with respect to $a$, dividing by $L^2$ and multiplying by $-1$ gives the force density acting on the domain wall at $a$:

$$\begin{aligned} f^{(a)} = 2\tilde{v}&\left(\int\frac{d^4p}{(2\pi)^4}\frac{m}{p^2+m^2} - \tilde{v}\int\frac{d^3p}{(2\pi)^3}\frac{m^2}{\omega_p^3}\right)(\theta_2^2-\theta_1^2) \\ &+ 2\tilde{v}^2\int\frac{d^3p}{(2\pi)^3}\frac{1}{\omega_p}\mathrm{e}^{-2\omega_p(b-a)}(\theta_2-\theta_1)^2\,. \end{aligned} \tag{4.104}$$

The first line of this expression still contains divergent integrals. We expect these to contribute to the renormalization of the fermion mass $m$. In (4.99) we already gave the result for the partition function with its exponent exact to linear order in $e^{-S}$. Hence, the term linear in $e^{-S}$, i.e. linear in $\tilde{v}$, in (4.104) must coincide with the force calculated from (4.99) for consistency. The second term in the first line of (4.104) provides a higher order correction to this. However, we also find a completely new contribution to the force at the 2-instanton level proportional to $(\theta_1 - \theta_2)^2$. This new term is exponentially suppressed in the distance of the two domain walls. For $m = 0$ the exponential suppression vanishes and the momentum integral can be performed exactly: The result is a Coulomb force law, i.e. it is proportional to $(b - a)^{-2}$.

# 5 Summary and Outlook

In this thesis we have studied QG constraints on axions as imposed by two so-called swampland constraints: the SDC and the WGC. Naively, they forbid super-Planckian field ranges and give certain non-trivial constraints on axion potentials. Since these conclusions, and in particular the SDC and WGC themselves, lack a rigorous proof and given their potential phenomenological importance, we tried to shed some light on SDC and WGC for axions by analyzing them via different approaches and in different setups such as string theory or EFT. Some of the central questions we considered while doing so are:

(1) Is it possible to construct explicit counterexamples to the SDC and WGC for axions in string theory?

(2) Can one obtain natural inflation from string theory?

(3) What is, if any, the correct formulation of the magnetic WGC for axions?

(4) What kind of pathologies does the violation of the magnetic WGC for axions imply for the theory at hand?

(5) How does QG constrain axion potentials?

In the following we summarize our main results.

We start with Question (1) which is equivalent to whether it is possible to have super-Planckian axion field ranges in string theory. Indeed, in Chapter 2 we were able to construct mildly super-Planckian axion field ranges in the moduli space of a very simple toroidal compactification of type IIB string theory. The essential ingredient of this model is a flux-induced reduction of the toroidal symmetry SL(2,$\mathbb{Z}$) which enhances the axion periodicity to super-Planckian values by a factor of the flux number. A tadpole cancellation condition bounds this enhancement from above.

Although we have succeeded in constructing a super-Planckian axionic direction, the true, geodesic distance between points on this trajectory grows only logarithmically with the flux number. This is due to the fact that the axionic direction is not a geodesic from the point of view of the full moduli space. Motivated by this observation we analyzed the model from a 4d low energy perspective and find that the corresponding EFT breaks down once one tries to traverse large field distances. This breakdown is due to a KK tower of states which becomes light at large field distances and is very similar to what the SDC requires.

The model described above obviously has two important drawbacks. On the one hand the achieved axionic field range is only mildly super-Planckian so that one can not definitely exclude that parametrically large field ranges are still unfeasible in string theory. Although it is plausible that more complicated compactifications may evade the bound on the axionic field range imposed by the tadpole cancellation condition, an explicit proof of this is missing. On the other hand, we have not tried to build a realistic model of natural inflation which would include moduli stabilization in general and in particular the generation of a potential that keeps the axion on its long trajectory and provides a small mass for it so that slow-roll inflation can take place.

In Chapter 3 we analyzed the magnetic WGC for axions in some detail from a low energy perspective. A naive extrapolation of the original magnetic WGC for charged particles suggests that the corresponding axionic version strictly forbids super-Planckian axion decay constants. An alternative formulation forbids the existence of black strings which are the analogues of black holes. Now the question is whether these two formulations are really equivalent and, even more importantly, what the correct formulation is.

To answer these questions we studied well-known solutions of strings which are charged under the axion field. Since strings have co-dimension two in four spacetime dimensions we certainly expect strong gravitational backreaction. Indeed, such strings have either physical singularities at a finite distance from the string core or give rise to a non-static inflating spacetime. While sub-Planckian axion decay constants admit at least a static field configuration this is no longer true for super-Planckian decay constants. Instead such strings give rise to topological inflation. A black string in the sense of the presence of an event horizon similar to that of a black hole does not seem to exist.

In the absence of a black string it is not clear what the magnetic WGC for axions should look like. The central question is whether topological inflation poses any fundamental problem for the theory and should hence be forbidden. In this case the magnetic WGC for axions would forbid super-Planckian decay constants as naively expected. Alternatively, accepting topological inflation as a proper UV completion of axionic strings would lead to inconsistencies with the WGC for charged particles. We therefore tend to prefer the WGC to censor topological inflation and hence super-Planckian axion decay constants. However, a decisive conclusion remains elusive.

By considering composite strings we tried to avoid topological inflation for super-Planckian decay constants but a very naive and approximate analysis suggests that this does not work. A quantitative understanding of the dynamics and the gravitational backreaction of such composite strings would help to make this conclusion more rigorous.

In Chapter 4 we discussed the axion potential in Yang-Mills theory and pointed out the well-known fact that massless fermions will make this potential exactly flat. This can be traced back to the emergence of a global symmetry involving a shift in the axion and an anomalous U(1) rotation of the fermions. A similar situation is encountered when the fermions remain massive via Yukawa couplings but are lacking a hard mass term. From a quantum gravitational point of view such theories with a global symmetry are presumably in the swampland.

We conjectured that these global symmetries are generically broken by higher-dimensional fermion operators which in particular break the anomalous U(1) transformation explicitly. Furthermore, since axion potentials can only be non-vanishing if the axionic shift symmetry is broken, we proposed a lower bound on axion potentials in order to provide a more quantitative measure for how strongly the shift symmetry is expected to be broken by quantum gravitational effects. Note, however, that our bound does not exclude vanishing axion potentials but they are possible in the presence of other massless degrees of freedom as is realized for example in $\mathcal{N} = 2$ supergravity. In such cases we expect the axionic shift symmetry to be broken by the additional massless degree of freedom. We also showed that such a bound on axion potentials can be used to infer a corresponding bound on hard fermion masses. More generally, this procedure should provide constraints on any fermion operator that breaks the shift symmetry of an axion.

In addition we considered K3 manifolds as gravitational instantons and argued that they generate fermion interactions. These interactions are the analogues of the well-known 't Hooft interactions which are induced by SU($N$) gauge instantons and turn out to be extremely feeble as they are suppressed by almost the Planck scale.

Besides the discussion of axionic shift symmetries and fermion operators we examined the effective 3-form description of instantons and in particular studied the effect of fermions on this description. We found that fermions without a hard mass term decouple the effective 3-form, i.e. the 3-form gauge coupling vanishes in such a scenario.

After having summarized the results of this thesis let us finally put them in relation to the general status of the research on swampland constraints. First, recall that a large portion of the recently increased interest in swampland constraints such as the WGC or SDC was due to the possibility to apply them to models of large field inflation such as natural inflation. Since models of large field inflation predict a large and hence potentially observable tensor-to-scalar ratio this was particularly interesting and exciting: If the swampland conjectures are indeed correct, large field models would have been the first phenomenologically relevant theories that are directly constrained by string theory.

Unfortunately, by now the upper bound on the tensor-to-scalar ratio has become so strong that large field inflation is already in quite some tension with experimental data. Consequently, much of the original motivation to study swampland constraints in general and constraints on axions and large field ranges in particular is lost. It is, nevertheless, interesting and worthwhile to further explore the swampland paradigm in order to gain a better understanding of string theory and quantum gravity, even though its phenomenological relevance may have decreased.

One of the most urgent and seemingly most difficult problems in the research on swampland constraints is to provide rigorous proofs for the swampland conjectures. There are several attempts to provide such a proof for the WGC but there are contradictory results. All of them are based on certain assumptions and are valid only in special settings [55–61]. It is extremely important to put the conjectures on firm ground as soon as possible. As long as this task has not been accomplished, all results obtained from the conjectures can not be fully trusted and it is questionable how useful it is to pursue more and more consequences of so far very speculative conjectures.

In this thesis we have seen that, based on examples from string theory and EFT, it is difficult to obtain a clear picture of swampland constraints, in particular of those on axions. On the one hand, we have found evidence that super-Planckian axion decay constants may be problematic in EFTs and verified a SDC-like property of an explicit moduli space. These findings support the WGC for axions and the SDC. On the other hand, we were also able to construct an axion with super-Planckian field range which is in tension with the WGC for axions. The status of these conjectures is therefore very inconclusive and calls for a more rigorous treatment. Furthermore, even if all or at least some of the conjectures are true, it is not clear whether they necessarily give constraints in the phenomenologically relevant IR of a theory. These doubts are raised by the observation that a UV theory that obeys the WGC for charged particles can give rise to an IR theory that violates it [67].

Nevertheless, we would like to point out a few promising directions for future research. First of all one can try to explicitly construct super-Planckian axions in different stringy settings and improve the constructions we have obtained in Chapter 2. Indeed, recently there has been substantial progress in this direction. In [49] an axion with parametrically large field range has been realized very explicitly in the EFT of a type IIB compactification with fluxes using the idea of a winding trajectory [78, 92] in the field space of two or more axions. The problem of moduli stabilization has been addressed therein and a potential that keeps the axion on its long trajectory could be constructed. This strongly suggests that parametrically large axion field ranges are indeed possible in string theory. However, realizing inflation in this scenario is more demanding than just having a large axion field range but the authors express some hope that inflation may be realizable in a small parameter range.

Furthermore, in Chapter 3, we identified composite strings as a possible loophole to topological inflation and the associated problems with super-Planckian axion decay constants. Estimates of the tension of such strings suggest that they will lead so topological inflation but a more detailed analysis of the dynamics of composite strings could be performed in future work in order to make this discussion more reliable. The fact that there seems to be evidence for axions with clearly super-Planckian field range in string theory (cf. Chapter 2) is of course in tension with the preliminary conclusions of Chapter 3 that super-Planckian axion decay constants are potentially problematic in EFT. It would be interesting to see how the stringy examples with super-Planckian axion decay constants evade the problems with axionic strings.

We have stressed that many swampland constraints are speculative and evidence in favor of them is often based on examples in string theory and not so much on general arguments. A reasonable strategy to make progress in the research on swampland constraints may therefore be to concentrate on the best established ones, in particular the censorship of global symmetries. Indeed, recently this has been proven in a very special setting using holography [185, 186]. Motivated by this it seems to be particularly promising to study axionic shift symmetries and their relation to axion potentials, which is what we have done in Chapter 4.

Based on Chapter 4 the following directions for further research could be followed in order to improve our understanding of axion potentials and QG constraints. We did a first step towards identifying possible mechanisms by which QG could break axionic shift symmetries when we discussed gravitational instantons. This discussion could be extended by considering instantons with fluxes. Moreover we encountered an exactly flat axion in $\mathcal{N} = 2$ SUGRA although we expect the axionic shift symmetry to be broken in this stringy example. It would be important to understand how these two observations are compatible with each other in detail. Finally, it is worthwhile to explore the consequences of our proposed bound on axion potentials in phenomenologically relevant theories such as models of the QCD axion or axion dark matter.

# Own Publications


(1) H. Abramowicz *et al.* [H1 and ZEUS Collaborations], *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data Eur. Phys. J. C* **75** (2015) 580, [1506.06042].

(2) A. Hebecker, P. Henkenjohann and L. T. Witkowski, *What is the Magnetic Weak Gravity Conjecture for Axions?*, *Fortsch. Phys.* **65** (2017) 1700011, [1701.06553].

(3) A. Hebecker, P. Henkenjohann and L. T. Witkowski, *Flat Monodromies and a Moduli Space Size Conjecture*, *JHEP* **12** (2017) 033, [1708.06761].


This thesis is based on publications (2) and (3).

# Bibliography


[1] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, *Phys. Rev.* **D20** (1979) 2619–2625.

[2] R. Barbieri and G. F. Giudice, *Upper Bounds on Supersymmetric Particle Masses*, *Nucl. Phys.* **B306** (1988) 63–76.

[3] M. Dine, *Naturalness Under Stress*, *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 43–62, [1501.01035].

[4] S. Weinberg, *Implications of Dynamical Symmetry Breaking*, *Phys. Rev.* **D13** (1976) 974–996.

[5] S. Dimopoulos and H. Georgi, *Softly Broken Supersymmetry and SU(5)*, *Nucl. Phys.* **B193** (1981) 150–162.

[6] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, *Phys. Lett.* **B429** (1998) 263–272, [hep-ph/9803315].

[7] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370–3373, [hep-ph/9905221].

[8] N. Arkani-Hamed, A. G. Cohen and H. Georgi, *Electroweak symmetry breaking from dimensional deconstruction*, *Phys. Lett.* **B513** (2001) 232–240, [hep-ph/0105239].

[9] S. Rappoccio, *The experimental status of direct searches for exotic physics beyond the standard model at the Large Hadron Collider*, *Rev. Phys.* (2018) 100027, [1810.10579].

[10] G. F. Giudice, *The Dawn of the Post-Naturalness Era*, in *From My Vast Repertoire ...: Guido Altarelli's Legacy* (A. Levy, S. Forte and G. Ridolfi, eds.), pp. 267–292. 2019. 1710.07663. DOI.

[11] S. Hossenfelder, *Screams for Explanation: Finetuning and Naturalness in the Foundations of Physics*, 1801.02176.

[12] J. D. Wells, *Naturalness, Extra-Empirical Theory Assessments, and the Implications of Skepticism*, 1806.07289.

[13] M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory. Vol. 1: Introduction.* Cambridge Monographs on Mathematical Physics. 1988.

[14] M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory. Vol. 2: Loop Amplitudes, Anomalies and Phenomenology.* Cambridge Monographs on Mathematical Physics. 1988.

[15] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2007, 10.1017/CBO9780511816079.

[16] J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2007, 10.1017/CBO9780511618123.

[17] C. D. Hoyle, D. J. Kapner, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, U. Schmidt et al., *Sub-millimeter tests of the gravitational inverse-square law*, *Phys. Rev.* **D70** (2004) 042004, [hep-ph/0405262].

[18] G. D. Coughlan, W. Fischler, E. W. Kolb, S. Raby and G. G. Ross, *Cosmological Problems for the Polonyi Potential*, *Phys. Lett.* **131B** (1983) 59–64.

[19] J. R. Ellis, D. V. Nanopoulos and M. Quiros, *On the Axion, Dilaton, Polonyi, Gravitino and Shadow Matter Problems in Supergravity and Superstring Models*, *Phys. Lett.* **B174** (1986) 176–182.

[20] B. de Carlos, J. A. Casas, F. Quevedo and E. Roulet, *Model independent properties and cosmological implications of the dilaton and moduli sectors of 4-d strings*, *Phys. Lett.* **B318** (1993) 447–456, [hep-ph/9308325].

[21] T. Banks, D. B. Kaplan and A. E. Nelson, *Cosmological implications of dynamical supersymmetry breaking*, *Phys. Rev.* **D49** (1994) 779–787, [hep-ph/9308292].

[22] M. Grana, *Flux compactifications in string theory: A Comprehensive review*, *Phys. Rept.* **423** (2006) 91–158, [hep-th/0509003].

[23] M. R. Douglas and S. Kachru, *Flux compactification*, *Rev. Mod. Phys.* **79** (2007) 733–796, [hep-th/0610102].

[24] R. Bousso and J. Polchinski, *Quantization of four form fluxes and dynamical neutralization of the cosmological constant*, *JHEP* **06** (2000) 006, [hep-th/0004134].

[25] L. Susskind, *The Anthropic landscape of string theory*, hep-th/0302219.

[26] W. Taylor and Y.-N. Wang, *The F-theory geometry with most flux vacua*, *JHEP* **12** (2015) 164, [1511.03209].

[27] K. Becker, M. Becker and J. H. Schwarz, *String theory and M-theory: A modern introduction.* Cambridge University Press, 2006.

[28] D. Baumann and L. McAllister, *Inflation and String Theory.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2015, 10.1017/CBO9781316105733.

[29] M. Dine, *Supersymmetry and String Theory.* Cambridge University Press, 2016.

[30] C. Vafa, *The String landscape and the swampland*, hep-th/0509212.

[31] H. Ooguri and C. Vafa, *On the Geometry of the String Landscape and the Swampland*, *Nucl. Phys.* **B766** (2007) 21–33, [hep-th/0605264].

[32] F. Baume and E. Palti, *Backreacted Axion Field Ranges in String Theory*, *JHEP* **08** (2016) 043, [1602.06517].

[33] D. Klaewer and E. Palti, *Super-Planckian Spatial Field Variations and Quantum Gravity*, *JHEP* **01** (2017) 088, [1610.00010].

[34] E. Palti, *The Weak Gravity Conjecture and Scalar Fields*, *JHEP* **08** (2017) 034, [1705.04328].

[35] I. Valenzuela, *Backreaction Issues in Axion Monodromy and Minkowski 4-forms*, *JHEP* **06** (2017) 098, [1611.00394].

[36] R. Blumenhagen, I. Valenzuela and F. Wolf, *The Swampland Conjecture and F-term Axion Monodromy Inflation*, *JHEP* **07** (2017) 145, [1703.05776].

[37] M. Cicoli, D. Ciupke, C. Mayrhofer and P. Shukla, *A Geometrical Upper Bound on the Inflaton Range*, *JHEP* **05** (2018) 001, [1801.05434].

[38] R. Blumenhagen, D. Klwer, L. Schlechter and F. Wolf, *The Refined Swampland Distance Conjecture in Calabi-Yau Moduli Spaces*, *JHEP* **06** (2018) 052, [1803.04989].

[39] R. Blumenhagen, *Large Field Inflation/Quintessence and the Refined Swampland Distance Conjecture*, *PoS* **CORFU2017** (2018) 175, [1804.10504].

[40] M. Dias, J. Frazer, A. Retolaza, M. Scalisi and A. Westphal, *Pole N-flation*, 1805.02659.

[41] T. W. Grimm, E. Palti and I. Valenzuela, *Infinite Distances in Field Space and Massless Towers of States*, *JHEP* **08** (2018) 143, [1802.08264].

[42] B. Heidenreich, M. Reece and T. Rudelius, *Emergence of Weak Coupling at Large Distance in Quantum Gravity*, *Phys. Rev. Lett.* **121** (2018) 051601, [1802.08698].

[43] T. W. Grimm, C. Li and E. Palti, *Infinite Distance Networks in Field Space and Charge Orbits*, 1811.02571.

[44] P. Corvilain, T. W. Grimm and I. Valenzuela, *The Swampland Distance Conjecture for Kähler moduli*, 1812.07548.

[45] P. Draper and S. Farkas, *Gravitational Instabilities and Censorship of Large Scalar Field Excursions*, 1901.00515.

[46] A. Hebecker, P. Henkenjohann and L. T. Witkowski, *Flat Monodromies and a Moduli Space Size Conjecture*, *JHEP* **12** (2017) 033, [1708.06761].

[47] A. Hebecker, S. Leonhardt, J. Moritz and A. Westphal, *Thraxions: Ultralight Throat Axions*, 1812.03999.

[48] G. Buratti, J. Caldern and A. M. Uranga, *Transplanckian Axion Monodromy !?*, 1812.05016.

[49] A. Hebecker, D. Junghans and A. Schachner, *Large Field Ranges from Aligned and Misaligned Winding*, 1812.05626.

[50] N. Arkani-Hamed, L. Motl, A. Nicolis and C. Vafa, *The String landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[51] L. F. Abbott and M. B. Wise, *Wormholes and Global Symmetries*, *Nucl. Phys.* **B325** (1989) 687–704.

[52] S. R. Coleman and K.-M. Lee, *Wormholes Made Without Massless Matter Fields*, *Nucl. Phys.* **B329** (1990) 387–409.

[53] R. Kallosh, A. D. Linde, D. A. Linde and L. Susskind, *Gravity and global symmetries*, *Phys. Rev.* **D52** (1995) 912–935, [hep-th/9502069].

[54] T. Banks and N. Seiberg, *Symmetries and Strings in Field Theory and Gravity*, *Phys. Rev.* **D83** (2011) 084019, [1011.5120].

[55] G. Shiu, P. Soler and W. Cottrell, *Weak Gravity Conjecture and Extremal Black Hole*, 1611.06270.

[56] S. Hod, *A proof of the weak gravity conjecture*, *Int. J. Mod. Phys.* **D26** (2017) 1742004, [1705.06287].

[57] Z. Fisher and C. J. Mogni, *A Semiclassical, Entropic Proof of a Weak Gravity Conjecture*, 1706.08257.

[58] C. Cheung, J. Liu and G. N. Remmen, *Proof of the Weak Gravity Conjecture from Black Hole Entropy*, *JHEP* **10** (2018) 004, [1801.08546].

[59] Y. Hamada, T. Noumi and G. Shiu, *Weak Gravity Conjecture from Unitarity and Causality*, 1810.03637.

[60] A. Urbano, *Towards a proof of the Weak Gravity Conjecture*, 1810.05621.

[61] M. Montero, *A Holographic Derivation of the Weak Gravity Conjecture*, 1812.03978.

[62] C. Cheung and G. N. Remmen, *Naturalness and the Weak Gravity Conjecture*, *Phys. Rev. Lett.* **113** (2014) 051601, [1402.2287].

[63] B. Heidenreich, M. Reece and T. Rudelius, *Sharpening the Weak Gravity Conjecture with Dimensional Reduction*, *JHEP* **02** (2016) 140, [1509.06374].

[64] B. Heidenreich, M. Reece and T. Rudelius, *Evidence for a sublattice weak gravity conjecture*, *JHEP* **08** (2017) 025, [1606.08437].

[65] M. Montero, G. Shiu and P. Soler, *The Weak Gravity Conjecture in three dimensions*, *JHEP* **10** (2016) 159, [1606.08438].

[66] A. Hebecker, P. Mangat, S. Theisen and L. T. Witkowski, *Can Gravitational Instantons Really Constrain Axion Inflation?*, *JHEP* **02** (2017) 097, [1607.06814].

[67] P. Saraswat, *Weak gravity conjecture and effective field theory*, *Phys. Rev.* **D95** (2017) 025013, [1608.06951].

[68] A. Hebecker and P. Soler, *The Weak Gravity Conjecture and the Axionic Black Hole Paradox*, *JHEP* **09** (2017) 036, [1702.06130].

[69] A. Hebecker, P. Henkenjohann and L. T. Witkowski, *What is the Magnetic Weak Gravity Conjecture for Axions?*, *Fortsch. Phys.* **65** (2017) 1700011, [1701.06553].

[70] S. Andriolo, D. Junghans, T. Noumi and G. Shiu, *A Tower Weak Gravity Conjecture from Infrared Consistency*, *Fortsch. Phys.* **66** (2018) 1800020, [1802.04287].

[71] K. Freese, J. A. Frieman and A. V. Olinto, *Natural inflation with pseudo - Nambu-Goldstone bosons*, *Phys. Rev. Lett.* **65** (1990) 3233–3236.

[72] A. de la Fuente, P. Saraswat and R. Sundrum, *Natural Inflation and Quantum Gravity*, *Phys. Rev. Lett.* **114** (2015) 151303, [1412.3457].

[73] T. Rudelius, *On the Possibility of Large Axion Moduli Spaces*, *JCAP* **1504** (2015) 049, [1409.5793].

[74] T. Rudelius, *Constraints on Axion Inflation from the Weak Gravity Conjecture*, *JCAP* **1509** (2015) 020, [1503.00795].

[75] M. Montero, A. M. Uranga and I. Valenzuela, *Transplanckian axions!?*, *JHEP* **08** (2015) 032, [1503.03886].

[76] J. Brown, W. Cottrell, G. Shiu and P. Soler, *Fencing in the Swampland: Quantum Gravity Constraints on Large Field Inflation*, *JHEP* **10** (2015) 023, [1503.04783].

[77] T. C. Bachlechner, C. Long and L. McAllister, *Planckian Axions and the Weak Gravity Conjecture*, *JHEP* **01** (2016) 091, [1503.07853].

[78] A. Hebecker, P. Mangat, F. Rompineve and L. T. Witkowski, *Winding out of the Swamp: Evading the Weak Gravity Conjecture with F-term Winding Inflation?*, *Phys. Lett.* **B748** (2015) 455–462, [1503.07912].

[79] J. Brown, W. Cottrell, G. Shiu and P. Soler, *On Axionic Field Ranges, Loopholes and the Weak Gravity Conjecture*, *JHEP* **04** (2016) 017, [1504.00659].

[80] D. Junghans, *Large-Field Inflation with Multiple Axions and the Weak Gravity Conjecture*, *JHEP* **02** (2016) 128, [1504.03566].

[81] B. Heidenreich, M. Reece and T. Rudelius, *Weak Gravity Strongly Constrains Large-Field Axion Inflation*, *JHEP* **12** (2015) 108, [1506.03447].

[82] E. Palti, *On Natural Inflation and Moduli Stabilisation in String Theory*, *JHEP* **10** (2015) 188, [1508.00009].

[83] A. Hebecker, F. Rompineve and A. Westphal, *Axion Monodromy and the Weak Gravity Conjecture*, *JHEP* **04** (2016) 157, [1512.03768].

[84] A. D. Linde, *Chaotic Inflation*, *Phys. Lett.* **129B** (1983) 177–181.

[85] L. Boubekeur, *On the Scale of New Physics in Inflation*, 1312.4768.

[86] Planck collaboration, Y. Akrami et al., *Planck 2018 results. X. Constraints on inflation*, 1807.06211.

[87] BICEP2 collaboration, P. A. R. Ade et al., *BICEP2 II: Experiment and Three-Year Data Set*, *Astrophys. J.* **792** (2014) 62, [1403.4302].

[88] BICEP2 collaboration, P. A. R. Ade et al., *Detection of B-Mode Polarization at Degree Angular Scales by BICEP2*, *Phys. Rev. Lett.* **112** (2014) 241101, [1403.3985].

[89] BICEP2, Keck Array collaboration, P. A. R. Ade et al., *BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season*, *Phys. Rev. Lett.* **121** (2018) 221301, [1810.05216].

[90] T. Banks, M. Dine, P. J. Fox and E. Gorbatov, *On the possibility of large axion decay constants*, *JCAP* **0306** (2003) 001, [hep-th/0303252].

[91] D. H. Lyth, *What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy?*, *Phys. Rev. Lett.* **78** (1997) 1861–1863, [hep-ph/9606387].

[92] J. E. Kim, H. P. Nilles and M. Peloso, *Completing natural inflation*, *JCAP* **0501** (2005) 005, [hep-ph/0409138].

[93] S. Dimopoulos, S. Kachru, J. McGreevy and J. G. Wacker, *N-flation*, *JCAP* **0808** (2008) 003, [hep-th/0507205].

[94] L. McAllister, E. Silverstein and A. Westphal, *Gravity Waves and Linear Inflation from Axion Monodromy*, *Phys. Rev.* **D82** (2010) 046003, [0808.0706].

[95] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[96] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279–282.

[97] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223–226.

[98] J. E. Kim, *Weak Interaction Singlet and Strong CP Invariance*, *Phys. Rev. Lett.* **43** (1979) 103.

[99] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Can Confinement Ensure Natural CP Invariance of Strong Interactions?*, *Nucl. Phys.* **B166** (1980) 493–506.

[100] A. R. Zhitnitsky, *On Possible Suppression of the Axion Hadron Interactions. (In Russian)*, *Sov. J. Nucl. Phys.* **31** (1980) 260.

[101] M. Dine, W. Fischler and M. Srednicki, *A Simple Solution to the Strong CP Problem with a Harmless Axion*, *Phys. Lett.* **104B** (1981) 199–202.

[102] J. Preskill, M. B. Wise and F. Wilczek, *Cosmology of the Invisible Axion*, *Phys. Lett.* **B120** (1983) 127–132.

[103] L. F. Abbott and P. Sikivie, *A Cosmological Bound on the Invisible Axion*, *Phys. Lett.* **B120** (1983) 133–136.

[104] M. Dine and W. Fischler, *The Not So Harmless Axion*, *Phys. Lett.* **B120** (1983) 137–141.

[105] W. Hu, R. Barkana and A. Gruzinov, *Cold and fuzzy dark matter*, *Phys. Rev. Lett.* **85** (2000) 1158–1161, [astro-ph/0003365].

[106] L. Hui, J. P. Ostriker, S. Tremaine and E. Witten, *Ultralight scalars as cosmological dark matter*, *Phys. Rev.* **D95** (2017) 043541, [1610.08297].

[107] D. J. E. Marsh, *Axion Cosmology*, *Phys. Rept.* **643** (2016) 1–79, [1510.07633].

[108] R. Gregory, *Nonsingular global strings*, *Phys. Rev.* **D54** (1996) 4955–4962, [gr-qc/9606002].

[109] R. Gregory and C. Santos, *Space-time structure of the global vortex*, *Class. Quant. Grav.* **20** (2003) 21–36, [hep-th/0208037].

[110] J. P. Conlon and S. Krippendorf, *Axion decay constants away from the lamppost*, *JHEP* **04** (2016) 085, [1601.00647].

[111] U. H. Danielsson, N. Johansson and M. Larfors, *The World next door: Results in landscape topography*, *JHEP* **03** (2007) 080, [hep-th/0612222].

[112] M. C. Johnson and M. Larfors, *Field dynamics and tunneling in a flux landscape*, *Phys. Rev.* **D78** (2008) 083534, [0805.3705].

[113] P. Ahlqvist, B. R. Greene, D. Kagan, E. A. Lim, S. Sarangi and I.-S. Yang, *Conifolds and Tunneling in the String Landscape*, *JHEP* **03** (2011) 119, [1011.6588].

[114] A. P. Braun, N. Johansson, M. Larfors and N.-O. Walliser, *Restrictions on infinite sequences of type IIB vacua*, *JHEP* **10** (2011) 091, [1108.1394].

[115] A. R. Brown, *Hyperinflation*, 1705.03023.

[116] S. Mizuno and S. Mukohyama, *Primordial perturbations from inflation with a hyperbolic field-space*, 1707.05125.

[117] A. Achcarro, V. Atal, C. Germani and G. A. Palma, *Cumulative effects in inflation with ultra-light entropy modes*, *JCAP* **1702** (2017) 013, [1607.08609].

[118] A. Achcarro, R. Kallosh, A. Linde, D.-G. Wang and Y. Welling, *Universality of multi-field $\alpha$-attractors*, 1711.09478.

[119] T. Kobayashi and S. Mukohyama, *Effects of Light Fields During Inflation*, *Phys. Rev.* **D81** (2010) 103504, [1003.0076].

[120] S. Cremonini, Z. Lalak and K. Turzynski, *On Non-Canonical Kinetic Terms and the Tilt of the Power Spectrum*, *Phys. Rev.* **D82** (2010) 047301, [1005.4347].

[121] S. Cremonini, Z. Lalak and K. Turzynski, *Strongly Coupled Perturbations in Two-Field Inflationary Models*, *JCAP* **1103** (2011) 016, [1010.3021].

[122] C. van de Bruck and M. Robinson, *Power Spectra beyond the Slow Roll Approximation in Theories with Non-Canonical Kinetic Terms*, *JCAP* **1408** (2014) 024, [1404.7806].

[123] S. Renaux-Petel and K. Turzynski, *On reaching the adiabatic limit in multi-field inflation*, *JCAP* **1506** (2015) 010, [1405.6195].

[124] S. Kachru, M. B. Schulz and S. Trivedi, *Moduli stabilization from fluxes in a simple IIB orientifold*, *JHEP* **10** (2003) 007, [hep-th/0201028].

[125] J. Gomis, F. Marchesano and D. Mateos, *An Open string landscape*, *JHEP* **11** (2005) 021, [hep-th/0506179].

[126] M. Arends, A. Hebecker, K. Heimpel, S. C. Kraus, D. Lust, C. Mayrhofer et al., *D7-Brane Moduli Space in Axion Monodromy and Fluxbrane Inflation*, *Fortsch. Phys.* **62** (2014) 647–702, [1405.0283].

[127] A. Landete, F. Marchesano, G. Shiu and G. Zoccarato, *Flux Flattening in Axion Monodromy Inflation*, *JHEP* **06** (2017) 071, [1703.09729].

[128] H. A. Verrill, *Algorithm for Drawing Fundamental Domains*, *http://wstein.org/Tables/fundomain/index2.html* (2001) .

[129] W. A. Stein, *lecture notes on 'Modular Abelian Varieties'*, *http://wstein.org/edu/Fall2003/252/lectures/index.html* (2003) .

[130] A. Hebecker, S. C. Kraus and L. T. Witkowski, *D7-Brane Chaotic Inflation*, *Phys. Lett.* **B737** (2014) 16–22, [1404.3711].

[131] R. Blumenhagen, D. Herschmann and E. Plauschinn, *The Challenge of Realizing F-term Axion Monodromy Inflation in String Theory*, *JHEP* **01** (2015) 007, [1409.7075].

[132] I. Garca-Etxebarria, T. W. Grimm and I. Valenzuela, *Special Points of Inflation in Flux Compactifications*, *Nucl. Phys.* **B899** (2015) 414–443, [1412.5537].

[133] N. Cabo Bizet, O. Loaiza-Brito and I. Zavala, *Mirror quintic vacua: hierarchies and inflation*, *JHEP* **10** (2016) 082, [1605.03974].

[134] R. Donagi and E. Sharpe, *On the global moduli of Calabi-Yau threefolds*, 1707.05322.

[135] A. Gerhardus and H. Jockers, *Quantum periods of CalabiYau fourfolds*, *Nucl. Phys.* **B913** (2016) 425–474, [1604.05325].

[136] H. Abe, T. Kobayashi and H. Otsuka, *Natural inflation with and without modulations in type IIB string theory*, *JHEP* **04** (2015) 160, [1411.4768].

[137] K. Kooner, S. Parameswaran and I. Zavala, *Warping the Weak Gravity Conjecture*, *Phys. Lett.* **B759** (2016) 402–409, [1509.07049].

[138] M. J. Dolan, P. Draper, J. Kozaczuk and H. Patel, *Transplanckian Censorship and Global Cosmic Strings*, *JHEP* **04** (2017) 133, [1701.05572].

[139] T. Higaki, K. S. Jeong, N. Kitajima, T. Sekiguchi and F. Takahashi, *Topological Defects and nano-Hz Gravitational Waves in Aligned Axion Models*, *JHEP* **08** (2016) 044, [1606.05552].

[140] A. G. Cohen and D. B. Kaplan, *The Exact Metric About Global Cosmic Strings*, *Phys. Lett.* **B215** (1988) 67–72.

[141] J. Polchinski, *Open heterotic strings*, *JHEP* **09** (2006) 082, [hep-th/0510033].

[142] A. D. Linde, *Monopoles as big as a universe*, *Phys. Lett.* **B327** (1994) 208–213, [astro-ph/9402031].

[143] A. Vilenkin, *Topological inflation*, *Phys. Rev. Lett.* **72** (1994) 3137–3140, [hep-th/9402085].

[144] M. Berg, E. Pajer and S. Sjors, *Dante's Inferno*, *Phys. Rev.* **D81** (2010) 103535, [0912.1341].

[145] I. Ben-Dayan, F. G. Pedro and A. Westphal, *Towards Natural Inflation in String Theory*, *Phys. Rev.* **D92** (2015) 023515, [1407.2562].

[146] G. Dvali, *Three-form gauging of axion symmetries and gravity*, hep-th/0507215.

[147] N. Kaloper and L. Sorbo, *A Natural Framework for Chaotic Inflation*, *Phys. Rev. Lett.* **102** (2009) 121301, [0811.1989].

[148] N. Kaloper, A. Lawrence and L. Sorbo, *An Ignoble Approach to Large Field Inflation*, *JCAP* **1103** (2011) 023, [1101.0026].

[149] W. A. Hiscock, *Exact Gravitational Field of a String*, *Phys. Rev.* **D31** (1985) 3288–3290.

[150] J. R. Gott, III, *Gravitational lensing effects of vacuum strings: Exact solutions*, *Astrophys. J.* **288** (1985) 422–427.

[151] P. Laguna and D. Garfinkle, *Space-time of Supermassive U(1) Gauge Cosmic Strings*, *Phys. Rev.* **D40** (1989) 1011–1016.

[152] B. Linet, *On the supermassive U(1) gauge cosmic strings*, *Class. Quant. Grav.* **7** (1990) L75–L79.

[153] M. E. Ortiz, *A New look at supermassive cosmic strings*, *Phys. Rev.* **D43** (1991) 2521–2526.

[154] I. Cho, *Inflation and nonsingular space-times of cosmic strings*, *Phys. Rev.* **D58** (1998) 103509, [gr-qc/9804086].

[155] I. Cho and A. Vilenkin, *Space-time structure of an inflating global monopole*, *Phys. Rev.* **D56** (1997) 7621–7626, [gr-qc/9708005].

[156] G. Shiu, W. Staessens and F. Ye, *Widening the Axion Window via Kinetic and Stckelberg Mixings*, *Phys. Rev. Lett.* **115** (2015) 181601, [1503.01015].

[157] T. Vachaspati and A. Achucarro, *Semilocal cosmic strings*, *Phys. Rev.* **D44** (1991) 3067–3071.

[158] A. Aurilia, *The Problem of Confinement: From Two-dimensions to Four-dimensions*, *Phys. Lett.* **81B** (1979) 203–206.

[159] M. Luscher, *The Secret Long Range Force in Quantum Field Theories With Instantons*, *Phys. Lett.* **78B** (1978) 465–467.

[160] P. Di Vecchia and G. Veneziano, *Chiral Dynamics in the Large n Limit*, *Nucl. Phys.* **B171** (1980) 253–272.

[161] G. Dvali and L. Funcke, *Small neutrino masses from gravitational θ-term*, *Phys. Rev.* **D93** (2016) 113002, [1602.03191].

[162] G. Dvali and L. Funcke, *Domestic Axion*, 1608.08969.

[163] G. Dvali, *Topological Origin of Chiral Symmetry Breaking in QCD and in Gravity*, 1705.06317.

[164] A. Hebecker, T. Mikhail and P. Soler, *Euclidean wormholes, baby universes, and their impact on particle physics and cosmology*, 1807.00824.

[165] G. Shiu and W. Staessens, *Strong Dynamics and Natural Inflation*, *Phys. Rev.* **D98** (2018) 083504, [1807.00620].

[166] G. Shiu and W. Staessens, *Phases of Inflation*, *JHEP* **10** (2018) 085, [1807.00888].

[167] G. 't Hooft, *Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle*, *Phys. Rev.* **D14** (1976) 3432–3450.

[168] G. 't Hooft, *How Instantons Solve the U(1) Problem*, *Phys. Rept.* **142** (1986) 357–387.

[169] S. W. Hawking, *Space-Time Foam*, *Nucl. Phys.* **B144** (1978) 349–362.

[170] S. W. Hawking and C. N. Pope, *Symmetry Breaking by Instantons in Supergravity*, *Nucl. Phys.* **B146** (1978) 381–392.

[171] E. Silverstein and A. Westphal, *Monodromy in the CMB: Gravity Waves and String Inflation*, *Phys. Rev.* **D78** (2008) 106003, [0803.3085].

[172] F. Marchesano, G. Shiu and A. M. Uranga, *F-term Axion Monodromy Inflation*, *JHEP* **09** (2014) 184, [1404.3040].

[173] R. Blumenhagen and E. Plauschinn, *Towards Universal Axion Inflation and Reheating in String Theory*, *Phys. Lett.* **B736** (2014) 482–487, [1404.3542].

[174] A. I. Vainshtein, V. I. Zakharov, V. A. Novikov and M. A. Shifman, *ABC's of Instantons*, *Sov. Phys. Usp.* **25** (1982) 195.

[175] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1.*, *Phys. Rev.* **122** (1961) 345–358.

[176] Y. Nambu and G. Jona-Lasinio, *DYNAMICAL MODEL OF ELEMENTARY PARTICLES BASED ON AN ANALOGY WITH SUPERCONDUCTIVITY. II*, *Phys. Rev.* **124** (1961) 246–254.

[177] D. G. Caldi, *Quark Mass Generation by Instantons*, *Phys. Rev. Lett.* **39** (1977) 121.

[178] R. D. Carlitz, *Bound States from Instantons*, *Phys. Rev.* **D17** (1978) 3225.

[179] R. D'Auria, S. Ferrara and P. Fre, *Special and quaternionic isometries: General couplings in N=2 supergravity and the scalar potential*, *Nucl. Phys.* **B359** (1991) 705–740.

[180] R. Delbourgo and A. Salam, *The gravitational correction to pcac*, *Phys. Lett.* **40B** (1972) 381–382.

[181] T. Eguchi and P. G. O. Freund, *Quantum Gravity and World Topology*, *Phys. Rev. Lett.* **37** (1976) 1251.

[182] T. Eguchi, P. B. Gilkey and A. J. Hanson, *Gravitation, Gauge Theories and Differential Geometry*, *Phys. Rept.* **66** (1980) 213.

[183] S. W. Hawking, D. N. Page and C. N. Pope, *Quantum Gravitational Bubbles*, *Nucl. Phys.* **B170** (1980) 283–306.

[184] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Instanton Density in a Theory with Massless Quarks*, *Nucl. Phys.* **B163** (1980) 46–56.

[185] D. Harlow and H. Ooguri, *Constraints on symmetry from holography*, 1810.05337.

[186] D. Harlow and H. Ooguri, *Symmetries in quantum field theory and quantum gravity*, 1810.05338.