Instabilities of string vacua and cosmological spacetimes

Aalsma, L.

Instabilities of String Vacua and Cosmological Spacetimes

Lars Aalsma

Instabilities of String Vacua and Cosmological Spacetimes
Lars Aalsma

Instabilities of String Vacua and Cosmological Spacetimes

This work has been accomplished at the Institute for Theoretical Physics (ITFA) of the University of Amsterdam (UvA) and is part of the research program of the former Foundation for Fundamental Research on Matter (now NWO-I), which is part of the Netherlands Organization for Scientific Research (NWO).

ISBN: 978-94-6323-677-5

Cover: fotografierende

Instabilities of String Vacua and Cosmological Spacetimes

Academisch Proefschrift

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties

ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op dinsdag 2 juli 2019, te 14:00 uur

door

Lars Aalsma

geboren te Beverwijk

Promotiecommissie

Promotor

dr. J.P. van der Schaar	Universiteit van Amsterdam

Copromotor

dr. B.W. Freivogel	Universiteit van Amsterdam

Overige leden

dr. A. Castro Anich	Universiteit van Amsterdam
prof. dr. D.D. Baumann	Universiteit van Amsterdam
prof. dr. E.P. Verlinde	Universiteit van Amsterdam
prof. dr. D. Roest	Rijksuniversiteit Groningen
dr. B.S.E. Vercnocke	KU Leuven
prof. dr. G. Shiu	University of Wisconsin-Madison

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Publications

This thesis is based on the following publications:

Presented in Chapter 2:

[1] *Extremal Tunneling and Anti-de Sitter Instantons*,
L. Aalsma and J. P. van der Schaar,
JHEP **03** (2018) 145, `arXiv:1801.04930 [hep-th]`.

Presented in Chapter 3:

[2] *Back(reaction) to the Future in the Unruh-de Sitter State*,
L. Aalsma, M. Parikh, and J. P. van der Schaar,
`arXiv:1905.02714 [hep-th]`.

Presented in Chapter 4:

[3] *Constrained superfields on metastable anti-D3-branes*,
L. Aalsma, J. P. van der Schaar, and B. Vercnocke
JHEP **05** (2017)089, `arXiv:1703.05771 [hep-th]`.

[4] *Supersymmetric embedding of antibrane polarization*,
L. Aalsma, M. Tournoy, J. P. van der Schaar, and B. Vercnocke,
JHEP **03** (2018) 145, `arXiv:1807.03303 [hep-th]`.

Other publications by the author:

[5] *Weak Gravity Conjecture, Black Hole Entropy, and Modular Invariance*,
L. Aalsma, A. Cole, and G. Shiu,
`arXiv:1905.06956 [hep-th]`.

Contribution of the author to the publications:

[1] All computations in sections 2, 3 and the appendix were performed by me and I wrote a large part of these sections. Interpreting our results and connecting it to the Weak Gravity Conjecture was a joint effort.

[2] I suggested using the trace anomaly in two dimensions to determine the energy-momentum tensor in the Unruh state and performed all computations of the energy-momentum tensor in two and four dimensions. I proposed to use the second law to check the consistency of our result. All members of the collaboration contributed to conceptual discussions.

[3] Most of the computations in section 2, 3 and 4 were performed jointly by me and BV. I suggested using the non-Abelian perspective in section 3.2 to determine if the modified Dirac operator contains a zero mode. I performed the computations in appendix A.2.1 and contributed to all conceptual discussions.

[4] I performed the computations in section 2 and 3 and wrote most of section 1, 2, 3 and 4. All members of the collaboration contributed to conceptual discussions.

Contents

1 Introduction

1.1 Motivation

Imagine you are riding your bike, going home after a long night out. Suddenly, you start to wonder what time it is and if you will make it home before sunrise. To answer this question, it suffices to just use classical physics. The behaviour of the world at smaller scales, such as the structure of subatomic matter, is irrelevant. Indeed, while you are lost in thought about the microscopic world, night will still become day regardless of the detailed behaviour of quarks and gluons.

This *separation of scales* made it possible for physicists in the 20^{th} century to develop accurate theories describing a wide range of phenomena. For example, at (sub)atomic scales, *quantum field theory* describes the world of elementary particles and their interactions and this framework has been successfully applied to describe three of the four fundamental forces of nature: the strong, weak and electromagnetic interaction. The fourth force, gravity, is negligible at these scales due to its weakness and its effect only becomes apparent at the scale of humans, planets, solar systems and galaxies. Here, gravity dominates and its properties are described by Einstein's theory of *general relativity*.

Despite the success of quantum field theory and general relativity, there are situations where the physics at different scales cannot be disentangled. In Figure 1.1 the *Bronstein cube* is displayed, whose axes correspond to three fundamental constants of nature: the speed of light c, Planck's constant $\hbar$ and Newton's constant G_N. By moving along an axis starting from the classical physics origin, the effects parametrized by the constant on this axis become more important and we see how different regimes correspond to different physical theories.

Most of the corners of this cube are well understood. The one exception is the corner indicated with red where effects in c, $\hbar$ and G_N are all relevant. This is the regime of *quantum gravity* and the length scale at which quantum gravity effects

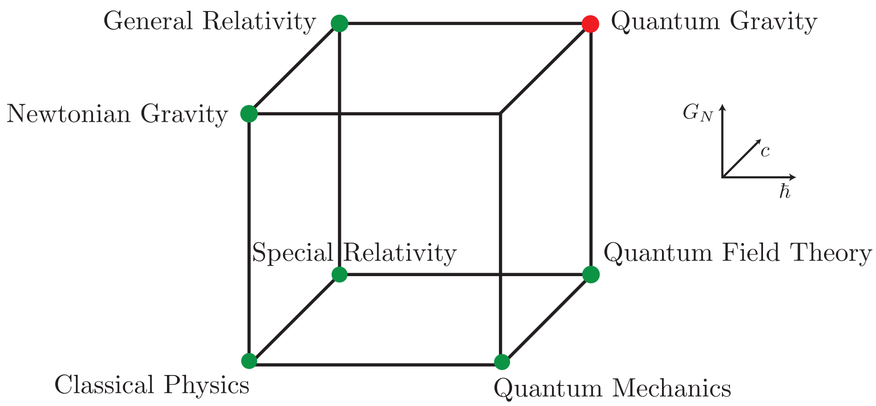

Figure 1.1: *The Bronstein cube describes the validity of different physical theories as a function of the importance of effects measured by three fundamental constants: the speed of light (c), Planck's constant ($\hbar$) and Newton's constant (G_N). One corner is left empty, which corresponds to a quantum mechanical version of Newtonian gravity that will not play any role in this thesis.*

become important is given by the *Planck length*, a length scale given by

$$\ell_p = \sqrt{\frac{\hbar G_N}{c^3}} = 1.6 \cdot 10^{-35}\,\text{m} \ . \tag{1.1}$$

One might expect that because this scale is so small, quantum gravity effects are always negligible. Surprisingly, this turns out to be false and there are systems even in our own galaxy where quantum gravity cannot be ignored, such as the black hole at the center of our galaxy.

Of course, black holes are solutions of general relativity but that does not mean that they can be completely described within classical general relativity. On the contrary, we know that general relativity breaks down at the singular center of a black hole where spacetime curvature becomes large and possibly even beyond. Moreover, as is illustrated by the black hole information paradox, even though the curvature at the horizon of large black holes is low compared to the Planck scale, this does not mean quantum gravity effects always decouple from low-energy physics at the horizon and naively applying effective field theory might yield incorrect results. Instead, a complete description of the black hole must come from a theory of quantum gravity.

In cosmology, something similar happens. The current (accelerated) expansion of the universe implies that in the past, the universe must have been smaller. When tracing back the expansion of the universe using general relativity, one finds a gravitational singularity not completely unlike the one inside a black hole. Also here, a complete understanding of the initial conditions that gave rise to our universe must come from a theory of quantum gravity. Moreover, just as in the black hole case low-energy physics does not seem to decouple from high-energy quantum effects. To explain the smallness of the cosmological constant for example, one needs to understand why quantum effects do not renormalize the cosmological constant to a much larger value.

We therefore have evidence from both black hole physics and cosmology that, even though quantum gravity effects are guaranteed to become important at the Planck scale, its effects propagate to much lower energies. This is good news, because it makes quantum gravity predictive as it implies that not any garden variety effective theory can be consistently coupled to it. Instead, for an effective theory to be successfully embedded into quantum gravity it must satisfy certain consistency conditions. The conclusion is that in many situations of interest, scale separation fails and one needs a theory of quantum gravity that a unifies the physics at small and large scales.

The most successful and best understood proposal for such a theory is *string theory*. By replacing elementary point particles with extended objects such as *strings* and *D-branes*, string theory consistently combines quantum field theory and general relativity into a unified framework. This has led to a precise understanding of quantum gravity in particular simple backgrounds. String theory has also been applied to cosmological solutions such as de Sitter space, a vacuum solution of general relativity that approximates the accelerated expansion of our universe well. Strikingly, despite the fact that the landscape of all string vacua is expected to be enormous[1], it is notoriously difficult to find (meta)stable de Sitter solutions in string theory. This has led to an active debate amongst string theorists whether or not string theory vacua are stable under quantum effects. As a result, not only the stability of de Sitter space has been questioned, but it has also been suggested that nonsupersymmetric anti-de Sitter space is unstable as well.

Over the last few years, there has been mounting evidence that quantum gravity effects indeed can destabilize vacuum solutions, unless they are protected by a symmetry. In this thesis, we will carefully study the instability of black holes, anti-de Sitter space and de Sitter space. It turns out that these kind of instabilities are rather universal and that they might be viewed as low-energy predictions of quantum gravity. This provides an opportunity to study quantum gravity without having to build an experiment that directly probes Planckian energies, which is extremely challenging if not impossible. The hope is that by studying these universal predictions, we will obtain a better understanding of the low-energy properties of

[1] A recent estimate of the number of F-theory vacua is $10^{272\,000}$ [6].

quantum gravity and learn which effective theories can be consistently embedded into quantum gravity.

Outline of this thesis. In the rest of this chapter, we will introduce some concepts and technicalities that will be of use in the upcoming chapters. We first discuss quantum effects of matter on top of a fixed gravitational background and the swampland conjectures. This will serve as a basis for Chapter 2 and Chapter 3 where we study quantum instabilities of black holes, anti-de Sitter space and de Sitter space. After that, we review some aspects of flux compactifications of string theory and highlight some important details regarding a particular construction of de Sitter vacua in string theory. This will prepare the reader for Chapter 4, where we construct the effective description of de Sitter vacua in string theory using constrained superfields. We will close this thesis in Chapter 5 with a summary and outlook.

1.2 Quantum field theory on a fixed background

Systems in which the curvature of spacetime is low compared to the length scale of interest need not to be described by a full theory of quantum gravity. Instead, we can treat gravity semi-classically by keeping the background fixed and only considering matter on top of the background as being quantum mechanical. We will now use this approach to derive some quantum effects in Rindler space, de Sitter space and black holes.

1.2.1 Rindler space

The simplest example where one can study quantum effects in gravity is Rindler space, which is the coordinate system in Minkowksi space that is natural from the point of view of an accelerating observer. Due to the acceleration of the Rindler observer, Rindler space has event horizons beyond which there is no causal contact with the Rindler observer. Hence, Rindler space only covers a quarter of the global Minkowksi Penrose diagram, see Figure 1.2.

For simplicity, we will focus our attention on two-dimensional Rindler space which is sufficient for our purposes. Consider two-dimensional Minkowski space.

$$ds^2 = -dt^2 + dx^2 \ . \tag{1.2}$$

Rindler coordinates in the left (‘L’) and right Rindler wedge (‘R’) can be defined

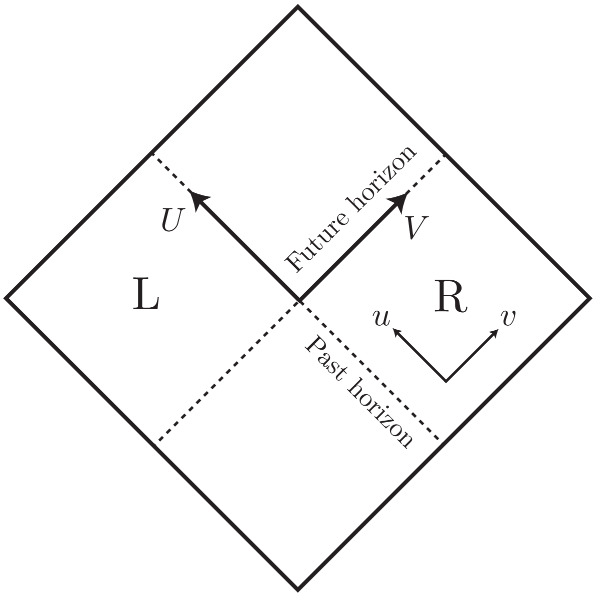

Figure 1.2: *Penrose diagram of Minkowski space with the left and right Rindler wedge indicated with 'L' and 'R'.*

as follows.

$$L: \quad t = -\frac{1}{a} e^{a\xi} \sinh(a\tau) \ , \qquad x = -\frac{1}{a} e^{a\xi} \cosh(a\tau) \ , \tag{1.3}$$

$$R: \quad t = \frac{1}{a} e^{a\xi} \sinh(a\tau) \ , \qquad x = \frac{1}{a} e^{a\xi} \cosh(a\tau) \ . \tag{1.4}$$

The metric in Rindler coordinates in a single wedge is given by

$$ds^2 = e^{2a\xi} \left(-d\tau^2 + d\xi^2 \right) \ . \tag{1.5}$$

Here, a is the acceleration of the Rindler observer. It will be convenient to also introduce lightcone coordinates with respect to Minkowksi and Rindler coordinates seperately.

$$U = t - x \ , \qquad u = \tau - \xi \ . \qquad V = t + x \ , \qquad v = \tau + \xi \ , \tag{1.6}$$

The direct relation between Minkowkski and Rindler lightcone coordinates is

$$U = -\frac{1}{a}e^{-au} \ , \qquad V = \frac{1}{a}e^{av} \ . \tag{1.7}$$

The metric in lightcone coordinates now reads

$$ds^2 = -dUdV = -e^{a(v-u)}dudv \ . \tag{1.8}$$

Having specified the background, we now introduce a massless scalar field $\varphi(t, x)$ on top of this background, whose action is given by

$$S = -\frac{1}{2}\int d^2x\sqrt{-g}\ \partial_\mu\varphi\partial^\mu\varphi \ . \tag{1.9}$$

Notice that the two-dimensional action of a massless scalar field on any two-dimensional spacetime is invariant under a conformal transformation.

$$g_{\mu\nu} \to \Omega(x)^2 g_{\mu\nu} \ . \tag{1.10}$$

This immediately implies that the equation of motion of the massless scalar field on any two-dimensional spacetime is given by the wave equation in flat space.

$$\partial_U\partial_V\varphi = 0 \ . \tag{1.11}$$

This allows us to treat incoming and outgoing modes that solve the wave equation separately where we define incoming and outgoing with respect to an asymptotic observer. A complete set of modes that are positive frequency with respect to Minkowski time t are given by

$$\varphi_\omega^{(\text{in})} = \frac{1}{\sqrt{4\pi\omega}}e^{-i\omega U} \ , \qquad \varphi_\omega^{(\text{out})} = \frac{1}{\sqrt{4\pi\omega}}e^{-i\omega V} \ . \tag{1.12}$$

Here, ω is the energy of the mode and the normalization is chosen such as to obey the standard Klein-Gordon norm

$$(\varphi_\omega, \varphi_{\omega'}) \equiv \int dt\left(\varphi^*_{\omega'}\overleftrightarrow{\partial}_t\varphi_\omega\right) = \delta(\omega - \omega') \ . \tag{1.13}$$

As usual, we can now expand the scalar field in terms of these mode functions and creation/annihilation operators.

$$\hat{\varphi} = \int_0^\infty \frac{d\omega}{\sqrt{4\pi\omega}}\left[e^{-i\omega U}\ \hat{a}_\omega + e^{+i\omega U}\ \hat{a}^\dagger_\omega\right] + (U \leftrightarrow V) \ . \tag{1.14}$$

The creation/annihilation operators obey the canonical commutation relations.

$$[\hat{a}_\omega, \hat{a}^\dagger_{\omega'}] = \delta(\omega - \omega') \ . \tag{1.15}$$

The Poincaré-invariant Minkowski vacuum is defined as being annihilated by $\hat{a}_k$ for both incoming and outgoing modes.

$$\hat{a}_\omega \left|0_M\right\rangle = 0 \; . \tag{1.16}$$

However, there is also an alternative way of quantizing the scalar field which makes use of Rindler coordinates. The wave equation is then given by

$$\partial_u \partial_v \varphi = 0 \; , \tag{1.17}$$

and the properly normalized mode functions in 'R' by

$$\begin{aligned} \varphi_\omega^{(\mathrm{in})} &= \frac{1}{\sqrt{4\pi\omega}} e^{-i\omega u} = \frac{1}{\sqrt{4\pi\omega}} (-aU)^{i\omega/a} \; , \\ \varphi_\omega^{(\mathrm{out})} &= \frac{1}{\sqrt{4\pi\omega}} e^{-i\omega v} = \frac{1}{\sqrt{4\pi\omega}} (aV)^{i\omega/a} \; . \end{aligned} \tag{1.18}$$

The mode expansion of the scalar field is now given by

$$\hat{\varphi} = \int_0^\infty \frac{d\omega}{\sqrt{4\pi\omega}} \left((-aU)^{i\omega/a} \hat{b}_\omega + (-aU)^{-i\omega/a} \hat{b}_\omega^\dagger \right) + (U \leftrightarrow V) \; , \tag{1.19}$$

with

$$[\hat{b}_\omega, \hat{b}_{\omega'}^\dagger] = \delta(\omega - \omega') \; . \tag{1.20}$$

The Rindler vacuum is defined as being annihilated by both incoming and outgoing Rindler modes.

$$\hat{b}_\omega \left|0_R\right\rangle = 0 \; . \tag{1.21}$$

The two different vacua can be related to each other by a so-called Bogoliubov transformation.

$$\hat{b}_\omega = \int_0^\infty d\omega' \left(\alpha_{\omega\omega'} \hat{a}_{\omega'} + \beta_{\omega\omega'} \hat{a}_{\omega'}^\dagger \right) \; . \tag{1.22}$$

A non-zero Bogoliubov coefficient $\beta_{\omega\omega'}$ indicates particle production, which can be seen by computing the expectation value of the number operator. Let us say we are interested in the number density of particles in the Minkowski vacuum as seen by a Rindler observer. He or she would compute

$$\left\langle 0_M \right| \hat{N}_\omega \left| 0_M \right\rangle = \left\langle 0_M \right| \hat{b}_\omega^\dagger \hat{b}_\omega \left| 0_M \right\rangle = |\beta_\omega|^2 \; , \tag{1.23}$$

where we used (1.22). So when $|\beta_\omega|^2 \neq 0$, it indicates that a Rindler observes sees the Minkowski vacuum as a state that contains particles. By equating the two mode expansions (1.14) and (1.19) it follows that

$$|\beta_\omega|^2 = \frac{1}{e^{2\pi\omega/a} - 1} \; . \tag{1.24}$$

We recognize this as the Bose-Einstein distribution of a thermal gas. The Rindler observer thus sees the Minkowksi vacuum as a thermal state with a temperature $T = a/(2\pi)$, which is the famous Unruh effect [7]. As we will see later, in more general situations we can attribute a temperature to an event horizon given by

$$T = \frac{\kappa}{2\pi} \ , \tag{1.25}$$

where κ is the surface gravity at the horizon.

To analyze the energy density in these two vacua we compute the vacuum expectation value of the energy-momentum tensor, given by

$$\langle T_{\mu\nu} \rangle = \langle \partial_\mu \varphi \partial_\nu \varphi \rangle - \frac{1}{2} \langle \partial_\rho \varphi \partial^\rho \varphi \rangle \eta_{\mu\nu} \ . \tag{1.26}$$

The expression for the energy-momentum tensor becomes particularly simple in lightcone coordinates, where we find

$$\begin{aligned} \langle T_{UU} \rangle &= \langle (\partial_U \varphi)^2 \rangle \ , \\ \langle T_{VV} \rangle &= \langle (\partial_V \varphi)^2 \rangle \ , \\ \langle T_{UV} \rangle &= 0 \ . \end{aligned} \tag{1.27}$$

A finite result can now easily be obtained by comparing the energy-momentum tensor in the Rindler and Minkowski vacuum. Using the mode expansion (1.19) and the Bogoliubov transformation we reproduce the following well-known result.

$$\begin{aligned} \langle 0_R | T_{UU} | 0_R \rangle - \langle 0_M | T_{UU} | 0_M \rangle &= - \int_0^\infty \frac{d\omega \ \omega}{2\pi a^2 U^2} |\beta_\omega|^2 = - \frac{1}{48\pi U^2} \ , \\ \langle 0_R | T_{VV} | 0_R \rangle - \langle 0_M | T_{VV} | 0_M \rangle &= - \int_0^\infty \frac{d\omega \ \omega}{2\pi a^2 V^2} |\beta_\omega|^2 = - \frac{1}{48\pi V^2} \ . \end{aligned} \tag{1.28}$$

By employing a regularization procedure, such as normal ordering with respect to the $\hat{a}_\omega$ operators, we can find the regularized expression for the energy-momentum tensor.

$$\begin{aligned} \langle 0_R | : T_{UU} : | 0_R \rangle &= - \frac{1}{48\pi U^2} \ , \qquad \langle 0_M | : T_{UU} : | 0_M \rangle = 0 \ , \\ \langle 0_R | : T_{VV} : | 0_R \rangle &= - \frac{1}{48\pi V^2} \ , \qquad \langle 0_M | : T_{VV} : | 0_M \rangle = 0 \ . \end{aligned} \tag{1.29}$$

Let us highlight some important aspects of this result.

The energy-momentum tensor is negative and diverges at the future ($U = 0$) and past ($V = 0$) horizon of the Rindler observer. The appearance of negative energy can be understood by writing the Minkowski vacuum as a thermofield double state between the left and right Rindler wedge.

$$|0_M\rangle = \prod_\omega \frac{1}{|\alpha_\omega|} \sum_n \left(e^{-\beta\omega} \right)^{n/2} |n\rangle_L \, |n\rangle_R \ . \tag{1.30}$$

Here, $|n\rangle_{L/R}$ is the Rindler state with n particles in the left/right wedge and β is the inverse Unruh temperature. This shows that the Minkowksi vacuum is a careful equilibrium between incoming and outgoing fluxes, such that the total energy-momentum vanishes. The Rindler vacuum then corresponds to the removal of both incoming and outgoing particles such that an accelerating observer sees an empty state. The result of course is that the energy associated with these quanta is also removed, such that the energy density becomes negative. In fact, the energy density violates the null energy condition $\langle T_{\mu\nu}\rangle k^\mu k^\nu \geq 0$, where k^μ is a null vector. Nevertheless, it can be shown that the total energy in the Rindler vacuum remains positive due to a singular, positive contribution at the horizon [8]. Due to the singularities in the Rindler vacuum, it is typically not considered as a reasonable physical state.

This finishes our discussion of Rindler space. As we will see next, many properties of Rindler space naturally generalize to de Sitter space and black holes.

1.2.2 De Sitter space

De Sitter space is maximally symmetric vacuum solution of Einstein's equations with a positive cosmological constant $\Lambda > 0$ and in four dimensions it has the isometry group $O(1,4)$, which is the Lorentz group in five dimensions. It can be defined as an embedding in Minkowski space in one dimension higher. So four-dimensional de Sitter space is defined as

$$-X_0^2 + X_1^2 + X_2^2 + X_3^2 + X_4^2 = \frac{1}{H^2} \ , \tag{1.31}$$

where $H = \sqrt{\Lambda/3}$ is the Hubble parameter. A convenient coordinate system in which de Sitter space is static is given by

$$\begin{aligned} X^0 &= H^{-1}\sqrt{1-H^2r^2}\sinh(Ht) \ , \\ X^1 &= H^{-1}\sqrt{1-H^2r^2}\cosh(Ht) \ , \\ X^2 &= r\sin\theta\cos\phi \ , \\ X^3 &= r\sin\theta\sin\phi \ , \\ X^4 &= r\cos\theta \ . \end{aligned} \tag{1.32}$$

In these coordinates, the metric becomes

$$ds^2 = -\left(1-H^2r^2\right)dt^2 + \left(1-H^2r^2\right)^{-1}dr^2 + r^2 d\Omega_2^2 \ . \tag{1.33}$$

Here, $d\Omega_2$ is the volume element of the unit two-sphere. We see that there is an event horizon located at $r = 1/H$. Static coordinates only cover a quarter of the global de Sitter Penrose diagram, indicated with blue in Figure 1.3.

It will be useful to introduce the tortoise coordinate

$$r_* = \int_0^r dr' \frac{1}{1 - H^2 r'^2} = \frac{1}{2H} \log\left(\frac{1+Hr}{1-Hr}\right) , \tag{1.34}$$

in which the metric becomes

$$ds^2 = \text{sech}^2(Hr_*)\left(-dt^2 + dr_*^2\right) + \frac{1}{H^2}\tanh^2(Hr_*) d\Omega_2^2 . \tag{1.35}$$

We will now consider the wave equation for a massless scalar field on this background

$$\partial_\mu\left(\sqrt{-g} g^{\mu\nu}\partial_\nu \varphi\right) = 0 , \tag{1.36}$$

and only take the scalar field to have a dependence on the time and radial coordinate. We then write

$$\varphi(t,r) = \frac{\phi(t,r)}{r} , \tag{1.37}$$

and obtain the following equation for $\phi(t,r)$

$$\left(-\partial_t^2 + \partial_{r_*}^2 + V(r)\right)\phi(t,r) = 0 . \tag{1.38}$$

Here, the potential is given by

$$V(r) = 2H^2(1 - H^2 r^2) . \tag{1.39}$$

We notice that close to the horizon, $V(r)$ vanishes. By introducing the lightcone coordinates[2]

$$u = t + r_* , \qquad v = t - r_* , \tag{1.40}$$

the wave equation in the near-horizon limit then reduces to

$$\partial_u \partial_v \phi(u,v) = 0 . \tag{1.41}$$

By comparing with (1.17) we see that this is simply the wave equation in Rindler space! The near-horizon region of de Sitter space can therefore be approximated by two-dimensional Rindler space, see Figure 1.3. This implies that we can derive some properties of de Sitter space by drawing the analogy with Rindler space.

Let us also introduce lightcone coordinates that can be extended to give a global cover of de Sitter space. These are defined as

$$U = -\frac{1}{H}e^{-Hu} , \qquad V = \frac{1}{H}e^{Hv} . \tag{1.42}$$

The wave equation near the horizon is of course

$$\partial_U \partial_V \varphi = 0 . \tag{1.43}$$

[2]Notice that compared to Rindler space, our definition of u, v is slightly different. This to ensure that the direction of the lightcone coordinates is still the same.

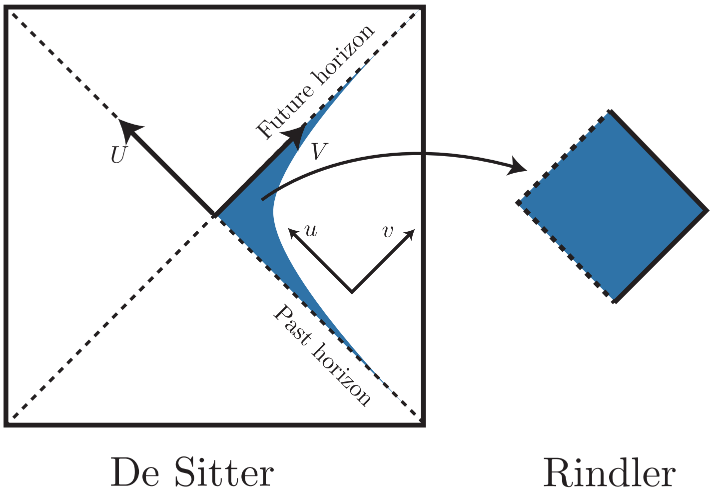

Figure 1.3: *Penrose diagram of de Sitter space. Static coordinates cover a quarter of de Sitter space. The near-horizon region is well approximated by Rindler space, which is shown by blue.*

In similar fashion as in Rindler space, we can now expand the scalar field in modes that are positive frequency with respect to either the static time coordinate or the global time coordinate and this will define different vacuum states. By comparing the expression for lightcone coordinates in Rindler space (1.7) with the analogous expression in de Sitter space (1.42), we see that we can think of the near-horizon geometry of de Sitter as a Rindler space where the acceleration of the Rindler observer is given by $a = H$. By doing so, we immediately conclude that a static observer in the static patch of de Sitter will see a thermal spectrum of particles at a temperature

$$T_{\mathrm{dS}} = \frac{H}{2\pi} , \tag{1.44}$$

which can be verified by explicit computation of the Bogoliubov transformation. In Chapter 3, we will further elaborate on the different vacuum states that one can define in de Sitter space and show how they affect the global evolution of de Sitter space.

1.2.3 Charged black holes

For Schwarzschild black holes, which are completely characterized by their mass, an entirely similar story as in de Sitter space goes through. We can introduce lightcone coordinates that are natural for an asymptotic observer outside of the black hole and also coordinates that can be extended to cover the maximally extended Schwarzschild solution. The near-horizon geometry becomes well approximated by a Rindler space with acceleration $a = 1/(4M)$ leading us to assign a temperature of

$$T_{\rm BH} = \frac{1}{8\pi M} \ , \tag{1.45}$$

to the black hole horizon, which indeed is the Hawking temperature.

When we add charge to the black hole, we find an additional feature. Consider the four-dimensional electrically charged Reissner-Nordström black hole.

$$ds^2 = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right) dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2 d\Omega_2^2 \ . \tag{1.46}$$

Here, M is the mass and Q the charge of the black hole. This black hole has two horizons located at

$$r_\pm = M \pm \sqrt{M^2 - Q^2} \ . \tag{1.47}$$

We see that the horizon is real only when $M \geq Q$, which is known as the extremality bound. When $M > Q$, the story is similar as for Schwarzschild black holes and the near-horizon geometry looks like Rindler space. As before, we can assign a temperature to the outer horizon of the black hole that can be computed by using that the near-horizon geometry is approximated by Rindler space.

$$T_{\rm RN} = \frac{1}{2\pi} \frac{\sqrt{M^2 - Q^2}}{\left(M + \sqrt{M^2 - Q^2}\right)^2} \ . \tag{1.48}$$

However, when we consider a so-called extremal black hole with $M = Q$ something special happens. First of all, we notice that the temperature goes to zero in this limit suggesting that the black hole has reached its 'ground state'. Secondly, when we look at the metric we find that the inner and outer horizon become degenerate and that the metric components become a perfect square.

$$ds^2 = -\left(1 - \frac{Q}{r}\right)^2 dt^2 + \left(1 - \frac{Q}{r}\right)^{-2} dr^2 + r^2 d\Omega_2^2 \ . \tag{1.49}$$

By introducing a new coordinate z as

$$r = Q + z \ , \tag{1.50}$$

and expanding for $z \ll Q$ we take a near-horizon limit. The resulting metric is given by

$$ds^2 = -\frac{z^2}{Q^2} dt^2 + \frac{Q^2}{z^2} dz^2 + Q^2 d\Omega_2^2 \ . \tag{1.51}$$

This is not the metric of Rindler space, but that of the Bertotti-Robinson geometry which locally corresponds to $AdS_2 \times S^2$. The radius of AdS_2 and the S^2 are equal and given by $\ell^2_{\text{AdS}} = Q^2$. The Penrose diagram of the extremal Reissner-Nordström black hole is displayed in Figure 1.4.

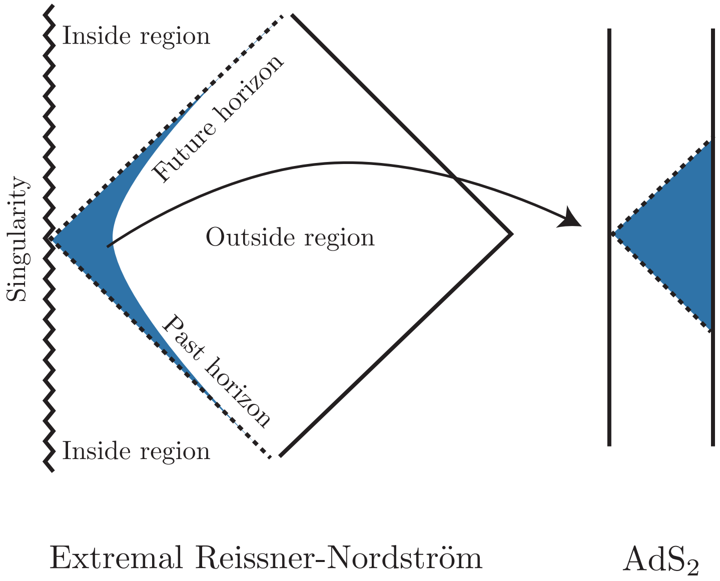

Figure 1.4: *Penrose diagram of the extremal Reissner-Nordström black hole, which extends indefinitely in the vertical direction. The blue region is the near-horizon geometry, which corresponds to* $AdS_2 \times S^2$.

We thus find that near-horizon physics of extremal charged black holes is not determined by Rindler space, but by the dynamics of AdS_2. Higher-dimensional extremal black holes also have an AdS_2 factor in their near-horizon geometry, but the sphere becomes higher dimensional. To obtain a higher-dimensional anti-de Sitter space as near-horizon factor, one needs to consider extended objects such as extremal black brane solutions.

These observations will be relevant in Chapter 2 where we show that a quantum instability of extremal charged black holes also implies that anti-de Sitter space is unstable.

1.3 Swampland conjectures

As we already briefly mentioned in section 1.1, high-energy quantum gravity effects do not always decouple from low-energy physics. This is most clearly illustrated by black hole physics and cosmological inflation. From the black hole information paradox, we know that we cannot fully trust effective field theory even at (large) horizon scales and the UV-sensitivity of inflation poses similar problems. While this complicates the treatment of black holes and cosmology, it also provides us with a unique opportunity because it suggests that we can learn about quantum gravity, without ever having to probe Planckian energies.

Viewed from a different perspective, this implies that quantum gravity puts non-trivial constrains on low-energy effective theories. This idea has been formulated more sharply in the context of the *swampland program* [9], whose goal is to distinguish low-energy effective theories that can and cannot be consistently coupled to quantum gravity. Theories that can be embedded into quantum gravity are said to be in the landscape, while theories that do not enjoy this property belong to the swampland, see Figure 1.5. To distinguish the swampland from the landscape several criteria, known as the *swampland conjectures*, have been put forward that a theory must satisfy for it to belong to the landscape, see [10, 11] for reviews. Most of these conjectures find their origin in string theory, but at least some of them are expected to hold more generally.

1.3.1 Weak Gravity Conjecture

Over the years, many different swampland conjectures have been put forward and one that is of particular relevance for this thesis is the Weak Gravity Conjecture (WGC) [12]. The WGC roughly states that in any quantum theory of gravity containing gauge forces, gravity should be the weakest force. More precisely, the WGC requires that a state exists that has the property that, if we take two of them, their gauge repulsion is always stronger than their gravitational attraction. A further refinement of the conjecture has been made by specifying the properties of the state(s) satisfying the WGC. This has led to different versions of the WGC such as the *mild form* (there should exist *some* state satisfying the WGC) and the *strong form* (the lightest charged state should satisfy the WGC). In later works, further sharpened versions were proposed such as the (sub)lattice [13] and tower WGC [14] for example.

But why should the WGC be true? Let us first give some loose arguments based on four-dimensional Einstein-Maxwell theory. In this case, the WGC requires the

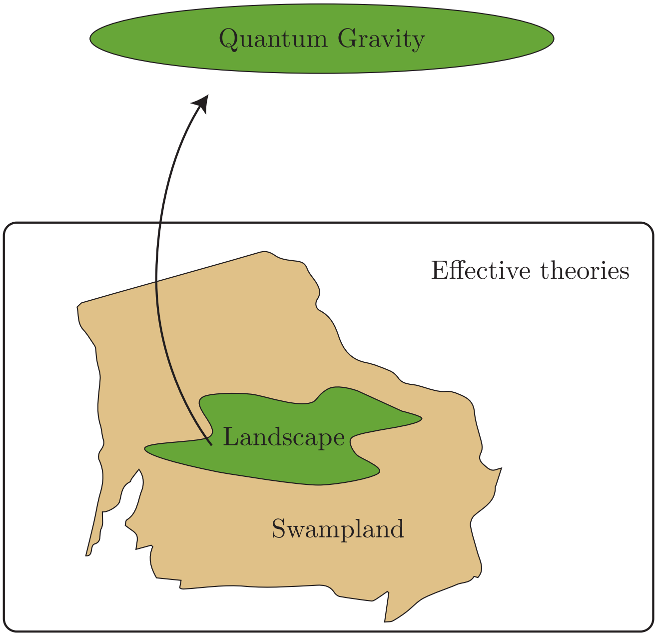

Figure 1.5: *In the space of all seemingly consistent low-energy effective theories only a subset can be coupled to quantum gravity. Theories that enjoy this property are said to belong the landscape, while theories that cannot be coupled to quantum gravity are in the swampland.*

existence of at least one massive, charged state with mass m and charge q whose mass-to-charge ratio obeys

$$m/q \leq \sqrt{g/G_N} \,, \tag{1.52}$$

where g is the gauge coupling. If we now take $g \to 0$ (but keep G_N finite) the gauge symmetry becomes indistinguishable from a global symmetry, because the interaction with the gauge field vanishes. For all practical purposes we have created a global symmetry, but we know global symmetries should not exist in quantum gravity [15–18]! For this reason, there should exist a lower bound on how small we can make the gauge coupling and the WGC precisely provides us with this, see (1.52). Exactly how strong (or weak) this bound is depends on the precise version of the WGC that is true. Notice that if the WGC-satisfying state is light enough, it is kinematically allowed for extremal charged black holes to decay. Hence, if the WGC would not be satisfied, one could imagine building a large tower of exactly stable (extremal) charged states, not neccesarily protected by any symmetry. The existence of such a large tower would be quite puzzling, as all these states would contribute to any scattering process by running in loops. This could potentially lead to pathologies, such as the violation of entropy bounds and renormalization of G_N to zero when the gauge coupling is sufficiently small [19].

Besides these somewhat heuristic arguments, also more precise derivations of a mild form of the WGC have been found by making use of IR constraints on scattering amplitudes [20, 21], black hole thermodynamics [22] and the AdS/CFT correspondence [23]. Finally, it has been observed that the WGC seems to be satisfied in all string compactifications.

In Chapter 2 we will relate the WGC to the decay of extremal charged black holes and see how this implies that the near-horizon anti-de Sitter space is unstable.

1.4 Vacuum solutions of string theory

In this section, we will introduce some of the key ideas used to find vacuum solutions of string theory and highlight some of the difficulties one faces when constructing them. First of all, (critical) string theories are defined in more dimensions than the four we observe at low energies. Hence, if we take ten-dimensional superstring theory, we have to compactify six of the dimensions on a compact space to arrive at a four-dimensional theory at low energies. Making sure that the resulting four-dimensional vacuum is stable is challenging and finding a vacuum that resembles the one we live in today is an even more difficult problem. Nevertheless, there has been progress in this direction and we now discuss some of the elements that are essential to obtain stable four-dimensional vacua.

1.4.1 Flux compactifications

To explain the idea of a compactification, we will consider a simple example (loosely based on examples in [24] and [25]) where we compactify the ten-dimensional Einstein-Hilbert action. The action is given by

$$S = \frac{1}{2\kappa^2} \int d^{10}X \sqrt{-G} e^{-2\Phi} R \ . \tag{1.53}$$

Here κ^2 is the ten-dimensional Newton's constant, Φ the dilaton and R the Ricci scalar constructed from the ten-dimensional metric given by

$$ds^2 = G_{MN} dX^M dX^N \ . \tag{1.54}$$

We now consider the situation where the ten-dimensional space $\mathcal{M}_{10}$ is a product of the form

$$\mathcal{M}_{10} = \mathcal{M}_4 \times Y_6 \ , \tag{1.55}$$

where $\mathcal{M}_4$ is a four-dimensional spacetime and Y_6 a compact six-dimensional space. We take the following ansatz for the metric

$$ds^2 = e^{-6\varphi(x)} g_{\mu\nu} dx^\mu dx^\nu + e^{2\varphi(x)} g_{mn} dy^m dy^n \ , \tag{1.56}$$

where $\mu, \nu = 0, .., 3$ and $m, n = 4, .., 9$. The field $\varphi(x)$ only depends on the four-dimensional spacetime and parametrizes the overall size of the compact space. We now reduce over the compact space by writing the action as

$$S = \frac{1}{2\kappa^2} \int d^4x \sqrt{-g_4} \int d^6y \sqrt{g_6} e^{-2\Phi} \ e^{-6\varphi} \left(\hat{R}_4 + \hat{R}_6 \right) \ . \tag{1.57}$$

Here, $\hat{R}_4$ and $\hat{R}_6$ are the Ricci scalars constructed from $e^{-6\varphi(x)} g_{\mu\nu}$ and $e^{2\varphi(x)} g_{mn}$. We can write this in terms of R_4 and R_6 constructed from $g_{\mu\nu}$ and g_{mn} respectively by using the fact that if two D-dimensional metrics are related by the Weyl rescaling

$$\hat{g}_{ab} = e^{2\omega} g_{ab} \ , \tag{1.58}$$

the Ricci scalars are related as

$$\hat{R} = e^{-2\omega} \left(R - 2(D-1)\nabla^2\omega - (D-2)(D-1)(\partial_a\omega\partial_b\omega) g^{ab} \right) \ . \tag{1.59}$$

Using this identity, we find

$$S = \frac{\mathcal{V}_6}{2\kappa^2} \int d^4x \sqrt{-g_4} e^{-2\Phi} \left(R_4 - 54 \partial_\mu\varphi\partial^\mu\varphi + e^{-8\varphi} R_6 \right) \ . \tag{1.60}$$

Here, $\mathcal{V}_6 = \int d^6y \sqrt{g_6}$ is the volume of the compact space and we assumed the six-dimensional Ricci scalar to be constant. We can canonically normalize φ and take the dilaton to be constant ($g_s = e^\Phi$) to find

$$S = \frac{M_p^2}{2} \int d^4x \sqrt{-g_4} \left(R_4 - \frac{1}{2} \partial_\mu\varphi\partial^\mu\varphi + e^{-\frac{4\varphi}{3\sqrt{3}}} R_6 \right) \ , \tag{1.61}$$

where we defined the four-dimensional Planck mass as $M_p^2 \equiv \frac{\mathcal{V}_6}{\kappa^2 g_s^2}$.

This example shows that we obtain the four-dimensional Einstein-Hilbert action, with in addition a scalar field with the potential $V(\varphi) = -e^{-\frac{4\varphi}{3\sqrt{3}}} R_6$. There are now three distinct possibilities.

1. $R_6 < 0$: In this case, the potential will drive $\varphi \to \infty$, such that the compact space will decompactify.

2. $R_6 = 0$: In this case, the potential vanishes and we have an exactly massless scalar field.

3. $R_6 > 0$: In this case, the potential will drive $\varphi \to 0$, such that we end up in an unbounded negative potential.

In all three cases, there is no stable vacuum where the scalar field φ obtains a positive mass squared. This is the problem of *moduli stabilization*. Moduli are scalar fields that parametrize deformations of the compact space and typically show up as light scalar fields in the lower-dimensional theory. Hence, to obtain a stable four-dimensional vacuum where the moduli are heavy enough such that they can be integrated out we need to include additional contributions to the action that generate a potential for the moduli.

In general, this presents an obstacle known as the *Dine-Seiberg problem* [26]. Let us say that φ is a modulus that controls a weak coupling expansion. Then, the leading-order result for the potential is given by the limit $\varphi \to \infty$ where the potential vanishes. For simplicity, let us consider $R_6 \neq 0$, but the argument also generalizes beyond this assumption. In this case, the leading correction to the potential leads to runaway behaviour, either to decompatification or to strong coupling. Hence, to obtain a stable minimum a higher-order correction needs to be included that gives a contribution similar in size. But if the next-to-leading order correction is of similar size, there is no reason to believe that we should not include the entire series of corrections. Unless the first two terms are accidentally similar in size and higher-order terms are small, a (meta)stable vacuum solution in string theory seems to be strongly coupled. For anti-de Sitter vacua we need competition between at least two terms, but for de Sitter vacua an additional third correction is required.

This problem can be evaded if we can find different sources of corrections to the potential that can be tuned independently of each other. As we will discuss now, one such source can come from fluxes.

Including fluxes

Additional contributions to the potential can come from higher-order (quantum) corrections in g_s and α' in the string action, but also fluxes. Because string theory comes with various p-form electric and magnetic fields, we can turn on fluxes in the compact space and as we will see this can lead to a stable vacuum. Let us say we turn on a constant three-form flux in the compact space. The Einstein-Hilbert action in ten dimensions is then given by

$$S = \frac{1}{2\kappa^2} \int d^{10}X \sqrt{-G} e^{-2\Phi} \left(R - e^{-6\varphi} |F_3|^2 \right) \ , \tag{1.62}$$

where indices on F_3 are raised and lowered with g_{mn}. By performing the same reduction to four dimensions, we find that the action is modified in the following way

$$S = \frac{M_p^2}{2} \int d^4x \sqrt{-g_4} \left(R_4 - 54 \partial_\mu \varphi \partial^\mu \varphi + e^{-8\varphi} R_6 - e^{-12\varphi} |F_3|^2 \right) \ . \tag{1.63}$$

The potential for the canonical scalar field is now given by

$$V(\varphi) = e^{-\frac{2\varphi}{\sqrt{3}}} |F_3|^2 - e^{-\frac{4\varphi}{3\sqrt{3}}} R_6 \ . \tag{1.64}$$

If the compact space has negative curvature ($R_6 < 0$) both terms contribute positively to the potential and the solution is driven to decompactification. On the other hand, if $R_6 > 0$ there is a competition between the two terms in the potential and we can obtain a stable vacuum with a negative cosmological constant, see Figure 1.6.

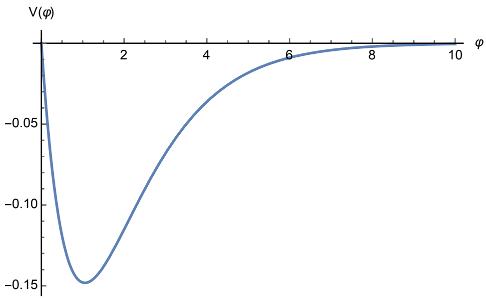

Figure 1.6: *The potential* (1.64) *for* $|F_3|^2 = R_6 = 1$, *which has a stable anti-de Sitter minimum.*

This was just a simple example, but it shows the essence of the challenge one faces when constructing (meta)stable vacua in string theory. As discussed before, to stabilize moduli there always needs to be a competition between different terms that contribute to the moduli potential and to evade the Dine-Seiberg problem it must be possible to tune these contributions separately.

Next, we will digress slightly and discuss $N = 1$ supersymmetry and its spontaneous breaking, which is a necessary step in any (semi-)realistic string compactification.

1.4.2 Supersymmetry breaking

Another crucial ingredient to obtain explicit and well-controlled vacua of string theory is supersymmetry. Supersymmetry is a powerful computational tool and vacua that preserve some amount of supersymmetry belong to the best-understood corners of string theory. Having said that, in our universe supersymmetry is necessarily broken due to the positive energy density of our vacuum and its breaking needs to be implemented in a controlled manner.

To still make use of (some amount of) unbroken supersymmetry along the way from ten to four dimensions, one typically considers compactifications that preserve (at least) $N = 1$ supersymmetry and only after that we introduce an ingredient that breaks supersymmetry at a relatively low energy scale. We are therefore mostly interested in theories that preserve $N = 1$ supersymmetry that contain a solution that breaks it spontaneously. We will now explain some useful techniques to describe spontaneous supersymmetry breaking assuming the reader has some familiarity with $N = 1$ supersymmetry. If not, the introductory review [27] will serve as a good basis.

Consider a theory with a single chiral superfield $S(\phi, \lambda, F)$ that contains a scalar field ϕ, a fermion λ and an auxiliary field F. A simple example of a globally supersymmetric theory for S is given by the following Kähler and superpotential

$$\begin{aligned} K &= S\bar{S} \ , \\ W &= fS \ . \end{aligned} \tag{1.65}$$

Here, f is a parameter with dimensions of (energy)^2. The scalar potential of this theory is given by

$$V = K^{A\bar{B}} \partial_A W \partial_{\bar{B}} W = f^2 \ . \tag{1.66}$$

We see that this theory breaks supersymmetry spontaneously as $\partial_S W \neq 0$ and the vacuum just corresponds to a flat direction where the scalar field is massless. Of course, because supersymmetry is broken spontaneously this also implies that the fermion λ is massless. This 'goldstone fermion' of supersymmetry breaking is usually referred to as the goldstino.

We could also imagine a similar theory, but now with additional higher-dimensional operators that give a contribution to the Kähler potential, an example that was considered in [28].

$$K = S\bar{S} - \frac{g_1}{\Lambda^2}(S\bar{S})^2 - \frac{g_2}{\Lambda^2}(S^3\bar{S} + \text{c.c}) + \mathcal{O}(\Lambda^{-3}) \ . \tag{1.67}$$

We take $g_1, g_2 > 0$. Notice that due to the higher-dimensional operators, this theory can only be viewed as an effective theory valid at energy scales $E \ll \Lambda$. The effect of the higher-dimensional operators is to induce a potential for ϕ, given by

$$V = f^2 \left(1 + 4\frac{g_1}{\Lambda^2}|\phi|^2 + 3\frac{g_2}{\Lambda^2}(\phi^2 + \bar{\phi}^2) + \mathcal{O}(\Lambda^{-3}) \right) \tag{1.68}$$

The mass of the scalar fields is given by

$$m^2 = \frac{f^2}{\Lambda^2}\left(4g_1 \pm 6g_2\right) \ , \tag{1.69}$$

and we impose $g_1 > \frac{3}{2}g_2$ to have a stable vacuum. We can now integrate out the scalar fields to obtain a theory that only contains the massless goldstino in its spectrum. The Lagrangian in terms of the component fields to lowest order in derivatives is given by [28]

$$\mathcal{L} = -f^2 + |F+f|^2 - \frac{g_1}{\Lambda^2}|2\phi F - \lambda|^2 - 3\frac{g_2}{\Lambda^2}\left(\left(\phi^2 F - \phi\lambda^2\right)\bar{F} + \text{c.c.}\right) \ . \tag{1.70}$$

The equation of motion for the scalar derived from this action is given by

$$\phi = \frac{\lambda^2}{2F} \ . \tag{1.71}$$

By plugging this back into the Lagrangian and also integrating out F, it reduces to

$$\mathcal{L} = -f^2 + (\text{derivatives of } \lambda) \ . \tag{1.72}$$

The goldstino is only derivatively coupled, which means that it must be invariant under $\lambda \to \lambda + \epsilon$, where ϵ is a constant spinor. Moreover, if we would have kept derivatives we would have obtained a Lagrangian that is invariant under the infinitesimal transformation [29]

$$\delta_\epsilon \lambda = \epsilon + (\lambda\sigma^\mu\bar{\epsilon})\partial_\mu\lambda \ . \tag{1.73}$$

We see that, despite the fact that supersymmetry is broken, the effective theory still enjoys a residual symmetry. Because the second term in the infinitesimal transformation is a term of order λ^2, we say that the goldstino realizes supersymmetry *nonlinearly*. This representation of the supersymmetry algebra was first discovered by Volkov and Akulov [30].

Furthermore, there also exists a convenient superfield procedure one can apply to directly arrive at this effective action. If we expand the superfield S in components as

$$S = \phi + \sqrt{2}\theta\lambda + \theta^2 F \ , \tag{1.74}$$

and square it we find

$$S^2 = \phi^2 + 2\phi\sqrt{2}\theta\lambda + \theta^2(2\phi F - \lambda^2) \ . \tag{1.75}$$

Now notice that when we plug in the scalar field equation of motion (1.71) in the expression for S^2, all three terms vanish independently (the first two due to the fact that fermions anticommute). We conclude that starting with the theory (1.65) and imposing

$$S^2 = 0 \ , \tag{1.76}$$

is a shortcut to directly integrate out the scalar degrees of freedom and to arrive at the effective action (1.72). Of course, the resulting theory is only valid at energies below the mass of the scalars that are integrated out. Hence, in general some UV-information is required (in this example knowledge of the value of Λ) to make sure that the mass of the degrees of freedom projected out by the constraint is large enough such that the resulting effective theory can be trusted at the energy scale of interest. If so, the upshot is that the resulting effective theory becomes universal, irrespective of its precise UV realization [28].

Constrained superfields can be generalized to supergravity, where the resulting expressions are more involved but the essence is the same as in the globally supersymmetric case. In Chapter 4, we will apply the formalism of constrained superfields to describe supersymmetry breaking by (anti-)D-branes. There, non-linear supersymmetry arises in a very natural way as D-branes are $\frac{1}{2}$-BPS objects. This means that half of the supersymmetries in ten-dimensions are realized linearly, while the other half is broken spontaneously and realized nonlinearly [31].

1.4.3 Brief review of KKLT

Now we put the different ingredients discussed in the previous subsections together and give a brief review of how they have been used to construct de Sitter solutions in string theory. We will focus on a particular construction that has been developed by Kachru, Kallosh, Linde and Trivedi (KKLT) in 2003 [32]. Even though we will only focus on the KKLT scenario, which will play a major role in Chapter 4, we should mention that after KKLT many other proposals for de Sitter vacua in string theory have been put forward (see for example [24]). Nevertheless, even 16 years after the KKLT paper was published there still is an active debate in the community about the consistency of de Sitter vacua in string theory. In this debate, the KKLT scenario serves as a concrete framework in which many of the concerns that have been raised can be addressed explicitly.

GKP compactifications

The starting point of many semi-realistic string compactifications are so-called *warped* compactifications, which means that there exists a region in the compact space that exhibits a large gravitational redshift (‘warping’). This allows one to make use of scale separation and obtain a low-energy four-dimensional vacuum in which stringy effects at high energies are decoupled. A particular metric ansatz for a warped compactification to four-dimensional flat space is of the form

$$ds^2 = e^{2A(y)}\eta_{\mu\nu}dx^\mu dx^\nu + e^{-2A(y)}g_{mn}dy^m dy^n \ . \tag{1.77}$$

Here, $e^{2A(y)}$ is the warp factor and $\eta_{\mu\nu}$ is the four-dimensional Minkowski metric. A concrete realization of such a compactification in type IIB supergravity was developed by Giddings, Kachru and Polchinski (GKP) [33], which we will review now closely following [24].

The GKP solution has non-zero self-dual five-form flux F_5 and three-form flux G_3, the latter of which is only turned on in the compact space. Explicitly, the five-form flux and three-form flux are given by

$$\begin{aligned} F_5 &= (1+\star_{10})d\alpha(y)\wedge dx^0\wedge dx^1\wedge dx^3\wedge dx^3 \ , \\ G_3 &= F_3 + ie^{-\Phi}H_3 \ . \end{aligned} \tag{1.78}$$

The ten-dimensional Einstein equations are now given by [24]

$$\nabla^2 e^{4A} = \frac{e^{8A}}{2\,\mathrm{Im}\,(\tau)}|G_3|^2 + e^{-4A}\left(|\partial\alpha|^2 + |\partial e^{4A}|^2\right) + 2\kappa^2 e^{2A}\mathcal{J}_{\mathrm{loc}} \ . \tag{1.79}$$

Here, ∇^2 is the six-dimensional Laplacian on the compact space, τ is the axiodilaton and $\mathcal{J}_{\mathrm{loc}}$ contains contributions to the energy-momentum tensor from localized objects (such as branes and orientifold planes) and is given by

$$\mathcal{J}_{\mathrm{loc}} = \frac{1}{4}\left(T^m_{\ \ m} - T^\mu_{\ \ \mu}\right) \ . \tag{1.80}$$

Without sources ($\mathcal{J}_{\mathrm{loc}} = 0$) the solution is trivial, which can be seen by integrating over the compact space: the left-hand side is zero and each term on the right-hand side must vanish identically as it consists of a sum of positive-definite terms. We therefore see that to obtain a non-trivial solution $\mathcal{J}_{\mathrm{loc}} < 0$.

An additional constraint on the solution comes from the Bianchi identify of F_5 which can be integrated over the compact space Y_6 to yield.

$$\frac{1}{2\kappa^2 T_{\mathrm{D3}}}\int_{Y_6} H_3\wedge F_3 + Q_3^{\mathrm{loc}} = 0 \ . \tag{1.81}$$

Here, T_{D3} is tension of an (anti-)D3-brane and Q_3^{loc} denotes the charge contribution of sources carrying D3-brane charge. Using the explicit form of F_5 and the integrated Bianchi identity, the ten-dimensional Einstein equations can be written as

$$\begin{aligned} \nabla^2\left(e^{4A}-\alpha\right) = \ &\frac{e^{8A}}{24\,\mathrm{Im}\,(\tau)}|iG_3 - \star_6 G_3|^2 + e^{-4A}|\partial(e^{4A}-\alpha)|^2 \\ &+ 2\kappa^2 e^{2A}\left(\mathcal{J}_{\mathrm{loc}} - \mathcal{Q}_{\mathrm{loc}}\right) \ , \end{aligned} \tag{1.82}$$

where $\mathcal{Q}_{\mathrm{loc}} = T_{\mathrm{D3}}\rho_3^{\mathrm{loc}}$ is the D3-brane charge density. A non-trivial solution to (1.82) is therefore only possible when all of the following conditions are satisfied.

1. $e^{4A} = \alpha$.
2. $\star_6 G_3 = iG_3$.
3. $\mathcal{J}_{\mathrm{loc}} = \mathcal{Q}_{\mathrm{loc}} < 0$.

Solutions satisfying these criteria are called imaginary self-dual (ISD).

As we have seen previously, turning on fluxes in the compact space can stabilize moduli. In the simple example we considered in subsection 1.4.1, we just considered a single modulus. In more realistic compactifications however, Y_6 is a Calabi-Yau three-fold that in addition to Kähler moduli (among which the volume modulus falls) also has complex structure moduli and an axiodilation that need to be stabilized.

By writing the ISD compactification in the language of $N = 1$ supersymmetry in terms of a Kähler and superpotential, it can be seen that the resulting theory contains a potential for the complex structure moduli and axiodilaton. The Kähler moduli remain unstabilized however, due to the *no-scale* structure of the potential, meaning the superpotential is independent of the Kähler moduli. Next, we will consider adding corrections and see how this can also stabilize the Kähler moduli.

Including quantum corrections

To stabilize the Kähler moduli, KKLT considered adding quantum corrections to the effective theory obtained from the GKP compactification. In the KKLT scenario, one makes use of the fact that the no-scale structure of the superpotential can be broken by non-perturbative effects. Two corrections that are typically considered and can generate a potential for Kähler moduli are gaugino condensation from wrapping D7-branes on 4-cycles or instantonic Euclidean D3-branes. It was argued in the original KKLT paper [32] that by including these non-perturbative effects, the Kähler moduli are indeed stabilized, although this has recently been challenged. For now, we will assume this procedure to be consistent, but we will come back to this important point later on.

After integrating out the massive fields and for simplicity only considering a single Kähler modulus T, the resulting Kähler and superpotential are of the form

$$\begin{aligned} K &= -3\log(T+\bar{T}) \\ W &= W_0 + \mathcal{A}e^{-aT} \ , \end{aligned} \tag{1.83}$$

where W_0 is a constant determined by the three-form flux. The exponential term in the superpotential parametrizes the non-perturbative effect and the value of $\mathcal{A}$ and a are set by the precise mechanism generating it. We can now compute the scalar potential

$$V = e^K \left(K^{T\bar{T}} D_T W D_{\bar{T}} \bar{W} - 3|W^2| \right) \ , \tag{1.84}$$

and write the result in terms of $\sigma = \frac{1}{2}(T+\bar{T})$. This leads to the following result

$$V = \frac{a\mathcal{A}e^{-a\sigma}}{2\sigma^2} \left(\left(1 + \frac{a}{3}\sigma\right) \mathcal{A}e^{-a\sigma} + W_0 \right) \ . \tag{1.85}$$

This potential has a minimum given by the solution to

$$D_T W = 0 \ , \tag{1.86}$$

showing that it is supersymmetric. Then, by plotting the potential in Figure 1.7 we find that we have obtained a stable vacuum!

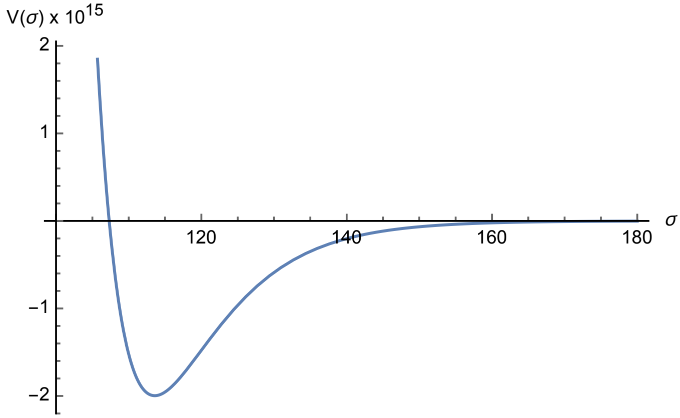

Figure 1.7: *The KKLT potential* (1.85) *for* $\mathcal{A} = 1, a = 0.1$ *and* $W_0 = -10^{-4}$. *The minimum is supersymmetric and corresponds to anti-de Sitter space.*

To summarize, by including a non-perturbative effect to the superpotential, we obtained a stable anti-de Sitter solution of string theory with all moduli stabilized. On top of this, by tuning fluxes appropriately (meaning taking W_0 small), the warp factor can be made exponentially small. This creates a throat-like region in the compact space where there is a large gravitational redshift between the region at the tip and top of the throat. Lets say that we have M units of F_3 flux and K units of H_3 flux with $K/Mg_s \gg 1$. Then, the warp factor at the tip of the throat is given by [33]

$$\exp(A_0) \simeq \exp\left(-\frac{2\pi K}{3g_s M}\right) \ll 1 \ . \tag{1.87}$$

Here, A_0 is the value of $A(y)$ at the tip. A cartoon of the resulting warped geometry is shown in Figure 1.8.

The small warp factor ensures that the physics in the low-energy four-dimensional anti-de Sitter vacuum can be decoupled from stringy effects at high energies.

Uplift to de Sitter

The final step in the KKLT scenario is to add a source to the supersymmetric anti-de Sitter vacuum that breaks supersymmetry and provides a positive contribution

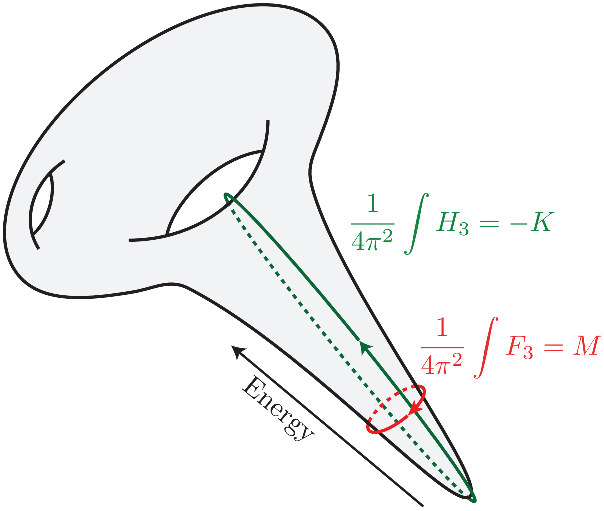

Figure 1.8: *Cartoon of the warped geometry in the compact space that is generated by putting M units of F_3 flux through the red cycle and K units of flux through the green cycle. By taking $K/Mg_s \gg 1$, there is a large gravitational redshift between the region at the tip and top of the throat.*

to the potential that 'uplifts' it from anti-de Sitter to de Sitter space. This is done by adding anti-D3-branes to the solution.

As can be seen from the (integrated) Bianchi identity (1.81), the flux of the $N = 1$ supersymmetric background contains D3-brane charge and adding D3-branes therefore preserves supersymmetry. Hence, a D3-brane obeys a 'no-force' condition meaning that the gravitational force it feels towards the tip of the throat exactly cancels with the force from the five-form flux. In contrast, anti-D3-branes carry an opposite charge and they are therefore naturally attracted towards the tip of the throat both gravitationally and by the five-form flux. At the tip of the throat, they therefore contribute a potential energy of (see (4.7) at $\psi = 0$ for example)

$$V = 2pT_{\mathrm{D3}}e^{4A_0} \ . \tag{1.88}$$

Here p is the number of anti-D3-branes and T_{D3} the tension of an (anti-)D3-brane. The factor of 2 signifies that the gravitational and electromagnetic potential energy add. Strictly speaking, we cannot just add this term to potential as we should take into account the dependence of the anti-D3-brane potential on the modulus σ, as was done in [32]. Also, we did not take into account the effect of brane polarization, which will lower the potential energy slightly, but we will come back to this point in Chapter 4. For the sake of illustration, we ignore these two subtleties. Then, adding the energy of anti-D3-branes to the potential (1.85), we find that for suitably chosen parameters the anti-de Sitter minimum is uplifted to a positive potential energy, corresponding to a metastable de Sitter vacuum in string theory.

At last, we have achieved our goal of constructing de Sitter vacua in string theory. Notice that the way we evaded the Dine-Seiberg problem is by balancing three different contributions to the effective potential: fluxes, non-perturbative quantum effects and anti-D3-branes.

Critical remarks

As mentioned before, we should be careful not to declare victory too soon as even today there still is an active debate concerning the consistency of constructions of de Sitter vacua in string theory, such as the KKLT scenario we just reviewed. Here, we make a selection of a few issues that have been raised and for a more complete overview we refer the reader to [34].

One point of concern has been raised in [35], where it has been argued that because supersymmetry is broken in the GKP solution (since $W_0 \neq 0$), one should include higher-order quantum corrections in the string action, even before adding anti-D3-branes or non-perturbative corrections. Although these corrections are not known completely in type IIB string theory, [35] took the point of view that this generically leads to a runaway behaviour of the potential that cannot be stabilized using an exponentially small non-perturbative effect. However, this conclusion has been challenged in [36], where effective field theory arguments were used to argue that by tuning fluxes, W_0 can be made very small such that the runaway behaviour due to non-zero W_0 can be safely ignored. To fully resolve the tension between both points of view, a better understanding of non-perturbative effects in time-dependent backgrounds is required, about which much is unknown currently. Recently, work towards this goal has been presented in [37].

A second issue that has received quite some attention recently is the description of non-perturbative effects, and in particular gaugino condensation, from a ten-dimensional point of view. This study was initiated in [38], where a tension was found between a ten-dimensional description of gaugino condensation and the four-dimensional results of [32]. However, these results were based on our incomplete knowledge of the ten-dimensional form of the action and in subsequent work [39–42], it was shown that this tension is resolved when a previously missed contribution to the ten-dimensional action is included (although this was challenged in [43]).

Furthermore, over the years it has also been questioned if the backreaction of supersymmetry-breaking anti-D3-branes in flux backgrounds is under control (see [44–59] for a subset of the relevant references). Anti-D3-branes have been shown to become singular when put inside the GKP background, and a fully backreacted supergravity solution in which it is revealed that this singularity is physical has never been constructed completely. Nonetheless, by imposing various consistency conditions on the solution, it has been argued that brane polarization can resolve the singularity, which allows for a consistent metastable solution [55, 59].

Finally, an issue that will be of particular relevance for this thesis is the nature of supersymmetry breaking by anti-D3-branes. As discussed in subsection 1.4.2, any theory that breaks $N = 1$ supersymmetry spontaneously should have a low-energy description in terms of constrained superfields. In particular, a universal prediction of spontaneous supersymmetry breaking is the presence of a goldstino that is exactly massless in the limit $M_p \to \infty$. It transform nonlinearly under the broken supersymmetry and can be captured by a nilpotent superfield. Hence, if supersymmetry breaking by anti-D3-branes in the KKLT scenario is spontaneous, as it should be for the resulting vacuum to be a bone fide solution of string theory, we should be able to describe the uplift procedure by making use of constrained superfields. However, in the way we explained how the uplift to de Sitter comes about we simply added the anti-D3-brane contribution to the effective potential. Instead, a proper treatment should reveal that their contribution can be described in a manifestly supersymmetric manner showing that supersymmetry is broken spontaneously and not explicitly.

Exactly how this works and what will be the implications for the effective theory describing de Sitter space in the KKLT scenario will be the main focus of Chapter 4.

2 Tunneling in Charged Black Holes

In this chapter, which is based on [1], we study quantum instabilities of charged black holes. We show how the *backreaction* of matter on the background geometry can be taken into account and illustrate how this leads to distinct instabilities. We calculate the decay rate of Reissner-Nordström black holes and find that it remains finite even after taking the extremal limit. As a consequence, the near-horizon anti-de Sitter space is also unstable, compatible with general expectations from the Weak Gravity Conjecture. This supports the conjecture that all nonsupersymmetric anti-de Sitter spaces are unstable [60, 61]. The different instabilities we uncover are all governed by the black hole entropy and are shown to have a unified and elegant description in terms of particles tunneling through the horizon.

2.1 Introduction

Ever since Hawking obtained his famous result for the thermal emission spectrum of black holes, an important question has been to understand, compute or estimate its leading corrections. The universal thermal nature of the spectrum is at the heart of the black hole information paradox and one unavoidable source of corrections is due to energy conservation: a black hole can only emit a particle with an energy at most equal to the mass of the black hole, implying the spectrum cannot be exactly thermal in any realistic microcanonical description.

In fact these backreaction corrections were first studied by Kraus and Wilczek by focusing on the dominant spherically symmetric sector of black hole emission [62, 63]. They imposed energy conservation by constructing a (non-local) effective action for the spherical shell in which the (radial) gravitational degrees of freedom are integrated out. It was subsequently suggested by Parikh and Wilczek that these results could also be interpreted, and more easily computed, in terms of the amplitude of a single particle tunneling through the horizon [64]. The tunneling approach clearly points towards a universal answer for the probability which is always equal to the change in the black hole horizon entropy before and after emission, which was already pointed out in earlier work by Massar and Parentani [65] using different methods. The fact that this probability is proportional to the change in the entropy supports the interpretation of the emission process, even after including backreaction, in terms of statistical thermodynamics [66].

Here, we would like to understand better the relation between the effective action approach of Kraus and Wilczek and the tunneling approach of Parikh and Wilczek, over which there has been some confusion over the years. In particular, the final result of the original Kraus-Wilczek paper does not match the (universal) tunneling result, although this was apparently remedied in [67] for the case of neutral spherical shells emitted from a Schwarzschild black hole. We will show, in a more general charged black hole setting, that the approach of Kraus and Wilczek is indeed equivalent to the tunneling approach. As a corollary, this provides a thorough and 'from first principles' effective action explanation for the validity of the tunneling approach, specified to the interesting case of charged particle shells. We will confirm that for a large range of parameters the probability for emission of charged spherical shells from a charged black hole indeed is proportional to

$$P_{\omega,q} \propto e^{\Delta S_{BH}} \ , \tag{2.1}$$

as in the neutral spherical shell case. Here $\Delta S_{BH} = S(M-\omega, Q-q) - S(M,Q)$ is the change in entropy of the black hole before and after emission of a spherical shell with energy ω and charge q. Although we will be considering charged emission from a four-dimensional charged black hole, the appearance of the entropy difference and its associated interpretation in terms of statistical thermodynamics strongly suggests this result also applies to higher-dimensional black holes and/or black branes (in the spherically symmetric sector).

After having carefully understood the detailed structure and universal nature of the result, we then study its implications in limits of interest. Specifically, we will show that the expression remains valid in the extremal limit of the black hole ($M = Q$) as long as the emitted particle shell satisfies $\omega \leq q$. The latter condition of course relates to the Weak Gravity Conjecture (WGC) [12], which essentially claims that in any consistent theory of quantum gravity there should exist a charged state whose mass is smaller than its charge in Planck units, that is $m \leq q$. This bound simply reflects the fact that an extremal black hole can only get rid of its charge by emitting a particle with $\omega \leq q$ to avoid creating a naked singularity. After all, one of the original motivations for the WGC was that an extremal black hole should be able to decay. We find that the probability to emit a charged particle satisfying the WGC from an extremal black hole is still nicely represented in terms of the entropy difference. To obtain a sensible result in the extremal limit crucially relies on including the backreaction of the shell. Moreover, we point out that the result remains applicable in a (non-thermal) regime of parameter space where the electrostatic potential energy dominates for some fixed particle charge q. At low enough energies the emission of particles of charge q enters the so-called 'superradiant' regime, where the entropy difference changes sign and the tunneling amplitude has to be reinterpreted. This low energy superradiant instability allows the black hole to quickly get rid of its charge, as originally noted and computed by Gibbons [68], but here we include the effect of backreaction.

With the generalized result for charged emission from charged black holes at our disposal we will then study its consequences in relation to a conjectured extension of the WGC, as put forward in [60, 61]. The claim of these authors is that only supersymmetric BPS states, which saturate the WGC bound, can be (meta)stable. If correct, this implies that not only (extremal) black holes, but also (nonsupersymmetric) anti-de Sitter vacua should feature universal decay channels. This conjectured extension of the WGC can be studied concretely in the context of an extremal Reissner-Nordström black hole, where the near-horizon geometry factorizes into AdS_2 times S^2. As we will show, the universal tunneling expression, when applied to the case of extremal to (non-)extremal emission, indeed implies a specific decay rate for non-extremal domain walls and at best metastability (fragmentation) for extremal domain walls in AdS spacetimes. When we expand our near-horizon result to leading order in backreaction, it exactly reproduces the general AdS_2 Euclidean instanton action of Maldacena, Michelson and Strominger which describes an instability in the nonsupersymmetric ($\omega < q$) case [69]. In the extremal case, this result matches the instanton first discovered by Brill [70], corresponding to AdS_2 fragmentation. As shown in [69], in the limit where one of the charges is very small, the fragmentation amplitude indeed coincides with the Euclidean action of the Brill instanton, which we explicitly relate to the tunneling amplitude. As a consequence, our results, which fully incorporate backreaction in the spherically symmetric sector, confirm and extend the existence of a family of gravitational instantons describing the decay of AdS vacua by the creation and subsequent expansion of super-extremal domain walls.

The rest of this chapter is organized as follows. In section 2.2 we employ the methods of Kraus and Wilczek to show that non-extremal charged black holes give rise to a universal decay rate of charged particles that is, after including backreaction, given by (2.1). We demonstrate that this is equivalent to the tunneling prescription of Parikh and Wilczek. Continuing, we then carefully study the extremal limit of this result in section 2.3 and identify a 'superradiant' region of parameter space where the charged emission is significantly enhanced as compared to the thermal regime. We go on to analyze the near-horizon limit of the tunneling calculation and interpret our results in terms of a family of gravitational instantons corresponding to instabilities of (nonsupersymmetric) *AdS* vacua. Finally, we discuss our results and present our conclusions in section 2.4. Some details on relevant integrals can be found in appendix 2.A.

2.2 Including gravitational backreaction

In order to study corrections to Hawking radiation from backreaction, Kraus and Wilczek used an effective action to derive their results [62, 63] whereas Parikh and Wilczek used a seemingly more ad-hoc approach by assuming particles tunneled through the horizon [64]. Obviously, since both approaches aim to incorporate the spherically symmetric part of the backreaction on the geometry of a black hole, the final result should be the same. However, on a technical level these approaches seem to be rather different and the expected agreement is far from obvious. Notably, the supplied boundary conditions, which are related to the physical interpretation, are different in the two cases. Despite these differences, we will in this section verify that the results are the same, and we will clarify some of the sources of confusion.

In order to include the effects of the energy of the spherical shell in the emission process one can either fix the total energy of the spacetime or the black hole mass. While Kraus and Wilczek fix the black hole geometry and let the ADM mass vary, Parikh and Wilczek fix the ADM mass and allow the black hole geometry to fluctuate. Because black hole evaporation corresponds to the loss of black hole mass-energy to the asymptotically flat space surroundings, one might be inclined to prefer the Parikh-Wilczek approach. However, we would like to emphasize that the effective action approach of Kraus and Wilczek, which might be considered a more rigorous derivation of the spherically symmetric dynamics including backreaction, can just as well be applied with the ADM mass fixed. As a consequence, one should be able to derive the (universal) Parikh-Wilczek tunneling answer from a first principles effective action method.

To illustrate this, we will first derive the decay rate of a charged spherical shell from a charged black hole using the Kraus-Wilczek effective action method and subsequently employ the tunneling perspective to arrive at the same result more directly. Along the way we will show how the Kraus-Wilczek computation [62, 63] reduces to the one performed by Parikh and Wilczek [64].

2.2.1 The effective action of a spherical shell

The central idea of the Kraus-Wilczek approach is that the dominant contribution to the emission flux in Hawking's original calculation is in the s-wave sector (spherical shells). Even though backreaction is in general hard, if not impossible, to keep track of, focusing on just the s-wave contribution allows backreaction to be incorporated. By constraining the gravitational degrees of freedom one arrives at a two-dimensional effective action for a spherical shell in the black hole background. Of particular importance are the boundary conditions that are needed to explicitly determine the on-shell action, which is then used in a WKB approximation to construct solutions to the (corrected) field equations. Using these corrected mode functions, the overlap can then be computed between appropriate energy eigenstates in terms of asymptotic Minkowski time and the Unruh vacuum state, which is selected by a specific initial condition for the mode functions near the horizon. Following through, one then arrives at Hawking's result plus corrections due to gravitational backreaction in the s-wave sector. In the original article [62] this was done for neutral emission from a Schwarzschild black hole, and in a follow-up article [63] the authors report to have worked out the result for charged emission as well. We will summarize the computation in the more general case of charged emission below and show that the corrected final result agrees with the elegant and universal answer that is naturally obtained and understood from a perspective of particle tunneling.

Lets us start by considering a four-dimensional Reissner-Nordström black hole with mass M and electric charge Q with the metric and gauge field A given by

$$\begin{aligned} ds^2 &= -f(r)dt^2 + f(r)^{-1}dr^2 + r^2 d\Omega_2^2 \,, \\ f(r) &= 1 - \frac{2M}{r} + \frac{Q^2}{r^2} \,, \\ A &= -\frac{Q}{r}dt \,. \end{aligned} \tag{2.2}$$

We introduce the standard notation for the inner and outer horizon of the black hole.

$$r_\pm = M \pm \sqrt{M^2 - Q^2} \,. \tag{2.3}$$

The metric (2.2) contains a coordinate singularity at $r_\pm$, so in order to construct regular mode functions for a freely falling observer we introduce coordinates that are regular across the horizon. A particularly useful choice are the Painlevé-Gullstrand coordinates [71]. We define a new time coordinate as

$$t_p = t + g(r) \,, \qquad g(r) = \int dr \, \frac{\sqrt{1 - f(r)}}{f(r)} \,, \tag{2.4}$$

such that the metric becomes (dropping the subscript p)

$$ds^2 = -f(r)dt^2 + 2\sqrt{1-f(r)}dtdr + dr^2 + r^2 d\Omega_2 \ , \tag{2.5}$$

which is regular at the horizon.

Quantization

Now that we have specified the details of the background, we turn to quantization of (spherically symmetric) fields in this background. Since we are interested in charged radiation, we consider a complex scalar field $\phi(t,r)$ and write down its mode expansion. Considering modes that are positive frequency with respect to the Killing time that is used by an asymptotic observer we write

$$\phi(t,r) = \int dk \, \left(\hat{c}_k u_k^q(r) e^{-i\omega_k t} + \hat{d}_k^\dagger \bar{u}_k^{-q}(r) e^{+i\omega_k t}\right) \ . \tag{2.6}$$

Here $u_k^q(r)$ denotes a particle mode function with positive charge q and $\bar{u}_k^{-q}(r)$ an anti-particle mode function with negative charge $-q$. The bar indicates complex conjugation. Furthermore, k is the wavenumber with ω_k the corresponding energy. We can now define the vacuum of an asymptotic observer as

$$\hat{c}_k \left|0_A\right\rangle = \hat{d}_k \left|0_A\right\rangle = 0 \ . \tag{2.7}$$

Alternatively, we can expand the scalar field in a different set of modes that are positive frequency with respect to a freely falling observer as

$$\phi(t,r) = \int dk \, \left(\hat{a}_k v_k^q(t,r) + \hat{b}_k^\dagger \bar{v}_k^{-q}(t,r)\right) \ , \tag{2.8}$$

and define the vacuum of a freely falling (Unruh) observer as

$$\hat{a}_k \left|0_U\right\rangle = \hat{b}_k \left|0_U\right\rangle = 0 \ . \tag{2.9}$$

The two sets of creation and annihilation operators are related by the following Bogoliubov transformations

$$\hat{c}_k = \int dk' \left(\alpha_{kk'} \hat{a}_k + \beta_{kk'} \hat{b}_k^\dagger\right) \ , \tag{2.10}$$

$$\hat{d}_k^\dagger = \int dk' \left(\bar{\alpha}_{kk'} \hat{b}_k^\dagger + \bar{\beta}_{kk'} \hat{a}_k\right) \ . \tag{2.11}$$

The Bogoliubov coefficients can be expressed in terms of the following integrals

$$\alpha_{kk'} = \frac{1}{2\pi u_k^q(r)} \int_{-\infty}^{\infty} dt \, e^{i\omega_{k'} t} v_k^q(t,r) \ , \tag{2.12}$$

$$\beta_{kk'} = \frac{1}{2\pi u_k^q(r)} \int_{-\infty}^{\infty} dt \, e^{i\omega_{k'} t} \bar{v}_k^{-q}(t,r) \ , \tag{2.13}$$

which satisfy the standard orthonormality and completeness constraints. Selecting the Unruh vacuum state, the amplitude for detecting n particles (and n anti-particles) with momentum k is determined by the overlap

$$\Gamma_k^n = \langle 0_U | n_k^q, n_{-k}^{-q} \rangle \ . \tag{2.14}$$

The average number of particles with momentum k in the Unruh state, introducing the number operator $\hat{N}_k \equiv \hat{c}_k^\dagger \hat{c}_k$, is given by

$$\langle 0_U | \hat{N}_k | 0_U \rangle = \int dk' |\beta_{kk'}|^2 \ . \tag{2.15}$$

Assuming the different mode expansions are defined on the same spatial slices, as will be our case of interest, the Bogoliubov matrices will be diagonal and the k' index can be dropped. Integrating over all modes and appropriately regulating the expression (by introducing a finite space-time volume) one arrives at the result for the (average) total flux of asymptotically observed particles. The integrand, corresponding to the average flux density with an energy between ω_k and $\omega_k + d\omega_k$, equals

$$F(\omega_k) = \frac{d\omega_k}{2\pi} \frac{\Omega(\omega_k)}{|\alpha_k|^2/|\beta_k|^2 - 1} \ . \tag{2.16}$$

Here an additional greybody factor $\Omega(\omega_k)$ was introduced that describes the effects of rescattering off the potential. In the case that we would ignore backreaction, assuming the emission were exactly thermal and the background perfectly transparent ($\Omega(\omega_k) = 1$), the ratio $|\beta_k|^2/|\alpha_k|^2$ would equal the Boltzmann factor characterizing the Bose-Einstein distribution. Throughout this chapter, we will ignore the effects of a non-trivial greybody factor, implying that the probability for the black hole to emit a single quantum with momentum k equals

$$P_k = |\Gamma_k^1|^2 = \frac{|\beta_k|^2}{(1 + |\beta_k|^2)^2} = \frac{1}{|\alpha_k|^2} \frac{|\beta_k|^2}{|\alpha_k|^2} \ , \tag{2.17}$$

where in the final equality we used the normalization $|\alpha_k|^2 - |\beta_k|^2 = 1$ which, as we will explain later, cannot be assumed for charged particle emission at low enough energies. For a thermal distribution this probability is of course proportional to the Boltzmann factor.

In principal, we need to know the explicit form of the mode functions $v_k(t, r)$ in order to calculate the Bogoliubov coefficients, which is not straightforward. However, as was argued by [62, 63], the mode functions take on a simple form in the WKB approximation, which is valid for short wavelengths. Because modes near the horizon are infinitely blueshifted with respect to an asymptotic observer, this approximation should be valid as long as the modes are close enough to the horizon, which is then sufficient to determine the Bogoliubov coefficients describing the emission process. Thus, the mode functions are assumed to be of the following WKB form

$$v_k^q(t,r) = e^{iS_k^q(t,r)} \,, \qquad \bar{v}_k^{-q}(t,r) = e^{-iS_k^{-q}(t,r)} \ . \tag{2.18}$$

Here, $S_k^{\pm q}(t,r)$ is the classical action of the shell and the superscript $\pm q$ indicates the charge of the solution. To obtain an explicit and useful expression for the effective action, we will make use of a Hamiltonian formalism.

Effective action from a Hamiltonian formalism

In [62,63] a Hamiltonian formalism is used to derive the effective action of a particle in the s-wave approximation, that is describing the dynamics of a shell in a black hole background incorporating the backreaction of the shell. As is well known, in order for the variation of the gravitational action to vanish when evaluated on the equations of motion, it is necessary to supplement the action with boundary terms that cancel the ones that are induced by the variation, see for example [72]. For non-extremal black holes in asymptotically flat spacetimes there are two types of surface terms that require cancellation. One of these is defined asymptotically and yields the ADM mass of the spacetime, whereas the second one is defined on the black hole horizon and is related to its area and surface gravity [73].

In [62,63] the geometry of the black hole is kept fixed and the ADM mass is allowed to vary to satisfy the Hamiltonian constraints. This means that the boundary term on the horizon vanishes and we only need to subtract the asymptotic boundary term from the action to have a well-defined variational principle. However, we could also have fixed the ADM mass as was done in [64]. In this case, we should add the boundary term defined on the horizon to the action.

For all practical purposes, this means that in the Kraus-Wilczek approach the evolution of the shell is determined by the ADM mass and total charge of the system (and therefore the geometry *outside* the shell), whereas in the Parikh-Wilczek method it is the mass and charge of the black hole that determines the evolution (the geometry *inside* the shell). This is precisely the difference in boundary conditions that we alluded to before. For the purpose of consistently comparing with [63] we will fix the black hole geometry for now, but it should be stressed that this is just a choice and we could just as well have fixed the ADM mass.

By solving the Hamiltonian constraints, it was found in [63] that the introduction of a massless shell with energy ω and charge q splits the spacetime into two parts with mass parameter $\mathcal{M}(r)$, which appears in the metric as $f(r) = 1 - 2\mathcal{M}(r)/r$.

$$\mathcal{M}(r) = \begin{cases} M - \frac{Q^2}{2r} & (r < \hat{r}) \ , \\ \\ M + \omega - \frac{(Q+q)^2}{2r} & (r > \hat{r}) \ . \end{cases} \tag{2.19}$$

Here, $\hat{r}$ is the position of the shell. The classical action of the shell can be written as [62, 63]

$$S_k^q(t, r(t)) = S_k^q(0, r(0)) + \int_{r(0)}^{r(t)} dr\, p_c - (M_+ - M)t \ . \tag{2.20}$$

In this expression, $r(t)$ is trajectory of the shell, p_c its canonical momentum whose explicit form is given in (2.30) and M_+ the ADM mass of the spacetime. We subtracted the contribution of the black hole from the action such that the Hamiltonian is that of the shell: $(M_+ - M) = \omega$. From Hamilton's equations we find that the equation of motion of the shell is

$$\dot{r} = \frac{\partial H}{\partial p_c} = 1 - \sqrt{\frac{2(M+\omega)}{r} - \frac{(Q+q)^2}{r^2}} \ . \tag{2.21}$$

This is the equation of motion of an outgoing null geodesic in a spacetime with a black hole of mass $M + \omega$ and charge $Q + q$, as can be seen by solving

$$g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} = 0 \ , \tag{2.22}$$

in Painlevé-Gullstrand coordinates. This indeed agrees with our earlier observation that the imposed boundary conditions on the ADM mass and total charge should determine the evolution of the shell.

To find an explicit expression for the trajectory $r(t)$, we need to specify the initial position $r(t = 0)$. The natural choice is to demand the standard positive and negative frequency modes for a freely falling observer that crosses the horizon (the Unruh vacuum), that is we impose that the mode functions take the form of a standard plane wave at $t = 0$:

$$S_k^q(0, r(0)) = kr(0) \ . \tag{2.23}$$

This means that the (diagonal) Bogoliubov coefficients can be written as

$$\alpha_k = \frac{1}{2\pi u_k^q(r)} \int_{-\infty}^{\infty} dt\, e^{i\omega_k t + iS_k^q(t,r)} \ , \tag{2.24}$$

$$\beta_k = \frac{1}{2\pi u_k^q(r)} \int_{-\infty}^{\infty} dt\, e^{i\omega_k t - iS_k^{-q}(t,r)} \ . \tag{2.25}$$

We can compute these integrals by a saddle point approximation. The saddle points of these integrals are solutions to

$$\omega_k \pm \frac{\partial S_k^{\pm q}(t,r)}{\partial t} = 0 \ , \tag{2.26}$$

where the plus sign corresponds to α_k and the minus sign to β_k. This simply indicates that the different saddle point trajectories have opposite energy. Because $\partial S_k^q/\partial t$ is minus the Hamiltonian we find that the solution to (2.26) is given by

$$M_+ = M \pm \omega_k \ , \tag{2.27}$$

and the Bogoliubov coefficients at the saddle points are given by

$$\alpha_k \propto \exp\left(ikr(0) + i\int_{r(0)}^{r(t)} dr\, p_c \right) \ , \tag{2.28}$$

$$\beta_k \propto \exp\left(-ikr(0) - i\int_{r(0)}^{r(t)} dr\, p_c \right) \ . \tag{2.29}$$

We will evaluate these expressions as close as possible to the horizon to minimize corrections to the WKB approximation. At the same time we should remain slightly outside the horizon for the mode functions $u_k^q(r)$ to be regular, so we take $r(t)$ to be just outside of the horizon, that is $r(t) = r_+(M+\omega, Q+q) + \epsilon$ with $\epsilon \ll r_+(M+\omega, Q+q)$. Before we can evaluate the Bogoliubov coefficients, we will need some explicit details of the trajectory of the shell towards which we will move our attention next.

Evaluating the Bogoliubov coefficients

The canonical momentum of an outgoing shell in our background is given by [63]

$$p_c(r(t)) = \sqrt{2Mr - Q^2} - \sqrt{2M_\pm r - Q_\pm^2} - r\log\left(\frac{r - \sqrt{2M_\pm r - Q_\pm^2}}{r - \sqrt{2Mr - Q^2}} \right) \ . \tag{2.30}$$

The upper sign of $M_\pm, Q_\pm$ denotes the saddle point for α_k and the lower sign the saddle point for β_k. Null geodesics can be found by introducing lightcone coordinates v and u as

$$\begin{aligned} u &= t - r_* = \text{constant} \ , \\ v &= t + r_* = \text{constant} \ , \end{aligned} \tag{2.31}$$

where we introduced the tortoise coordinate r_*.

$$r_* = \int dr \, \frac{1}{f(r)} = r + \frac{r_+^2}{r_+ - r_-} \log\left(r - r_+\right) - \frac{r_-^2}{r_+ - r_-} \log\left(r - r_-\right) \ . \tag{2.32}$$

It is now straightforward to obtain an explicit expression for t and k, given the initial position $r(0)$ of the shell. By setting $u(t, r(t)) = u(0, r(0))$ it follows that

$$t = r - r(0) + \frac{r_+^2}{r_+ - r_-} \log\left(\frac{r - r_+}{r(0) - r_+}\right) - \frac{r_-^2}{r_+ - r_-} \log\left(\frac{r - r_-}{r(0) - r_-}\right) . \tag{2.33}$$

Using the initial condition (2.23) we obtain an expression for k.

$$k = p_c(0, r(0)) . \tag{2.34}$$

The last details we need are the value of $r(0)$ and t at the saddle points. Because the modes are infinitely blueshifted near the horizon as compared to an asymptotic observer, we only require the solution in the limit $k \to \infty$. In this limit we can invert (2.34) to find an expression for $r(0)$ and plug this into (2.33) to obtain an expression for t. This then leads to

$$r(0) = \begin{cases} r_+(M_+, Q_+) + \mathcal{O}(e^{-k/r_+}) & (\text{for } \alpha_{kk'}) \\ \\ r_+(M_-, Q_-) - \mathcal{O}(e^{-k/r_+}) & (\text{for } \beta_{kk'}) \end{cases} \tag{2.35}$$

$$\mathrm{Im}\,(t) = \begin{cases} 0 & (\text{for } \alpha_{kk'}) \\ \\ -\pi \frac{r_+(M_-, Q_-)^2}{r_+(M_-,Q_-) - r_-(M_-,Q_-)} & (\text{for } \beta_{kk'}) . \end{cases} \tag{2.36}$$

It is important to notice that generically, in the parameter space of the shell spanned by q and ω, $r(0)$ lies outside of the black hole horizon for α_k. As a consequence, the action at the saddle point of α_k is completely real, and describes a classically allowed trajectory. On the other hand, the initial position of the shell is inside the horizon for β_k. This implies that the shell travels on a classically forbidden trajectory and the action picks up an imaginary part. Since the quantity of interest is the ratio of the absolute values of the Bogoliubov coefficients, implying that only the imaginary parts of the action contribute, we arrive at the following result

$$\frac{|\beta_k|^2}{|\alpha_k|^2} \propto \exp\left[2\,\mathrm{Im}\left(\int_{r(0)}^{r(t)} dr\, p_c\right)\right] , \tag{2.37}$$

which is evaluated at the saddlepoint for β_k and where $r(t)$ is taken as close as possible, but slightly outside the horizon. Details on how to evaluate this integral and the correct pole prescription can be found in appendix 2.A. Here, we simply quote the result:

$$\mathrm{Im}\left(\int_{r(0)}^{r(t)} dr\ p_c\right) = -\pi \int_{r_+(M_-,Q_-)}^{r_+(M,Q)} dr\, r = \frac{1}{2}\pi(r_+(M_-, Q_-)^2 - r_+(M, Q)^2) . \tag{2.38}$$

This expression equals half the difference of the black hole entropy before and after emission, which is typically negative and reduces to the Boltzmann factor in the limit where the backreaction can be neglected. We therefore conclude that the probability of the black hole to emit a particle with charge q and energy ω, assuming ΔS_{BH} is negative, equals

$$P(k) \propto \frac{|\beta_k|^2}{|\alpha_k|^2} \propto e^{\pi(r_+(M_-,Q_-)^2 - r_+(M,Q)^2)} = e^{\Delta S_{BH}} , \tag{2.39}$$

where $\Delta S_{BH} = S_{BH}(M-\omega, Q-q) - S_{BH}(M,Q)$.

For both the neutral and charged case this does not exactly reproduce the result of the original articles [62, 63] due to a technical error, which was in fact corrected in [67] for the neutral case. Here we extended it to also include charged emission. The fact that the shell (generically) follows a classically forbidden trajectory clearly suggests that we should be able to reproduce the result (2.39) directly by doing a tunneling calculation, an idea that was worked out in [64]. In the next subsection we demonstrate that this approach is indeed equivalent and verify explicitly that the same Hamiltonian formalism used in [62, 63] underlies this computation.

2.2.2 The tunneling perspective

In the previous section we saw that the computation of the Bogoliubov coefficients reduced to calculating the imaginary part of the classical action, which is what one would compute in a tunneling calculation in a WKB approximation. This was done in [64] for the emission of neutral massless radiation. Here we generalize their calculation to charged emission and make clear that this method is equivalent to, and can be derived from, the Kraus-Wilczek effective action approach.

As mentioned, an important difference between the two approaches is that the former keeps the black hole mass fixed and allows the ADM mass to vary, while the latter keeps the ADM mass fixed and allows the black hole mass to vary. Before we continue to discuss the tunneling calculation of Parikh and Wilczek, we first discuss some details of the effective action computation if one would have fixed the ADM mass. In that case the mass parameter $\mathcal{M}(r)$ for a shell at position $\hat{r}$ is given by

$$\mathcal{M}(r) = \begin{cases} M - \omega - \frac{(Q-q)^2}{2r} & (r < \hat{r}) , \\ \\ M - \frac{Q^2}{2r} & (r > \hat{r}) . \end{cases} \tag{2.40}$$

Since we fixed the ADM mass M, it is now the geometry inside the shell that determines the evolution. Again, we could use Hamilton's equations to obtain the result that the shell travels on a null geodesic with mass parameter $\mathcal{M}(r < \hat{r})$ as the Hamiltonian of the shell is still given by $M_{ADM} - (M - \omega) = \omega$. Another difference is related to the initial condition of the shell which is now given by

$$r(0) = \begin{cases} r_+(M,Q) + \mathcal{O}(e^{-k/r_+}) \quad (\text{for } \alpha_k) \ , \\ \\ r_+(M,Q) - \mathcal{O}(e^{-k/r_+}) \quad (\text{for } \beta_k) \ . \end{cases} \tag{2.41}$$

So for α_k the shell now starts just outside of, and for β_k just inside of the *initial* horizon. It is important to notice that the parameter M that now appears in all relevant expressions is the ADM mass and not the black hole mass. At the end of the day, we want to interpret the probability for shell emission in terms of the change in entropy of the black hole. Hence, we should write the canonical momentum in terms of $M_{BH} = M_{ADM} - \omega$. After this simple shift, the calculation becomes equivalent to the Kraus-Wilczek computation with the black hole mass fixed. We therefore conclude that when fixing the ADM mass the Kraus-Wilczek effective action approach also leads to the same result (2.39).

In addition, this result can now be compared directly with the tunneling method of [64]. The starting point of Parikh and Wilczek is the fact that in a WKB approximation the tunneling probability is given by the exponential of the classical action, which reduces to the integral

$$P \propto \exp\left[-2 \ \mathrm{Im}\left(\int_{r_i}^{r_f} dr \ p_c\right)\right] \ , \tag{2.42}$$

where r_i and r_f correspond to the initial and final position respectively of the particle that is tunneling through a potential barrier. Based on the previous section, we recognize it as the final expression for the integral in the Kraus-Wilczek approach. In fact, it can be directly related to the expression (2.37) obtained by fixing M_{ADM} and using the boundary conditions (2.41). The relative minus sign between these expressions can be explained by the fact that in the tunneling integral (2.42) $r_f < r_i$, since the shell is taken to tunnel from just inside the initial horizon to just outside the final horizon. In contrast, in the Kraus-Wilczek computation $r(0)$ is always smaller than $r(t)$, when expressed in terms of the black hole mass. Because these expressions are written in terms of the canonical momentum and do not (explicitly) depend on the details of the background, we expect this result to remain universally valid as long as spherical symmetry is imposed.

Making use of Hamilton's equations, we can manipulate (2.42) to write it as

$$\mathrm{Im}\left(\int_{r_i}^{r_f} dr \, p_c\right) = \mathrm{Im}\left(\int_{r_i}^{r_f} dr \int_{H(0)}^{H(\omega)} dH \, \frac{1}{\dot{r}}\right) \ . \tag{2.43}$$

Here H is the Hamiltonian of the geometry seen by the shell. Since the shell follows a null geodesic in a geometry with mass $M-\omega$ and charge $Q-q$, the Hamiltonian is identified as the mass of the black hole, such that $dH = -d\omega$. The equation of motion for the outgoing positive energy shell in Painlevé-Gullstrand coordinates is given by

$$\dot{r} = 1 - \sqrt{\frac{2(M-\omega)}{r} - \frac{(Q-q)^2}{r^2}} \; . \tag{2.44}$$

The boundaries of the integral are taken such that we integrate the shell from just inside the initial horizon to just outside the final horizon. We now find that the integral of (2.43) contains a pole, determined by the position of the outer horizon, that is $r_+(M-\omega, Q-q)$. In order to evaluate this integral, we need a prescription that tells us how to deform the contour around the pole. Different choices correspond to different boundary conditions. We show in appendix 2.A that the prescription that supplies the (physically) correct boundary conditions is given by the (Feynman) deformation $\omega \to \omega - i\epsilon$, which was also used in [64].

Evaluating the integral using the prescribed contour deformation and taking the boundaries as the position of the initial and final horizon, one arrives at

$$\mathrm{Im}\left(\int_{r_i}^{r_f} dr\ p_c\right) = -\pi \int_{r_+(M,Q)}^{r_+(M_-,Q_-)} dr\ r = -\frac{1}{2}\pi\left(r_+(M_-,Q_-)^2 - r_+(M,Q)^2\right) \; , \tag{2.45}$$

where we used the results of appendix 2.A. The fact that these are the correct boundaries to take can be seen by switching the order of integration, which leads to the same result [64]. So we see that the tunneling method indeed gives the same result, as it should, for generic parameters ω and q implying the following universal decay probability (for sufficiently large energies ω)

$$P(k) \propto e^{\Delta S_{BH}} \; , \tag{2.46}$$

where we (again) ignored the appropriate normalization factor, which for large enough negative values of ΔS_{BH} is approximately one. This universal expression can now be employed to study different physical scenarios. In [64], where neutral radiation was considered, the result was used to identify the (leading) correction to Hawking radiation, capturing a deviation from perfectly thermal behavior, but consistent with an interpretation in terms of statistical thermodynamics. We will instead use this generalized expression to study charged decay channels in certain limits of parameter space that are of interest to us and where the inclusion of backreaction is crucial. We will in particular be considering limits where the emission of charged quanta does not (only) occur through a thermal Hawking process, but is dominated by a charged Schwinger-like process.

2.2.3 Superradiant emission and the tunneling integral

A particular limit of interest is that of low-energy charged emission, which is well known to display superradiant behavior. Indeed, in the regime of parameters where one expects superradiance the entropy difference ΔS_{BH} becomes positive, implying that the standard interpretation in terms of a tunneling probability is invalid. The appearance of a superradiant regime in the (incorrect) expression for the emission probability including backreaction was noticed in [63], but not elaborated upon. Here we will provide the appropriate interpretation and application of the tunneling integral in the low-energy superradiant regime.

To remind the reader, usually superradiance is associated to (and described by) a scattering process, and as a consequence one introduces transmission and reflection coefficients instead of Bogoliubov coefficients. When scattering an incoming wave $v_{1,k}(t, r)$ on the horizon, conservation of flux implies

$$v_{1,k} + R\ v_{2,k} = T\ v_{3,k}\ , \tag{2.47}$$

where v_1 and v_2 are respectively the right-moving and left-moving wave functions inside the horizon and v_3 is the right-moving wave outside the horizon. The reflection and transmission coefficients R and T are normalized as

$$|R|^2 + |T|^2 = 1\ . \tag{2.48}$$

It is then straightforward to show that the transmission and reflection coefficients can be related to the Bogoliubov coefficients in the following way [74]

$$|\alpha_k|^2 = 1/|R|^2\ , \qquad \frac{|\beta_k|^2}{|\alpha_k|^2} = |T|^2\ , \tag{2.49}$$

where the standard normalization condition for the Bogoliubov coefficients has been assumed

$$|\alpha_k|^2 - |\beta_k|^2 = 1\ . \tag{2.50}$$

We conclude that the transmission coefficient T can be associated to the tunneling probability P, which as we have seen is expressed in terms of the entropy difference between the final and initial state of the black hole.

Obviously an interpretation in terms of a probability requires the ratio of Bogoliubov coefficients to be smaller than one. For charged emission there exists a parameter regime at low enough energies where the sign of ΔS_{BH} actually becomes positive. For charged decay channels obeying

$$\omega + \frac{q^2}{2r_+} < q\frac{Q}{r_+}\ , \tag{2.51}$$

the change in entropy becomes positive and the tunneling integral is exponentially enhanced instead of suppressed. We recognize the left-hand side as the total energy of the shell (including its electromagnetic self-energy) and the right-hand side as the electromagnetic potential of the black hole that the shell couples to. Notably, in the extremal limit $M = Q$, the effect of backreaction is to shift the superradiant regime to lower (super-extremal $\omega < q$) values for the energy of the emitted particle. When (2.51) is satisfied this clearly signals a (thermodynamic) instability, as it becomes possible for the black hole to radiate away charge, while nevertheless increasing the entropy of the black hole. This superradiant instability was first discovered in the process of partial wave scattering off rotating black holes. For rotating black holes it is absent in the s-wave sector (and therefore more suppressed), but for charged black holes it remains present when restricting to the s-wave sector at low enough energies. In the appropriate superradiant scattering process, the normalization condition for reflection and transmission coefficients for a particle with frequency ν and charge q is affected in the following way [75].

$$|R|^2 = 1 - \frac{\nu - qQ/r_+}{\nu}|\mathcal{T}|^2 \tag{2.52}$$

So effectively this corresponds to the replacement

$$|T|^2 \to \frac{\nu - qQ/r_+}{\nu}|\mathcal{T}|^2 \ . \tag{2.53}$$

Comparing this to (2.48) we observe that when the frequency obeys the bound

$$\nu < q\frac{Q}{r_+} \ , \tag{2.54}$$

the reflection coefficient exceeds unity, that is $|R|^2 > 1$, meaning that the particle that scatters off the black hole takes away some of its mass and charge [75]. This frequency agrees with (2.51) in the limit $q/2Q \ll 1$, that is when ignoring back-reaction. However, instead of particles scattering off black holes, we would like to consider spontaneous emission in this superradiant regime of parameter space. One observes that in the superradiant regime apparently $|\alpha_k|^2 = 1/|R|^2 < 1$, suggesting that the Bogoliubov coefficients should be interchanged ($\alpha_k \leftrightarrow \beta_k$) to still obey the normalization condition. Equivalently, one can interpret this as a change in the sign of the normalization condition for the Bogoliubov coefficients. This can be traced back to the fact that what was previously defined to be a positive frequency mode at asymptotic infinity in the superradiant regime turns into a negative frequency mode, and vice versa. As a consequence, in the super-radiant regime the probability of emission $P(k)$ for a charged particle should be re-evaluated and is related in a more indirect way to the tunneling integral, as we will see below.

To determine the probability distribution we start with the appropriate expression for the average number of particles per mode in the superradiant regime [76]. This is most easily derived by applying a change of sign for the normalization of the Bogoliubov coefficients, that is what one means with positive and negative frequency modes. This results in the following expression for the expectation value of the number operator in the superradiant regime

$$\langle N_k \rangle = \frac{-1}{|\alpha_k|^2/|\beta_k|^2 - 1} \ . \tag{2.55}$$

The change of sign in the numerator ensures that the average number of particles remains positive in the superradiant regime where the ratio $|\alpha_k|^2/|\beta_k|^2 = e^{-\Delta S_{BH}(k)} < 1$. It is important to note that at the transition from superradiant to ordinary (Hawking) emission it is crucial to take into account the greybody factor $\Omega(\omega_k)$ to ensure appropriately continuous behavior, but for our purposes here we can safely ignore this issue. The relevant probability distribution for observing n particles in a mode k can be written in terms of the average number as follows

$$P_k(n) = \frac{\langle N_k \rangle^n}{(\langle N_k \rangle + 1)^{n+1}} \ . \tag{2.56}$$

As an easy check this indeed reproduces the standard (Bose-Einstein) distribution for single particle emission when ΔS_{BH} is negative. Using the superradiant expression for the average number of particles, one then arrives at the following probability for emitting a single particle in mode k

$$P(k)_{SR} = (1 - e^{-\Delta S_{BH}(k)}) \frac{1}{(2 - e^{-\Delta S_{BH}(k)})^2} \ , \tag{2.57}$$

where we expressed the probability explicitly in terms of the ratio $|\alpha_k|^2/|\beta_k|^2 = e^{-\Delta S_{BH}(k)} < 1$. This superradiant expression clearly differs from the standard (Bose-Einstein) distribution and generalizes the known result without backreaction [76]. Typically $\Delta S_{BH} > 0$ over a considerable range of superradiant frequencies and the probability distribution is very flat. As a consequence a charged black hole quickly radiates away its charge.

We conclude that in addition to the direct connection to the probability of emission in the high energy (charged) Hawking regime, the universal result for the tunneling integral also appears in the (modified) expression for the emission probability in the superradiant regime, which can be interpreted in terms of (generalized) Schwinger pair creation in the electric field near the horizon of the charged black hole. Indeed, for large black holes it was shown in [68] that the emission of charged quanta is dominated by Schwinger pair production, rather than the Hawking process, and allows charged black holes to quickly get rid of their charge. The derived probability distribution generalizes that result by taking into account the backreaction which, as before, can be expressed in terms of the change of the black hole entropy.

The superradiant regime and the inclusion of backreaction will play an important role in the next section. To be precise, so far we only considered non-extremal black holes for which the tunneling rate describes both thermal (neutral) radiation as well as charged (superradiance). Next, we will take the extremal limit of charged black holes. Since extremal black holes have a vanishing temperature, one expects the neutral (thermal) emission to shut down but charged decay channels should remain present. Clearly, to avoid creating a naked singularity only tunneling of (super-)extremal particles with $m \leq \omega \leq q$ is allowed. We will see that the tunneling calculation in the extremal limit not only confirms this expectation but in addition suggests the existence of a family of (non-extremal) gravitational instantons in the near-horizon $AdS_2 \times S^2$ limit, in which the superradiant regime is decoupled.

2.3 Extremal and near-horizon limits

We would now like to study the universal result for charged emission from a charged black hole in the extremal limit. As is well known, the temperature of an extremal black hole vanishes, which is reflected by the fact that the emission rate (2.46) for neutral particles becomes zero in the extremal limit $M = Q$. However, we are interested to see what happens to the charged decay channels in the extremal limit. After a careful examination and regularization of the extremal limit, we will conclude that those decay channels are still captured by the universal expression for the tunneling integral. Once that has been established we will consider the near-horizon limit and relate the tunneling decay rate to gravitational instantons describing the spontaneous nucleation of domain walls in AdS.

2.3.1 Charged particle decay in the extremal limit

A description in terms of particles tunneling out of an extremal black hole, for which the inner and outer horizon overlap, naively seems to be problematic due to the absence of a tunneling barrier lying in between the inner and outer horizon. The latter seems to be required to allow for a proper interpretation and related derivation of the final tunneling integral. One should be careful just extrapolating the final result, as it might be inconsistent and the different steps in the derivation need to be understood properly as one takes the extremal limit.

To regularize the extremal limit, we will introduce a non-extremality parameter $\epsilon \ll Q$ defined as

$$\begin{aligned} r_+ &= Q + \epsilon \ , \\ r_- &= Q - \epsilon \ , \end{aligned} \tag{2.58}$$

such that the metric becomes

$$\begin{aligned} ds^2 &= -f(r)dt^2 + f(r)^{-1}dr^2 + r^2 d\Omega_2^2 \ , \\ f(r) &= \frac{(Q-r)^2}{r^2} - \frac{\epsilon^2}{r^2} \ . \end{aligned} \tag{2.59}$$

The extremal limit is then defined as $\epsilon \to 0$. To study the region $r_- \leq r \leq r_+$ we follow [77] by introducing the coordinates

$$r = Q - \epsilon \cos(\chi) \, , \qquad t = \frac{Q^2}{\epsilon}\psi \ . \tag{2.60}$$

In this region χ is a spacelike and ψ a timelike coordinate. Using these coordinates the metric becomes

$$\begin{aligned} ds^2 &= Q^2 \left(-h(\chi)^2 d\chi^2 + \frac{\sin^2(\chi)}{h(\chi)^2} d\psi^2 + h(\chi)^2 d\Omega_2^2 \right) \ , \\ h(\chi) &= 1 - \frac{\epsilon}{Q}\cos(\chi) \ . \end{aligned} \tag{2.61}$$

The proper distance between r_+ and r_- is given by

$$\tau = Q \int_0^\pi d\chi \ h(\chi) = \pi Q \ , \tag{2.62}$$

which is independent of ϵ. We conclude just as [77], perhaps somewhat surprisingly, that even in the extremal limit $\epsilon \to 0$ there remains a finite proper distance between the inner and outer horizon.

Now we can continue as before and compute the tunneling integral for (super-)extremal shells from an extremal black hole. For an extremal shell we find (in Painlevé-Gullstrand coordinates)

$$\begin{aligned} ds^2 &= -f(r)dt_p^2 + 2\sqrt{1-f(r)}dt_p dr + dr^2 + r^2 d\Omega_2 \ , \\ f(r) &= \frac{(r-Q+q)^2}{r^2} - \frac{\epsilon^2}{r^2} \ . \end{aligned} \tag{2.63}$$

The final result can be calculated by taking into account the substitutions (2.58) and in the end sending $\epsilon \to 0$. The result for the tunneling integral is

$$\exp\left(-2 \ \mathrm{Im}\left(\int_{r_i}^{r_f} dr \ p_c\right)\right) = e^{\pi\left((Q-q)^2 - Q^2\right)} = e^{\Delta S_{BH}} \ , \tag{2.64}$$

in full agreement with the universal expression. Similarly, we could also consider emission of super-extremal shells from an extremal black hole by making the substitutions

$$M \to Q - \omega \ , \qquad (2.65)$$
$$Q \to Q - q \ ,$$

in the metric to describe a non-extremal black hole as the final state. On the other hand, if we were to consider sub-extremal shells, r_+ becomes imaginary after emission of the shell, which implies that the tunneling integral vanishes. Therefore, the emission of a sub-extremal particle (that would create a naked singularity) is forbidden.

We conclude that the same universal expression in terms of the black hole entropy difference still applies in the extremal limit. Although for an extremal black hole neutral emission shuts down, it can still decay via charged particles and the probability for that to happen can be expressed in terms of the (negative) entropy difference, as anticipated. If we consider the emission of super-extremal shells satisfying

$$\omega + \frac{q^2}{2Q} < q \ , \qquad (2.66)$$

we notice that the entropy difference becomes positive and therefore this process is governed by the superradiant expression for the probability that was derived previously. In contrast, we note that by including backreaction a parameter window opens up for shells satisfying

$$q\left(1 - \frac{q}{2Q}\right) < \omega \leq q \ , \qquad (2.67)$$

that can be described by a (suppressed) tunneling amplitude, instead of the (lower energy) regime of superradiant emission. From a near-horizon point of view one might anticipate that these decay channels can be understood in terms of an instanton. In fact, in the near-horizon limit of a four-dimensional extremal Reissner-Nordström black hole [69] derived the action for an instanton with charge equal to its tension connecting an initial $AdS_2 \times S^2$ spacetime with charge Q to two $AdS_2 \times S^2$ with charge Q_1 and Q_2 while keeping the total charge $Q = Q_1 + Q_2$ fixed. They related this to the instanton found by Brill [70] resulting in the following decay rate

$$P \sim e^{-2\pi Q_1 Q_2} \ , \qquad (2.68)$$

which coincides with (2.64) in the limit $q \ll Q$, that is to leading order in the backreaction. In this extremal case this is appropriately described as fragmentation, since the two different vacua coexist peacefully and the domain wall separating them is flat and static.

Similarly, super-extremal domain walls should be related to the emission of super-extremal shells for which $q(1-q/2Q) < \omega < q$. Such a shell necessarily expands due to its electromagnetic repulsion describing an instability of the extremal near-horizon AdS geometry. Starting from an extremal black hole, for all q and ω satisfying $M = Q \geq q \geq \omega > q(1-q/2Q)$ the entropy difference is negative describing an exponentially suppressed tunneling rate. In the near-horizon limit this should be related to a decay of AdS space through the creation and subsequent expansion of a (super)extremal domain wall. In the next section we will make this connection to domain walls in the near-horizon AdS geometry explicit by using the near-horizon relation between the AdS energy parameter U, which we will define in a moment, and the asymptotic Minkowski space energy parameter ω.

2.3.2 The near-horizon limit, domain walls and gravitational instantons

In order to relate the extremal black hole tunneling rate to a near-horizon AdS instanton one needs to introduce the relevant near-horizon energy parameter, instead of the asymptotic Minkowski energy parameter ω that we have used so far. To derive an expression for the local energy density of the shell, let us reconsider the situation where an extremal black hole with charge Q emits an extremal shell with charge q. Before emission, the metric is given by

$$\begin{aligned} ds^2 &= f(r)dt^2 + f(r)^{-2}dr^2 + r^2 d\Omega_2^2 \ , \\ f(r) &= \frac{(r-Q)^2}{r^2} \ , \end{aligned} \tag{2.69}$$

and after emission the charge of the solution is reduced to $Q-q$. In order for these two geometries to be consistently joined together by the shell we need to satisfy Israel's junction conditions [78]. We place the shell at some fixed position r and label coordinates on the shell by x^i. In the thin-wall approximation the condition we have to satisfy is (working in units where $G_N = 1$)

$$8\pi S^i_j = (\Delta K)\delta^i_j - \Delta K^i_j \ . \tag{2.70}$$

Here S^i_j is the surface energy-momentum tensor of the shell and ΔK_{ij} is the difference between extrinsic curvature on both sides of the shell. The energy density ρ of a shell is then given by

$$\rho = \frac{1}{8\pi}\left(\Delta K - \Delta K^t_t\right) = \frac{1}{4\pi r}\left(\sqrt{f_-(r)} - \sqrt{f_+(r)}\right) \ , \tag{2.71}$$

where $f_-(r)$ denotes the geometry with mass $M-\omega$ and charge $Q-q$ and $f_+(r)$ the geometry with M and Q. The extremal ($\omega = q$) shell has an energy density equal to

$$\rho_{ext} = \frac{q}{4\pi r^2} \ . \tag{2.72}$$

If the shell can be viewed as a domain wall in the near-horizon limit, this energy density should be equal to the tension of an extremal domain wall. An extremal domain wall separating two (supersymmetric) vacua with vacuum energy V_1 and V_2 has a tension T_{ext} that is given by [79]

$$8\pi T_{ext} = \frac{2}{\sqrt{3}} \left(\sqrt{|V_1|} - \sqrt{|V_2|} \right) \ , \tag{2.73}$$

where we take $|V_1| > |V_2|$. For two AdS spaces of charge $Q_1 = Q - q$ and $Q_2 = Q$ the vacuum energy is given by $|V_1| = 3/(Q-q)^2$ and $|V_2| = 3/Q^2$. Thus, the tension of an extremal domain wall separating these two vacua is

$$T_{ext} = \frac{q}{4\pi Q(Q-q)} = \frac{q}{4\pi Q^2} + \mathcal{O}\left(q^2/Q^2\right) \ , \tag{2.74}$$

where we assumed the probe limit $q \ll Q$. This indeed matches with the tension of an extremal shell, as given by (2.72) in the near horizon limit $r \to Q$, provided $q \ll Q$. This confirms that extremal particle shells can be interpreted as flat, extremal domain walls from the point of view of the near-horizon geometry.

Similarly, in the near-horizon limit super-extremal shells should correspond to super-extremal domain walls whose tension is bounded by $T < T_{ext}$. To make this correspondence explicit let us derive an expression for the near-horizon AdS energy

$$U = r^2 \int d\Omega_2 \ \rho \ , \tag{2.75}$$

which is a function of the asymptotic energy ω and the charges Q and q. Here, $d\Omega_2$ is the volume element of the unit 2-sphere. Integrating the local energy density ρ in the spherical shell, as given by (2.71), and taking the near-horizon limit one derives

$$U^2 = q^2 - 2Q(q-\omega) \ . \tag{2.76}$$

Note that for shells satisfying $\omega < q$ this similarly implies $U < q$. This confirms that $T(\omega < q) < T_{ext}$ by recognizing that the tension can be expressed as $T = U/4\pi Q^2$.

Several additional comments are in order regarding the domain wall energy (2.76). In the extremal limit $\omega = q$ one indeed finds, as should be expected, that this implies $U = q$ as well. Inverting this relation gives $\omega = q(1 - q/2Q) + U^2/2Q$ and as a consequence the near-horizon domain wall energy U vanishes when the asymptotic energy is equal to $\omega = q(1 - q/2Q)$, which exactly corresponds to the transition point where the entropy difference vanishes and the decay turns superradiant. We conclude that the near-horizon limit decouples this regime, in the sense that for all $U \geq 0$ the asymptotic energy ω is always in the regime where the decay channel is described by a suppressed tunneling amplitude in terms of the inherited entropy difference. We also note that the probe limit ($q/Q \ll 1$) necessarily implies $\omega \sim q$ and therefore is related to a near-extremal particle decay channel of the parent extremal black hole.

To summarize the above, we explicitly related a family of domain wall instabilities of the near-horizon geometry to decay channels of the parent black hole. This seems to be a realization of an old conjecture made by Brill. He suggested that there should be an instanton that describes a single extremal Reissner-Nordström black hole splitting into two or more extremal black holes that agrees with the Brill instanton in the interior throat region [70]. Work towards this goal was presented in [80], where an instanton was found describing the splitting of the throat region into two or more connected throat regions. According to that work the probability for that specific process is only *half* the entropy difference. Our results suggest that the tunneling integral corresponds to the Lorentzian continuation of Brill's conjectured instanton. In fact, by taking backreaction into account the gravitational instanton related to the Hawking modes of a non-extremal black hole was first discussed in [65], where they indeed found a decay rate equal to

$$P \sim e^{\Delta A_{BH}/4} = e^{\Delta S_{BH}} \,, \tag{2.77}$$

in full agreement with the tunneling result (for $\Delta S_{BH} < 0$). To extend their results to the extremal near-horizon limit we can regulate it as before using (2.58), at the end sending $\epsilon \to 0$ and introducing U as the relevant near-horizon energy parameter. For the special case where the extremal black hole emits an extremal shell, the instanton involved in this process should correspond to Brill's conjectured instanton.

To be precise, we will now show explicitly that the expression for an extremal black hole emitting an extremal shell indeed reduces to the Brill instanton in the near-horizon limit. To do so, we continue the metric used in the tunneling calculation to Euclidean signature. Before we do this, it should be noted that Brill considered a magnetically charged solution whereas we are interested in an electrically charged solution. In order to obtain a real-valued instanton action, the electrical charge also has to be appropriately continued as $Q \to iQ$ [81]. The Euclidean metric is then given by

$$\begin{aligned} ds^2 &= f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2 d\Omega_2^2 \\ f(r) &= \frac{(r-\mathcal{Q})^2}{r^2} \,. \end{aligned} \tag{2.78}$$

Here we defined

$$\mathcal{Q} = \begin{cases} Q & (r < \hat{r}) \ , \\ \\ Q - q & (r > \hat{r}) \, , \end{cases} \tag{2.79}$$

where $\hat{r}$ denotes the position of the extremal shell. Both the geometry inside and outside the shell are Reissner-Nordström geometries, with a charge of respectively $Q - q$ and Q. We can therefore take the near horizon limit either inside or outside the shell. This limit is defined by writing

$$r = \mathcal{Q} + \chi \ , \tag{2.80}$$

and expanding around $\chi = 0$. The metric then becomes

$$ds^2 = \frac{\chi^2}{\mathcal{Q}^2} dt^2 + \frac{\mathcal{Q}^2}{\chi^2} d\chi^2 + \mathcal{Q}^2 d\Omega_2^2 \, . \tag{2.81}$$

This can be rewritten in a form used by Brill

$$\begin{aligned} ds^2 &= H^2 dt^2 + H^{-2} \left(dx^2 + dy^2 + dz^2 \right) \ , \\ H &= \frac{\mathcal{Q}}{|\vec{x}|} \ , \end{aligned} \tag{2.82}$$

where $|\vec{x}|^2 = x^2 + y^2 + z^2$. The Lorentzian version of this geometry, known as the Bertotti-Robinson geometry, corresponds to $AdS_2 \times S^2$ and is an exact solution to the Einstein-Maxwell equations. More general, there also exist solutions with N charges Q_i for which

$$H = \sum_{i=1}^{N} \frac{Q_i}{|\vec{x} - \vec{x}_i|} \ , \tag{2.83}$$

that are interpreted as a set of static extremal black holes with charge Q_i placed at $\vec{x}_i$.

We will now write down a particular two-centered black hole solution that agrees asymptotically with (2.82) by writing

$$V = \frac{q}{|\vec{x} - \vec{x}_1|} + \frac{Q - q}{|\vec{x}|} \, . \tag{2.84}$$

As we will see, we can interpret this geometry as an extremal black hole placed at $\vec{x} = 0$ and our extremal shell placed at $\vec{x}_1$. It has the following asymptotic behavior

$$\lim_{x \to \infty} V = \frac{Q}{|\vec{x}|} = H(r > \hat{r}) \ , \qquad \lim_{x \to 0} V = \frac{Q - q}{|\vec{x}|} = H(r < \hat{r}) \ . \tag{2.85}$$

We see that this geometry, at least asymptotically, agrees with (2.82). Furthermore, we notice that (2.85) is precisely the Brill instanton where an $AdS_2 \times S^2$ space of charge Q splits into two $AdS_2 \times S^2$ spaces with charge $Q - q$ and q. This leads us to the conclusion that the extremal near-horizon limit of the tunneling instanton found by [65] for extremal shells is indeed the Brill instanton [70], as conjectured.

Whereas the Brill instanton describes the fragmentation of AdS spaces, corresponding to the emission of an extremal shell from an extremal black hole, the tunneling expression should apply far more generally. In particular, it predicts that there should exist an entire family of gravitational instantons labeled by ω and q that satisfy $q(1-q/2Q) < \omega \leq q$, for which the associated black hole entropy difference is always negative and therefore corresponds to a suppressed tunneling amplitude. From the perspective of the near-horizon limit these instantons describe the decay of AdS vacua through the creation of super-extremal (expanding) domain walls connecting different vacua. The decay probability of these instantons should be provided by the tunneling integral for a finite window of parameters up until the extremal case. Indeed, the low-energy superradiant regime is decoupled in the near-horizon limit, as it would correspond to an imaginary near-horizon domain wall energy U.

In fact, in the context of string theory some AdS instantons of this type were already constructed in [69]. There it was also observed that when the charge of a particle (0-brane) equals its tension, the Euclidean action of the corresponding instanton reduces to the value given by the Brill instanton. However, [69] only derived this relation in the limit where the associated energy density (charge) of one of the AdS spaces was small. Our results do not have such a restriction, as the tunneling integral is valid as long as $\omega \leq q \leq Q \leq M$. Nevertheless, in the limit where backreaction is small, our result should reduce to those of [69]. By writing the tunneling amplitude for an extremal black hole emitting a shell with $\omega \leq q$ in terms of the domain wall energy U we find

$$\Delta S_{BH} = -2\pi Q \left(q - \sqrt{q^2 - U^2}\right) + \mathcal{O}(U^2/Q^2) + \mathcal{O}(q^2/Q^2) \,, \tag{2.86}$$

which as we already concluded reduces to the Brill instanton for $U = q$ and matches exactly with the AdS_2 instanton found for $U < q$ in [69]. From the AdS_2 point of view, this super-extremal emission corresponds to Schwinger pair production [82]. Since the superradiant regime is decoupled in the near-horizon limit, the general decay rate including backreaction in the spherically symmetric sector, should just be given by $P \sim e^{\Delta S_{BH}}$, extending the result of [69] beyond the probe limit.

We thus find that the instabilities of (nonsupersymmetric) AdS space, as conjectured in an extension of the WGC in [60, 61], are related to the (charged) decay channels that satisfy $q(1-q/2Q) < \omega < q$ of the parent extremal black hole geometry. The resulting decay rate can be expressed, including backreaction, in terms of the associated entropy difference. This result is not restricted to extremal shells, for which it was already noticed in [69, 70], but extends to super-extremal shells.

Due to the universal nature of the decay rate, it would be of interest to understand better how these results are related to constraints on low-energy effective potentials in the context of the landscape, if such a connection exists at all. Because our results imply constraints on AdS vacua, it is natural to wonder if and how this generalizes to arbitrary effective potentials.

2.4 Conclusions

One of the original motivations to study backreaction corrections to the Hawking process was to potentially shed some light on how it could be consistent with unitary evolution of an underlying microscopic description. Treating the emitted particles as spherically symmetric shells and imposing energy conservation, Kraus and Wilczek derived an effective action for the shells and indeed found that the emission probability deviates from being exactly thermal [62, 63]. In a similar spirit, Parikh and Wilczek imposed energy conservation to include backreaction by understanding the Hawking process in terms of a (spherically symmetric) quantum mechanical tunneling process. Their universal result [64] in terms of the difference of the black hole entropy before and after emission nicely supports a statistical thermodynamical interpretation of the transition, as was pointed out in earlier work by [65]. One important conclusion of our results is that the effective action approach reduces exactly to the tunneling integral. As a consequence both approaches are equivalent and the final result can always be expressed in terms of the entropy difference. This strongly suggests that the result can also be applied to describe the (spherically symmetric) decay of higher-dimensional black holes and/or black branes.

As our prime example of interest we then derived the probability for emission of a charged shell from a charged black hole in terms of the exponential of the entropy difference of the black hole before and after emission. We then clarified the interpretation of the result in a low-energy regime of charged emission where the entropy difference changes sign and becomes positive. In this superradiant regime the probability for emission, including backreaction, has to be reassessed and we derived an expression that again features the entropy difference and reduces to the known expression for superradiant emission in the absence of backreaction. We then studied the extremal limit, showing that the backreacted result for charged (super-extremal) particle emission remains valid. The absence of decay channels for which $\omega \geq m > q$ is consistent with the weak cosmic censorship conjecture, and assuming the existence of super-extremal particles in the spectrum, as conjectured by the Weak Gravity Conjecture (WGC), the extremal black hole will decay. Another noteworthy result is that in the extremal black hole limit the inclusion of backreaction (and charge conservation) implies that the threshold energy below which superradiant behavior kicks in is distinguishably lower than the particle's charge q, opening a window of suppressed tunneling in the super-extremal emission regime.

Having understood the extremal limit, corresponding to a near-horizon geometry of $AdS_2 \times S^2$, we then focused our attention on inherited decay channels of (non-supersymmetric) AdS spacetimes by identifying the relevant near-horizon AdS energy. In the near-horizon limit a positive domain wall energy will always be in the suppressed (negative entropy difference) regime. Recently, the WGC conjecture was extended in [60, 61] by suggesting that the bound is only saturated for BPS states in a supersymmetric theory. This would imply that all nonsupersymmetric AdS spaces are unstable and will decay. In particular, [60] motivated their conjecture by arguing that the WGC bounds the tension of (super-)extremal domain walls and therefore controls the stability of AdS vacua. Extending the result for charged emission from charged black holes to the extremal near-horizon region, we indeed confirm that domain walls satisfying the WGC constraint will be spontaneously produced resulting in the decay of the AdS geometry. The associated probability for this process is given by the universal expression in terms of the entropy difference of the parent black hole. Indeed, in the probe limit the tunneling amplitude exactly reproduces known results for super-extremal and extremal AdS instantons.

Strictly speaking our results only apply to AdS_2, but the universal form of the decay rate and its natural interpretation in terms of statistical thermodynamics suggests it applies equally well to higher-dimensional AdS spacetimes, providing a very general and precise expression for the decay rate of AdS through (super-)extremal domain walls, beyond the probe limit. It would of course be of interest to investigate this in more detail. In fact, higher-dimensional analogues of the Brill instanton that describe the fragmentation of higher-dimensional AdS spaces claim to have been constructed in [81], seemingly at odds with general expectations from the AdS/CFT correspondence. We find our results also particularly intriguing in light of the work of [83]. These authors claim that the instabilities they found for higher dimensional AdS spaces are higher-dimensional analogues of an instability discovered by Aretakis [84, 85], which seems to be closely related to the onset of superradiance [86]. In our approach, the domain walls associated to superradiant decay of the parent black hole would have imaginary tension. Furthermore, in [83] it is also argued that from the perspective of AdS the instability is actually perturbative, signalled by open string modes becoming tachyonic. It would certainly be interesting to explore the relation between our semi-classical results and their top-down constructions further.

Let us finally emphasize once more that our results are expected to be universally valid in the spherically symmetric sector and can be applied whenever (super-)extremal particles are present in the low-energy effective action. As such, assuming the WGC holds, it should accurately describe the instabilities of charged black holes, as well as those of the corresponding near-horizon AdS space in the extremal limit. Our findings therefore support the conjecture that all nonsupersymmetric AdS spaces are unstable and belong to the swampland, that is they cannot be consistently coupled to quantum gravity. What would be interesting to investigate in future work is if and how these results, derived from black hole physics, can be related to specific (constraints on) potentials in low-energy effective descriptions of AdS spaces. If this is possible, it is also natural to think about possible generalizations to arbitrary potentials, such as the ones that have recently been proposed for metastable de Sitter space in the context of the swampland program [87, 88].

2.A Pole prescription and relevant integrals

To calculate the rate of particles emitted by a charged black hole we need a pole prescription to evaluate the integrals (2.38) and (2.45). In this appendix we show how the correct prescription is determined by the physical process under consideration and calculate the relevant integrals.

Pole prescription

How to deal with the poles in the integrals we encountered can be understood by viewing the classical action as a propagator $K(x, x')$, which can be written as [89]

$$K(x,x') = \sum_{paths} e^{iS(x,x')} \ , \tag{2.87}$$

where $S(x, x')$ is the classical action connecting the points x and x'. Alternatively, the same propagator can also be viewed as a Green's function for the Klein-Gordon equation of a scalar field with mass m.

$$(\nabla_\mu \nabla^\mu - m^2) K(x,x') = -\delta(x,x') \ . \tag{2.88}$$

Focussing on flat space for now, we write the propagator for a massless field in momentum space and obtain the well-known Feynman propagator

$$K_F(x,x') = \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 - i\epsilon} e^{ik(x-x')} \ , \tag{2.89}$$

which corresponds to deforming the contour as

$$\begin{aligned} k^0 &= \omega_k - i\epsilon \qquad (\omega_k > 0) \ , \\ k^0 &= \omega_k + i\epsilon \qquad (\omega_k < 0) \ . \end{aligned} \tag{2.90}$$

We can evaluate (2.89) by making use of a contour integral. If $t - t' > 0$ we have to close the contour in the lower half-plane in order to be able to apply Jordan's lemma. This choice picks up the positive energy pole. We then obtain the well-known expression

$$K_F(x,x') = -\frac{i}{4\pi^2} \frac{1}{s(x,x') + i\epsilon} \ , \tag{2.91}$$

with $s(x,x')$ the square of the geodesic distance between x and x'. Similarly, if $t - t' < 0$ we have to close the contour in the upper half-plane which picks up the negative energy pole. We see that future directed propagation of positive energy particles corresponds to deforming the contour $\omega_k \to \omega_k - i\epsilon$ in momentum space and $t \to t - i\epsilon$ in position space.

Now we can use the results of [89] to obtain the analogous prescription for the Reissner-Nordström background. Also in this case, it was found that a future directed null geodesic corresponds to a deformation of the contour in the lower half t-plane, which in momentum space is equivalent to $\omega_k \to \omega_k - i\epsilon$, just as in flat space. We conclude that future propagation of positive energy particles requires $\omega_k \to \omega_k - i\epsilon$ and past propagation of negative energy particles $\omega_k \to \omega_k + i\epsilon$. This is also the prescription used in [64].

Parikh-Wilczek integral

The integral of interest is

$$S = -\int_{r_i}^{r_f} dr \int_0^{\omega} d\omega' \ \frac{1}{1 - \sqrt{2(M-\omega')/r - (Q-q)^2/r^2} - i\epsilon} \ , \tag{2.92}$$

where we used the prescription $\omega \to \omega - i\epsilon$. To calculate this integral, we first substitute $u = \sqrt{2(M-\omega)/r - (Q-q)^2/r^2} - 1$ to find

$$S = -\int_{r_i}^{r_f} dr \ r \int_{u(0)}^{u(\omega)} du \ \frac{u+1}{u+i\epsilon} = \int_{r_i}^{r_f} dr \ r \int_{u(\omega)}^{u(0)} du \ \frac{u+1}{u+i\epsilon} \ , \tag{2.93}$$

where in the right-hand side of the second equality we switched the boundaries of the integral because $u(0) > u(\omega)$. This integral has a pole at $u = -i\epsilon$. To evaluate it, we split it in terms of a principle value integral and a small (non-closed) contour that runs clockwise around the pole in the upper half-plane, as determined by the pole prescription. This contour is displayed in Figure 2.1. The

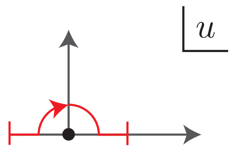

Figure 2.1: *Contour of the integral* (2.93), *which has a pole at* $u = -i\epsilon$.

integral now becomes

$$\begin{aligned}\lim_{\epsilon\to 0^+}\int_{u(\omega)}^{u(0)} du\ \frac{u+1}{u+i\epsilon} &= \mathcal{P}\int_{u(\omega)}^{u(0)} du\ \frac{u+1}{u} - i\pi\int_{u(\omega)}^{u(0)} du\ (u+1)\delta(u)\ , \\ &= \mathcal{P}\int_{u(\omega)}^{u(0)} du\ \frac{u+1}{u} - i\pi\ . \end{aligned} \tag{2.94}$$

Here, $\mathcal{P}$ denotes the principle value of the integral. It is straightforward to check that the principle value integral does not contribute an imaginary piece, and therefore we find

$$\mathrm{Im}\,(S) = -\pi\int_{r_i}^{r_f} dr\ r = -\frac{\pi}{2}(r_f^2 - r_i^2)\ . \tag{2.95}$$

Kraus-Wilczek integral

The second integral we need to evaluate appeared in section 2.2.

$$I = \int_{r(0)}^{r(t)} dr\ p_c\ , \tag{2.96}$$

with p_c given in (2.30) and the boundaries by $r(0) = r_+(M-\omega, Q-q) - \epsilon$ and $r(t) = r_+(M,Q)+\epsilon$. Notice that to evaluate this integral, we could use Hamilton's equations to rewrite this integral in the form of (2.92) to which we know the answer. Instead, for completeness and comparison with other references we will use the explicit expression of p_c.

Because we are only interested in the imaginary part of this integral, we can focus on the logarithmic piece in p_c, since only this term can contribute an imaginary piece. Thus, the integral of interest is

$$I = -\int_{r(0)}^{r(t)} dr\ r \log\left(\frac{r-\sqrt{2r(M-\omega)-(Q-q)^2}-i\epsilon}{r-\sqrt{2Mr-Q^2}}\right) . \tag{2.97}$$

This integral has branch points of the logarithm at

$$r_1 \equiv M-\omega+\sqrt{(M-\omega)^2-(Q-q)^2}+i\epsilon\ , \tag{2.98}$$

and

$$r_2 \equiv M+\sqrt{M^2-Q^2}\ . \tag{2.99}$$

Moreover, the argument of the logarithm has additional branch cuts at $r_3 \equiv (Q-q)^2/2(M-\omega)$ and $r_4 \equiv Q^2/2M$. We choose the following branch cut structure.

$$\begin{aligned} r_1 &: (-\infty, r_1]\ , \qquad & r_3 &: (-\infty, r_3]\ , \\ r_2 &: (-\infty, r_2]\ , \qquad & r_4 &: (-\infty, r_4]\ . \end{aligned} \tag{2.100}$$

To evaluate I, we will split the integral in different parts.

$$I = \lim_{\epsilon\to 0^+}\left(\int_{C_L}+\int_{C_\epsilon^1}+\int_{C_1}+\int_{C_\epsilon^2}+\int_{C_R}\right) dr\ h(r)\ , \tag{2.101}$$

where

$$h(r) = r\log\left(\frac{r-\sqrt{2r(M-\omega)-(Q-q)^2}-i\epsilon}{r-\sqrt{2Mr-Q^2}}\right) . \tag{2.102}$$

The structure of this integral in the complex plane is displayed in Figure 2.2. When we approach the real axis between $r_1 < r < r_2$ from below, as required by

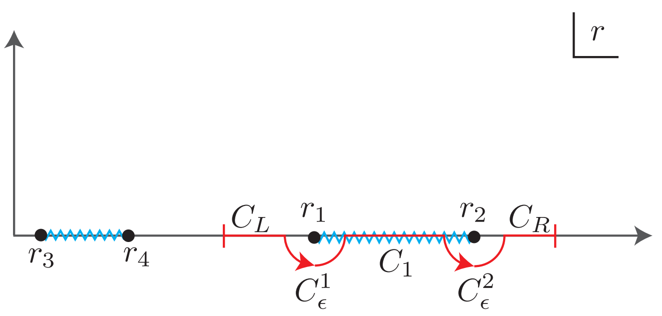

Figure 2.2: *Branch cut structure of the integral* (2.101)*. This integral has two branch cuts that connect the points r_1 to r_2 and r_3 to r_4. The piece over C_ϵ^1 and C_ϵ^2 vanishes in the limit $\epsilon \to 0$ and the parts over C_L and C_R are completely real.*

the pole prescription, the logarithm takes the following form.

$$\log\left(\frac{r-\sqrt{2r(M-\omega)-(Q-q)^2}}{r-\sqrt{2Mr-Q^2}}\right)=\log\left|\frac{r-\sqrt{2r(M-\omega)-(Q-q)^2}}{r-\sqrt{2Mr-Q^2}}\right|+i\pi \tag{2.103}$$

On the other hand, the imaginary part of the logarithm is zero on C_L and C_R. Because we are only interested in the imaginary part of I, we can ignore these pieces. Finally, by observing that the integrals over C_ϵ^1 and C_ϵ^2 vanish in the limit $\epsilon \to 0$, we obtain

$$\mathrm{Im}\,(I) = -\pi \int_{r_+(M_-,Q_-)}^{r_+(M,Q)} dr\ r = -\frac{\pi}{2}\left(r_+(M,Q)^2 - r_+(M_-,Q_-)^2\right)\ . \tag{2.104}$$

3 De Sitter Space in the Unruh State

In the previous chapter, we learned that backreaction effects in semi-classical gravity can lead to instabilities of black holes and anti-de Sitter space. In this chapter, which is based on [2], we study the (in)stability of de Sitter space in different quantum states. In particular, we construct the analogue of the Unruh state for black holes and compute the energy-momentum tensor in this state. One of our main observations is that the energy-momentum tensor carries a negative energy density at the horizon. As a result, pure de Sitter space is no longer a consistent solution when backreaction effects are included. We give a preliminary estimate of the effect of backreaction which suggests that the horizon area decreases, just as for black holes in the Unruh state. This instability puts a fundamental bound on the lifetime of de Sitter space in the Unruh state, set by its initial entropy. We point out that the Unruh-de Sitter state may be a natural initial state for patches of spacetime locally described by de Sitter space and comment on the possible cosmological implications.

3.1 Introduction

There have been contradictory claims about the quantum stability of de Sitter space.[1] In much of the literature, a common and very reasonable approach is to start out with perturbative quantum fluctuations in the Bunch-Davies state. Since that state is invariant under the de Sitter isometry group, the question boils down to whether or not (and how) the de Sitter symmetry becomes anomalous. Over the years different authors, using a varying range of techniques and methods, have reached different conclusions on the ultimate fate of de Sitter space. The results range from exact all-order stability [92] to (slow or fast) decay to flat space [93–96] and also evolution towards a phase where perturbative quantum field theory techniques break down [97]. The different approaches, whose consistency

[1] We refer the reader to [90, 91] for good reviews on de Sitter physics.

and interpretation is not always transparent, make it difficult to compare results and reach consensus on this important question. This is particularly confusing in light of the fact that one would expect to be able to make unambiguous statements, at least in perturbation theory, as long as the de Sitter curvature is small, since an effective field theory description should be sufficiently accurate then.

Here we will take yet another approach, motivated by lessons learned from the evaporation of black holes. Recall why it is that we are convinced that realistic black holes are quantum mechanically unstable. The emission of Hawking radiation from the horizon into the asymptotically flat geometry, where energy is well-defined and conserved, is not quite enough to conclude that a black hole has to evaporate: it is also necessary that there not be any special boundary conditions corresponding to compensating incoming energy flux. Indeed, for black holes there are two special quantum states that contain radiation: the Unruh and the Hartle-Hawking state. Although an asymptotic observer measures thermal radiation in both of these states, it is only in the Unruh state that the black hole decays. In the Hartle-Hawking state, the outgoing flux of energy is balanced by incoming flux from infinity and the maximally extended black hole geometry is in fact stable; this state is therefore typically used to describe the eternal black hole in exact thermal equilibrium, rather than the evaporating black hole, and the Hartle-Hawking state also preserves the symmetries of the original extended Schwarzschild geometry. In contrast, in the Unruh state there is an asymmetry between incoming and outgoing fluxes. The boundary conditions are such that there is no incoming flux from infinity and the outgoing flux is no longer compensated for. Therefore, the black hole is no longer in thermal equilibrium and evaporates. Whereas the vacuum expectation value of the energy-momentum tensor is regular everywhere in the Hartle-Hawking state, in the Unruh state it features a singularity on the past horizon. Despite this, the Unruh state is accepted as a physically viable state because it is regular on the future horizon, and the putative singularity on the past horizon is occluded by the collapsing matter forming the black hole.

We will translate these observations to the context of de Sitter space. The de Sitter counterpart of the Hartle-Hawking state is the Bunch-Davies state, a state annihilated by the full $O(1,4)$ group of four-dimensional de Sitter isometries. Here we propose that one can also consider a de Sitter analog of the Unruh state. As in the black hole case, this state imposes different boundary conditions for incoming and outgoing fluxes, breaking homogeneity by identifying a special static patch region. Correspondingly, the vacuum expectation value of the energy-momentum tensor is regular on the future horizon, but diverges on the past horizon. Nevertheless, as we will argue this state is consistent and even reasonable from a certain point of view.[2] Therefore, the first goal of this chapter is to explicitly construct the Unruh-de Sitter state.

Next, we would like to calculate the vacuum expectation value of the energy-momentum tensor in this state in order to examine the singularity structure of the vacuum and determine the backreaction on the geometry in semi-classical gravity. The absence of an asymptotically flat region in de Sitter space means that one cannot rely on global energy conservation arguments; instead, a careful calculation of the expectation value of the energy-momentum tensor is required to analyze the stability of the background. For the Bunch-Davies state, it is well known that the vacuum expectation value of the energy-momentum tensor is proportional to the metric; as it should if it preserves the de Sitter isometries. This can only renormalize the bare cosmological constant and does not destabilize de Sitter space. To study the stability of the Unruh-de Sitter state, we first need to calculate the expectation value of the energy-momentum tensor in the Unruh-de Sitter state. Unfortunately, as for black holes, directly calculating the expectation value of the energy-momentum tensor in the Unruh-de Sitter state is far from straightforward [99]. We therefore focus on 1+1 dimensional de Sitter space, where the calculation is more tractable and can be nicely deconstructed and interpreted for the different states. Under certain assumptions, which we spell out, the 1+1 results can then be generalized to the s-wave sector of 3+1 dimensional de Sitter space.

We find, as for black holes, that de Sitter space in the Unruh state is unstable. Even though the expectation value of the energy-momentum tensor in the Unruh state is inhomogeneous, it is a small correction and its averaged initial effect on a static patch region can be approximated by an analysis on the level of the Friedmann-Lemaître-Robertson-Walker (FLRW) equations. Analogous to an evaporating black hole, we find that the direction of the instability in the Unruh state is not towards Minkowski space but rather towards a singular geometry whose description lies outside the semi-classical regime. This evolution therefore corresponds to a quantum violation of the null energy condition (as occurs also for evaporating black holes). Again in parallel with the black hole case, the time scale of the instability is set by the gravitational entropy, that is $H\,t \sim M_p^2/H^2$. As such, the instability cannot explain the smallness of the cosmological constant[3] (being both too slow and in the wrong direction) but when applied in the context of eternal inflation it might prevent the anthropically required level of population of the landscape of string vacua. We will also comment on the relation to the recent conjectures about de Sitter space [87, 88] in the context of the swampland

[2]Earlier results supporting this point of view were described in [98], but did not explicitly involve the Unruh-de Sitter state.

[3]See [100] for a nice review.

program.

This chapter is organized as follows. In section 3.2, we recall some elementary results about quantum fields on Rindler and Schwarzschild black hole backgrounds. In particular, we review the different choices of state one can make. In section 3.3, we translate the problem to de Sitter space, defining the Unruh-de Sitter state and calculating the corresponding expectation value of the energy-momentum tensor. In section 3.4, we use the expectation value to analyze the instability of de Sitter space in the Unruh state. We give a bound on the maximal number of e-folds before the semi-classical description breaks down. For most of the paper we emphasize the mathematical consistency of the Unruh-de Sitter state, but here we will also argue that in many circumstances for which de Sitter space is used to approximate primordial or late-time cosmology, selecting the Unruh-de Sitter state might be a natural choice. Finally, in section 3.5, we consider the potential implications of our results for cosmological inflation and the string landscape.

3.2 States in Rindler and Schwarzschild geometries

To illustrate our key point regarding different states, their singularities and back-reaction, let us briefly recap some of the properties of Rindler space we saw in Chapter 1. Consider a massless minimally coupled scalar field in two-dimensional Minkowski space. To quantize the scalar field we expanded it in terms of incoming and outgoing modes that were either positive frequency with respect to Minkowski time t or Rindler time τ. There are now four special cases of vacuum states because we can select either the $\hat{a}$ vacuum (positive frequency with respect to t) or the $\hat{b}$ vacuum (positive frequency with respect to τ) for the incoming and outgoing modes independently. When both set of modes are chosen to be in the $\hat{a}$ vacuum, we call that the Poincaré-invariant vacuum, $|0_M\rangle$. When they are both chosen to be in the $\hat{b}$ vacuum, we call that the Rindler vacuum, $|0_R\rangle$. However, as we saw, the vacuum expectation value of the energy-momentum tensor in the Rindler vacuum is singular on the future ($U = 0$) and past ($V = 0$) Rindler horizons, which makes the Rindler vacuum unphysical.

The vacuum expectation value of the energy-momentum tensor in the Minkowski and Rindler vacuum is given by (see (1.29))

$$\begin{aligned} \langle 0_R|T_{UU}|0_R\rangle &= -\frac{1}{48\pi U^2} \ , \qquad \langle 0_M|T_{UU}|0_M\rangle = 0 \ , \\ \langle 0_R|T_{VV}|0_R\rangle &= -\frac{1}{48\pi V^2} \ , \qquad \langle 0_M|T_{VV}|0_M\rangle = 0 \ . \end{aligned} \tag{3.1}$$

We would like to emphasize that, as compared to the Minkowski vacuum, the Rindler vacuum has a negative energy density and the energy-momentum tensor is singular at the past ($V = 0$) and future ($U = 0$) acceleration horizons of the Rindler observer. Note also that in principle one could make a hybrid, asymmetric, choice to allow for states that are only singular at one of the horizons.

An entirely similar story applies for quantum fields in the background of black holes, which we take here to be a two-dimensional Schwarzschild black hole for simplicity. The counterpart of the Poincaré-invariant state is the Hartle-Hawking state, while the counterpart of the Rindler vacuum is the Boulware vacuum. In addition, for physical black holes we introduce the Unruh state. This is a hybrid state in which the outgoing (modes moving away from the past horizon) vacuum is chosen to be annihilated by the $\hat{a}$ operators, while the incoming (modes moving towards the future horizon) vacuum is chosen to be annihilated by the $\hat{b}$ operators. This state is singular on the past horizon but regular everywhere else. As already mentioned, we could also have chosen this hybrid state in Minkowski space, but there the singularity on the past horizon would still render it unphysical, much like the Rindler vacuum. In the context of a black hole that forms from a collapsing star, however, the Schwarzschild geometry is replaced at early times by the geometry of the collapsing star. Thus, for physical black holes, this state is acceptable and in fact natural. As we shall see the de Sitter counterpart of the black hole Unruh state is also well-defined on an entire planar patch and might even be a natural alternative to the commonly used Bunch-Davies state.

3.3 The Unruh-de Sitter state

Having described the Unruh state in two-dimensional Minkowski space and the Schwarzschild geometry, let us now define the Unruh state of a massless, minimally coupled scalar field in de Sitter space. For simplicity, we will first consider a scalar field in two-dimensional de Sitter space and later generalize our results to four dimensions. Recall the line element in static coordinates

$$ds^2 = -\left(1 - H^2 r^2\right) dt^2 + (1 - H^2 r^2)^{-1} dr^2 \ . \tag{3.2}$$

As we saw in Chapter 1, these coordinates only cover a quarter of de Sitter space and in terms of the tortoise coordinate

$$r_* = \frac{1}{2H} \log \left(\frac{1 + Hr}{1 - Hr} \right) \ , \tag{3.3}$$

one finds that two-dimensional de Sitter space is conformally flat

$$ds^2 = (1 - \tanh^2(Hr_*))(-dt^2 + dr_*^2) \ . \tag{3.4}$$

Thus, in terms of the lightcone coordinates

$$u = t + r_* \ , \qquad v = t - r_* \ , \tag{3.5}$$

the solutions to the wave equation are the same as in flat space.

$$\varphi(u, v) = \frac{1}{\sqrt{4\pi\omega}} \left(e^{-i\omega u} + e^{-i\omega v} \right) \ . \tag{3.6}$$

We also define lightcone coordinates that can be analytically continued to provide a global cover of de Sitter space, given by

$$U = -\frac{e^{-Ht}}{H} \sqrt{\frac{1 - Hr}{1 + Hr}} \ , \qquad V = \frac{e^{+Ht}}{H} \sqrt{\frac{1 - Hr}{1 + Hr}} \ . \tag{3.7}$$

The line element in these coordinates becomes

$$ds^2 = -\frac{4}{(H^2 UV - 1)^2} dU dV \ . \tag{3.8}$$

Using these coordinates, the wave equation has positive frequency solutions given by

$$\varphi(U, V) = \frac{1}{\sqrt{4\pi\omega}} \left(e^{-i\omega U} + e^{-i\omega V} \right) \ . \tag{3.9}$$

The Bunch-Davies state, understood as the analogue of the Hartle-Hawking vacuum for black holes, is now defined as the state that is annihilated by the $\hat{a}$ operators that multiply the positive frequency modes (3.9) for both incoming and outgoing modes, that is[4]

$$\hat{a}_{\text{in}} |0_{BD}\rangle = \hat{a}_{\text{out}} |0_{BD}\rangle = 0 \ . \tag{3.10}$$

The static vacuum, which is the counterpart of the Boulware state for black holes, can be defined as the state that is annihilated by the $\hat{b}$ operators that multiply the positive frequency modes (3.6) for both incoming and outgoing modes:

$$\hat{b}_{\text{in}} |0_S\rangle = \hat{b}_{\text{out}} |0_S\rangle = 0 \ . \tag{3.11}$$

Finally, similar as the Unruh state for black holes the Unruh-de Sitter state is the state that is annihilated by the $\hat{a}$ operator for incoming modes[5] (moving from the past horizon to an $r = 0$ observer), but annihilated by the $\hat{b}$ operator for outgoing modes (moving from an $r = 0$ observer to the future horizon):

$$\hat{a}_{\text{in}} |0_U\rangle = \hat{b}_{\text{out}} |0_U\rangle = 0 \ . \tag{3.12}$$

[4] Alternatively, the Bunch-Davies state can be defined by using mode functions that are positive frequency with respect to the time coordinate used in the flat planar slicing of de Sitter space. It is straightforward to show that the resulting vacuum state is equivalent.

[5] Notice that for black holes, incoming and outgoing is from the point of view of the black hole. In de Sitter space, incoming and outgoing refers to an $r = 0$ observer.

It is this state that we would like to study in more detail. It is essentially defined as a hybrid between the Bunch-Davies and the static de Sitter vacuum, making an asymmetric state choice for incoming (Bunch-Davies) and outgoing (static) modes. As a consequence this hybrid Unruh-like state breaks the de Sitter symmetries in the outgoing sector, by explicitly identifying a center, but is only singular at the past horizon. This will have important consequences for the analysis of the energy-momentum tensor expectation value and the corresponding stability of the state, to which we will now turn.

To evaluate the energy-momentum tensor in any particular state, one could expand the field φ in the different positive frequency mode functions, plug that into

$$\langle T_{\mu\nu} \rangle = \langle \partial_\mu \varphi \partial_\nu \varphi \rangle - \frac{1}{2} g_{\mu\nu} \langle \partial_\rho \varphi \partial^\rho \varphi \rangle \ , \tag{3.13}$$

and calculate the mode sums. The disadvantage of this direct procedure, however, is that a proper regularization scheme is required to obtain a finite answer, which is of course notoriously subtle in curved spacetimes [101]. Unlike flat space, there is no unique vacuum state with respect to which we can define a normal ordering prescription, making a standard normal-ordering procedure ambiguous at best. Hence, a more sophisticated regularization scheme would seem to be required.

We will actually avoid these subtleties by employing an approach pioneered in [102] and instead impose various consistency conditions that the components of $T_{\mu\nu}$ should satisfy, independent of the state under consideration. This allows us to construct the energy-momentum tensor in all generality, unambiguously and most importantly without the need for a regularization procedure that might obscure the relevant physics.

3.3.1 Vacuum expectation value of the energy-momentum tensor

The vacuum states that we would like to consider are all spherically symmetric and time independent, but not necessarily homogeneous. So we will only allow the energy-momentum tensor components to depend on the radius, that is $T_{\mu\nu} = T_{\mu\nu}(r)$. Then, in two dimensions the conservation of the energy-momentum tensor

$$\nabla_\mu T^{\mu\nu} = 0 \ , \tag{3.14}$$

leads to two differential equations for the components of the energy-momentum tensor, given by

$$\begin{aligned}\partial_r T^{tr} &= \frac{2H^2 r}{1-H^2r^2} T^{tr} \ , \\ \partial_r T^{rr} &= -H^2 r T^{\mu}_{\ \mu} \ .\end{aligned} \tag{3.15}$$

The first equation is solved by

$$T^{tr} = \frac{\Phi}{1-H^2r^2} \ , \tag{3.16}$$

where Φ is a constant. A non-zero off-diagonal component of the energy-momentum tensor implies that there is a net flux, so all states with $\Phi \neq 0$ contain a flux of energy momentum. The second equation can be solved by realizing that a massless scalar field in two dimensions is conformally invariant. This implies that the trace of the energy-momentum tensor picks up a conformal anomaly that is state independent. For two-dimensional de Sitter space the conformal anomaly is given by [103]

$$T^{\mu}_{\ \mu} = \frac{H^2}{12\pi} \ . \tag{3.17}$$

We then find

$$\begin{aligned}T_{tt} &= \frac{H^2}{24\pi}(H^2r^2-2) + \Omega \ , \\ T_{rr} &= \frac{1}{(1-H^2r^2)^2}\left(-\frac{H^4r^2}{24\pi} + \Omega\right) \ .\end{aligned} \tag{3.18}$$

As we will see in more detail next, the constant Ω is associated to the energy of particles present in the state under consideration without taking the flux Φ into account. Together with the flux parameter Φ, these two constants parametrize all possible two-dimensional conformally invariant states that obey the ansatz $T_{\mu\nu} = T_{\mu\nu}(r)$.

Transforming to lightcone coordinates the results are particularly transparant and read

$$\begin{aligned}T_{UU} &= -\frac{1}{48\pi U^2} + \frac{\Omega-\Phi}{2H^2U^2} \ , \\ T_{VV} &= -\frac{1}{48\pi V^2} + \frac{\Omega+\Phi}{2H^2V^2} \ , \\ T_{UV} &= -\frac{H^2}{12\pi(H^2UV-1)^2} \ .\end{aligned} \tag{3.19}$$

The physical interpretation of the different components is as follows. The off-diagonal T_{UV} term is completely fixed by the conformal anomaly and the diagonal components describe the (null) energy present in the state under consideration. The T_{UU} component describes the energy in incoming modes and T_{VV} captures the energy of outgoing modes. Remember that in de Sitter space, we defined incoming and outgoing from the point of view of a static observer at $r = 0$.

Different states can now be uniquely determined by specifying the value of the flux Φ and energy parameter Ω, or equivalently by specifying incoming and outgoing null energy. Notice that if we set the flux $\Phi = 0$, Ω completely determines the total energy. When the flux is non-zero ($\Phi \neq 0$) there is an additional energy contribution that contributes with an opposite sign to the diagonal components of the stress tensor, breaking the symmetry between incoming and outgoing modes. We therefore see that an incoming flux of positive energy particles ($\Phi < 0$), contributes negatively to the outgoing null energy and vice versa.

Because two-dimensional spacetimes are conformally flat, we can relate states constructed in Minkowski space to states in de Sitter space simply by a conformal transformation [103]. By performing a conformal transformation on the Rindler vacuum, we obtain a state in de Sitter space known as the static vacuum. This state is the natural vacuum for a static observer at $r = 0$ and looks empty. Hence, there is no flux and no energy associated with particles in this state and it is therefore given by

$$|0_S\rangle : \qquad \Phi = \Omega = 0 \ . \tag{3.20}$$

The vacuum expectation value of the energy-momentum tensor then only contains 'vacuum energy' and is given by

$$\begin{aligned} \langle 0_S | T_{UU} | 0_S \rangle &= -\frac{1}{48\pi U^2} \ , \\ \langle 0_S | T_{VV} | 0_S \rangle &= -\frac{1}{48\pi V^2} \ , \\ \langle 0_S | T_{UV} | 0_S \rangle &= -\frac{H^2}{12\pi (H^2 UV - 1)^2} \ , \end{aligned} \tag{3.21}$$

which indeed agrees with known results obtained from conformal transformation [103] or direct evaluation of mode sums [101]. The static vacuum is also the analogue of the Boulware vacuum for black holes. Just as the Rindler vacuum, the static vacuum is singular on both the past and future horizon. As a consequence it is not considered a reasonable physical state.

A different state is found by insisting the energy-momentum tensor to be regular on both the past and future horizon. This implies that the null energy vanishes and it singles out the Bunch-Davies state.

$$|0_{BD}\rangle : \qquad \Phi = 0 \ , \quad \Omega = \frac{H^2}{24\pi} \ . \tag{3.22}$$

Thus, the energy-momentum tensor in this state is given by

$$\langle 0_{BD}|T_{\mu\nu}|0_{BD}\rangle = \frac{H^2}{24\pi} g_{\mu\nu} \ , \tag{3.23}$$

again in agreement with standard results [101]. Because the off-diagonal component of the energy-momentum tensor vanishes in static coordinates, there is no net flux in the Bunch-Davies state, even though all free-falling observers will see a thermal spectrum of particles [104], as is indicated by a non-zero Ω. This means that the incoming flux from the horizon is compensated for by an equal but opposite amount of outgoing flux. In this sense, the Bunch-Davies state is a careful equilibrium of fluxes, just as the Hartle-Hawking state is for black holes. Furthermore, because the Bunch-Davies state preserves all de Sitter symmetries, which is immediately clear since it is proportional to the metric[6], the energy-momentum tensor does not result in backreaction and just renormalizes the bare cosmological constant.

Finally, we can also break the symmetry between incoming and outgoing modes. To construct the Unruh-de Sitter state, we pick the incoming sector to be in the Bunch-Davies vacuum and take the outgoing sector to be in the static vacuum. The null energy in the Unruh state should therefore agree with the incoming energy of the Bunch-Davies state and outgoing energy of the static vacuum. Due to this asymmetry, Φ is necessarily non-zero.

$$|0_U\rangle : \qquad \Phi = -\frac{H^2}{48\pi} \ , \quad \Omega = \frac{H^2}{48\pi} \ . \tag{3.24}$$

The energy-momentum tensor in the Unruh-de Sitter state becomes

$$\begin{aligned} \langle 0_U|T_{UU}|0_U\rangle &= 0 \ , \\ \langle 0_U|T_{VV}|0_U\rangle &= -\frac{1}{48\pi V^2} \ , \\ \langle 0_U|T_{UV}|0_U\rangle &= -\frac{H^2}{12\pi(H^2UV-1)^2} \ , \end{aligned} \tag{3.25}$$

which is only singular at the past horizon, as expected. Of course, the same energy-momentum tensor could also have been obtain by conformal transformation from the Unruh state for black holes or by explicitly evaluating mode sums.

Let us make a few comments regarding the form of the energy-momentum tensor in the Unruh-de Sitter state. First of all, notice that by construction an observer at $r = 0$ in this state only observes incoming radiation and no outgoing radiation. By removing the outgoing component of the radiation as compared to the

[6] To check if the energy-momentum tensor preserves an isometry we can act with a Lie derivative in the direction of the Killing vectors.

Bunch-Davies state, this has resulted in a negative energy density along a constant V slice that violates the null energy condition locally and is only singular at the past horizon. As a result, the Unruh state is well-defined on an entire planar patch and has the potential to be used as a physically acceptable state in this region. This is similar to what happens in the black hole case. There, the Unruh state is constructed by taking modes coming from the past horizon to be in the Harte-Hawking vacuum and modes moving through the future horizon to be in the Boulware vacuum. As a result, an asymptotic observer only measures radiation coming from the past horizon, which is singular and violates the null energy condition [102].

Of course, it would be interesting to investigate if a more general energy condition such as the averaged null energy condition does hold in the Unruh-de-Sitter state, which is expected [105]. In fact, as discussed extensively in [8] for the Rindler case, when integrating over the past horizon there are additional positive (singular) contributions that ensure that the integrated energy-momentum tensor is indeed positive.

Because the energy-momentum tensor in the Unruh state is only singular at the past horizon, it is well-defined on an entire planar patch. In planar coordinates, which are given by

$$ds^2 = \frac{1}{H^2\eta^2}\left(-d\eta^2 + dr^2\right) \ , \tag{3.26}$$

the energy-momentum tensor in the Unruh-de Sitter state reads

$$\begin{aligned} \langle 0_U|T_{\eta\eta}|0_U\rangle &= -\frac{1}{24\pi\eta^2} - \frac{1}{48\pi(\eta - r)^2} \ , \\ \langle 0_U|T_{rr}|0_U\rangle &= +\frac{1}{24\pi\eta^2} - \frac{1}{48\pi(\eta - r)^2} \ , \\ \langle 0_U|T_{\eta r}|0_U\rangle &= \frac{1}{48\pi(\eta - r)^2} \ . \end{aligned} \tag{3.27}$$

One might expect any difference from the Bunch-Davies vacuum to be redshifted away exponentially fast, but by transforming the time-time component to the standard planar time coordinate t (given by $\partial\eta/\partial t = -H\eta$) we find a source of energy-momentum that does not redshift away with time. Furthermore, in the limit $(-r/\eta) \gg 1$ (corresponding to superhorizon scales), the above expressions reduce to the energy-momentum tensor in the Bunch-Davies state. Therefore, the difference between the energy-momentum tensor in the Unruh-de Sitter state and the Bunch-Davies vacuum is largest in a Hubble-sized region and vanishes at large superhorizon scales where all de Sitter symmetries are restored. The relevant subhorizon contributions are finite and break the de Sitter symmetries (only rotational invariance and time translations are preserved), even when averaged over a Hubble-sized region. Of course, these results are strictly two-dimensional, so let us now consider how these results can be generalized to four dimensions.

3.3.2 Generalization to four dimensions

The method we used to construct the energy-momentum tensor in two dimensions in principle generalizes without much difficulty to four dimensions. The main difference is that the additional components of the energy-momentum tensor imply more free components, unless additional constraints are imposed. By constraining to emission in the s-wave sector the difference between the two and four-dimensional case can be captured by a greybody factor. As we will argue more precisely below, close to the horizon this ensures that the s-wave energy-momentum tensor in four dimensions is effectively two-dimensional.

Consider the action of a massless minimally-coupled scalar field in four-dimensional de Sitter space in static coordinates. The metric is given by

$$ds^2 = -\left(1 - H^2r^2\right)dt^2 + (1 - H^2r^2)^{-1}dr^2 + r^2\left(d\theta^2 + \sin^2\theta\ d\phi^2\right)\ , \qquad (3.28)$$

and the action for the scalar field on this background is

$$S = -\frac{1}{2}\int d^4x\sqrt{-g}\ \partial_\mu\varphi\partial^\mu\varphi\ . \qquad (3.29)$$

If we concentrate on just the s-wave ($\varphi = \varphi(t, r)$), we can integrate over the two-sphere to obtain an effective two-dimensional action.

$$S = -2\pi\int d^2x\sqrt{-g_2}\ r^2\partial_\mu\varphi\partial^\mu\varphi\ . \qquad (3.30)$$

Here g_2 is the determinant of the metric of two-dimensional de Sitter space. We can go to a canonically normalized field by rescaling $\varphi \to (4\pi r^2)^{-1/2}\varphi$ such that the action becomes

$$S = \int d^2x\sqrt{-g_2}\left(-\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V_{\text{eff}}(\varphi)\right)\ , \qquad (3.31)$$

where

$$V_{\text{eff}}(\varphi) = \frac{(1 - H^2r^2)}{2r^2}\left(1 - r\partial_r\right)\varphi^2\ . \qquad (3.32)$$

We conclude that the difference between the action of a scalar field in two and four dimensions is captured by the effective potential $V_{\text{eff}}(\varphi)$, which vanishes at the horizon $r = 1/H$. The radial potential modifies the propagation of modes away from the horizon to the center of de Sitter space, which can be taken into account by introducing a greybody factor. So as far as high frequency s-wave modes near the horizon are concerned, their energy should be related to the two-dimensional energy-momentum tensor results.

In general, under the assumption that $T_{\mu\nu} = T_{\mu\nu}(r)$ and that the only non-vanishing off-diagonal component is T_{tr}, conservation of the energy-momentum tensor

$$\nabla_\mu T^{\mu\nu} = 0 \ , \tag{3.33}$$

in four dimensions leads to the following (differential) equations

$$\begin{aligned} \partial_r T^{tr} &= -\frac{2\left(2H^2r^2-1\right)}{r\left(H^2r^2-1\right)} T^{tr} \ , \\ \partial_r T^{rr} &= -\frac{2}{r} T^{rr} + 2rT^{\theta\theta} - H^2 r T^\mu{}_\mu \ , \\ T^{\phi\phi} &= \sin^{-2}\theta \, T^{\theta\theta} \ . \end{aligned} \tag{3.34}$$

An important difference with two dimensions, where the massless scalar field is conformally invariant, is that the trace $T^\mu{}_\mu$ is not fixed by the conformal anomaly and as a consequence does not introduce additional constraints.

The most general solution to (3.34) is given by

$$T_{tt} = \frac{1}{r^2}\left(\Delta - H^2 \int_{1/H}^r dr' \ r'^3 T^\mu{}_\mu + \Theta(r) + 2(1-H^2r^2)T_{\theta\theta} - r^2(1-H^2r^2)T^\mu{}_\mu\right) \ , \tag{3.35}$$

$$\begin{aligned} T_{rr} &= \frac{1}{r^2(1-H^2r^2)^2}\left(\Delta - H^2\int_{1/H}^r dr' \ r'^3 T^\mu{}_\mu + \Theta(r)\right) \ , \\ T_{tr} &= -\frac{\Phi}{r^2(1-H^2r^2)} \ , \end{aligned}$$

with Φ, Δ constants and

$$\Theta(r) \equiv 2\int_{1/H}^r dr' \ \frac{T_{\theta\theta}}{r'} \ . \tag{3.36}$$

As before, we see that a non-zero Φ introduces a net flux and breaks the symmetry between incoming and outgoing modes. Also notice that the T_{tr} component has an apparent singularity at $r = 0$ and at $r = 1/H$. The singularity at the origin is just a consequence of the fact that in static coordinates the origin is a singular point. The flux density diverges at $r = 0$ due to the vanishing volume of the two-sphere, but the energy obtained by integrating over the two-sphere is of course finite and we could have also chosen to remove this coordinate singularity by an appropriate coordinate transformation. In contrast, the singularity at the horizon vanishes only in an appropriate (non-singular) quantum state evaluated in a coordinate system that is regular at the horizon.

We will now make use of the fact that the action for the four-dimensional massless minimally coupled scalar in the s-wave sector effectively reduces, at the de Sitter horizon, to a two-dimensional action of a massless minimally-coupled scalar field, obtained by integrating over the two-sphere and canonically normalizing the scalar field. This implies that close to the horizon the two energy-momentum tensors are related as follows

$$T^{(4d)}_{\mu\nu} = \frac{H^2}{4\pi} T^{(2d)}_{\mu\nu} \ , \tag{3.37}$$

where the entire expression should be evaluated at the future horizon in a non-singular coordinate system. In the Unruh-de Sitter state (see (3.25)), one obtains

$$\begin{aligned} \langle 0_U | T_{UU} | 0_U \rangle &= 0 \ , \\ \langle 0_U | T_{VV} | 0_U \rangle &= -\frac{H^2}{192\pi^2 V^2} \ , \\ \langle 0_U | T_{UV} | 0_U \rangle &= -\frac{H^4}{48\pi^2} \ , \end{aligned} \tag{3.38}$$

which has been evaluated at the future horizon ($U = 0$). As before, we see that the energy-momentum tensor in the Unruh-de Sitter state is singular at the past horizon and breaks de Sitter symmetries as it preserves only $SO(3) \times R$ corresponding to spatial rotations and time translations. The state is inhomogeneous by introducing just incoming flux, which results in a conserved energy-momentum tensor describing a net flow of energy and momentum. This is of course bound to result in a nontrivial backreaction effect on the initial de Sitter geometry. Also note that just as in the two-dimensional case, the energy density of the Unruh-de Sitter state is negative. In the class of (inhomogeneous) states the Bunch-Davies vacuum is now very special, corresponding to the case where the incoming and outgoing fluxes precisely cancel, restoring the de Sitter symmetries. It is straightforward to check that the energy-momentum tensor is then indeed proportional to the de Sitter metric, as it should.

Now that we have determined the vacuum expectation value of the four-dimensional s-wave energy-momentum tensor in the Unruh-de Sitter state, close to the horizon, let us next turn to an estimate of the backreaction effect on the de Sitter geometry it induces.

3.4 Backreaction and de Sitter evolution

Before we estimate the backreaction effect, we would like to compare the situation to computing backreaction in black hole spacetimes. An important difference with de Sitter space is that for black holes one can rely on a globally conserved energy, which makes estimating the backreaction effect more tractable. In the Unruh state, there is a flux of energy-momentum coming from the black hole, described by T_{tr} [102], and energy conservation immediately implies that the mass of the black hole should decrease. Notice that although the Unruh state for black holes preserves time translations, by imposing energy conservation we are forced to conclude that the black hole geometry evolves.

Although in de Sitter space there is no globally conserved energy, we can still use the semi-classical Einstein equations to learn about the evolution of de Sitter space. The effect of backreaction on the geometry is then estimated by putting the field in the Unruh-de Sitter state and plugging the energy-momentum tensor into the semi-classical Einstein equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N \langle T_{\mu\nu} \rangle \ , \tag{3.39}$$

which should be consistent when the backreaction effect is small and only evolves slowly as compared to the typical timescale set by the background. This is also the procedure one uses for black holes in the Unruh state. As mentioned before, notice that despite the fact that the energy-momentum tensor in the Unruh state is time-translation invariant, one finds by including backreaction that the (static) black hole geometry has to be modified to be consistent with the semi-classical Einstein equations. By doing so in an iterative procedure, time evolution is generated and it is found that the black hole decays.

The same procedure can be applied to de Sitter space in the Unruh state. We will use a semi-classical de Sitter space ($H \ll M_p$) as our initial background in order to calculate the right-hand side of the equation and then estimate how the small energy-momentum tensor correction affects the geometry on the typical Hubble timescale of the original de Sitter geometry. Once again, this iterative procedure is completely analogous to the situation for evaporating black holes. There $\langle T_{\mu\nu} \rangle$ is calculated for a static Schwarzschild spacetime, even though the outgoing Hawking flux in the Unruh state implies that the geometry is only approximately static on the typical Schwarzschild timescale.

Before we derive an estimate of the backreaction effect it is important to emphasize that we have only derived the four-dimensional energy-momentum tensor in the Unruh-de Sitter state close to the horizon. Nevertheless, to estimate the effects on the de Sitter geometry we will use a homogeneous and isotropic FLRW ansatz which is, strictly speaking, incorrect. In fact, conservation of a perfect fluid energy-momentum tensor for a homogeneous and isotropic FLRW ansatz forces any backreaction effects to redshift away quickly. Indeed, it is the inhomogeneous nature of the Unruh-de Sitter state that allows for a maintained flux of stress energy. Nevertheless, even though the energy-momentum tensor in the Unruh-de

Sitter state breaks homogeneity, its effects are dominant in a Hubble-sized region and to estimate backreaction the energy can be averaged over a single static patch. On large superhorizon scales, the energy-momentum tensor should reproduce the Bunch-Davies energy-momentum tensor, just as in the two-dimensional case.

Then, we can approximate the backreaction by using a homogeneous and isotropic Friedmann-Lemaitre-Robertson-Walker (FLRW) ansatz

$$ds^2 = -dt^2 + a^2(t)\left(dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right) \ , \tag{3.40}$$

where t is now the planar time coordinate and $a(t)$ the scale factor. Combining the time-time and spatial-spatial components of the Einstein equations, the evolution of the Hubble parameter is given by

$$\dot{H} = -\frac{4\pi G_N}{a(t)^2}\left(\langle T_{\eta\eta}\rangle + \langle T_{rr}\rangle\right) \ , \tag{3.41}$$

where the dot denotes a time derivative with respect to t and $T_{\eta\eta}$ is the time-time component of the energy-momentum tensor using conformal time ($\partial\eta/\partial t = -H\eta$).

We now use the value of the near-horizon energy-momentum tensor as a proxy for the average (integrated) energy-momentum in a Hubble-sized region. Again, in two dimensions this is an excellent approximation and we will assume this is also true in four dimensions. We then find (using (3.38))

$$\begin{aligned} \langle 0_U|T_{\eta\eta}|0_U\rangle &= -\frac{3H^2}{256\pi^2\eta^2} \ , \\ \langle 0_U|T_{rr}|0_U\rangle &= \frac{7H^2}{768\pi^2\eta^2} \ , \\ \langle 0_U|T_{\eta r}|0_U\rangle &= \frac{H^2}{768\pi^2\eta^2} \ . \end{aligned} \tag{3.42}$$

Furthermore, just as in the black hole case we will choose to ignore the off-diagonal component, since it does not affect the backreaction on the (local) Hubble parameter to leading order. In a more complete treatment it would be interesting to keep this term and calculate how it modifies the evolution.

Plugging in the small and approximately homogeneous and isotropic energy-momentum tensor correction, the evolution equation for the Hubble parameter is (with $N \equiv \log a$ the number of e-folds and $M_p^{-2} = 8\pi G_N$ the reduced Planck mass)

$$\dot{H} \approx \frac{H^4}{768\pi^2 M_p^2} \quad \to \quad \frac{1}{H}\frac{dH}{dN} \approx \frac{1}{768\pi^2}\frac{H^2}{M_p^2} \tag{3.43}$$

implying that the relative magnitude of the backreaction effect per e-fold is suppressed as $(H/M_p)^2$. The inverse can then be associated to a bound on the lifetime of de Sitter space that is of the order of the de Sitter entropy

$$N_{\rm max} = H\,t \sim M_p^2/H^2 \propto S_{dS}\ . \tag{3.44}$$

As mentioned earlier, the energy-momentum tensor in the Unruh-de Sitter state violates the null energy condition. By again referring to the analogy with black holes this should not come as a surprise: we know that black holes evaporate in the Unruh state and, by Hawking's area theorem, a decrease in the horizon area is only possible by violating the null energy condition (locally, at the horizon scale). Nevertheless, it would be interesting to study more general energy conditions such as the averaged null energy condition or the quantum null energy condition in this context. We conclude that the presence of negative vacuum energy in the Unruh-de Sitter state effectively causes the Hubble radius to *decrease*. The evolution can be followed as long as the effects are in the semi-classical regime, until it enters a highly curved inhomogeneous and presumably singular regime beyond an effective field theory description.

Let us make a few comments regarding the approximations we made. To obtain an estimate of the effect of backreaction on the background geometry, we used several assumptions regarding the form of the energy-momentum tensor in four dimensions. Based on the two-dimensional result, we expect the difference between the Unruh state and the Bunch-Davies vacuum to be dominant in a Hubble-sized region. In addition, we also assumed the near-horizon value of the energy-momentum tensor to give a good estimate of the true averaged energy-momentum and we ignored the off-diagonal component of the energy-momentum tensor. It is important to mention however that, despite the subtleties in trying to track the evolution of the backreacted geometry, the conclusion that pure de Sitter space is no longer a solution in the Unruh state holds regardless of these assumptions. Contrary to an exactly homogeneous and isotropic energy-momentum tensor the continuous flux of stress energy in the Unruh-de Sitter state does not simply redshift away. Because of this, energy-momentum present in the Unruh state backreacts on the geometry such that pure de Sitter space is no longer a consistent solution. Just as for black holes in the Unruh state, the Unruh-de Sitter state contains negative outgoing null energy at the future horizon (given by T_{VV}) which is present because of the removal of outgoing flux as compared to the Bunch-Davies vacuum. This negative energy violates the null energy condition and backreaction on the geometry causes the de Sitter horizon to shrink.

In future work it would be interesting to go beyond the approximations we made here to track the evolution of the instability, perhaps using an approach based on cosmological perturbation theory.

3.4.1 Consistency and interpretation

Although exact global de Sitter space is a very interesting spacetime, the typical cases of physical interest only require a geometry which is approximately de Sitter space in a few Hubble-sized regions of space and time. Examples include i) the interior of a bubble of false vacuum in global de Sitter space and ii) a patch of spacetime dominated by vacuum energy leading to inflation. In each of these cases, the local geometry is essentially that of de Sitter space. However, certain global features, such as the existence of past horizons or IR issues [106,107] should now be revisited. In these examples the existence of a singular $\langle T_{\mu\nu} \rangle$ at the past horizon is not necessarily problematic and just indicates our ignorance to how the universe started off in a state dominated by vacuum energy. Again this is analogous to physical black holes that form from collapsing matter. In these scenarios the inhomogeneous Unruh state therefore seems to be a reasonable candidate for an initial state. More generally, any linear combination of the Bunch-Davies state and the Unruh state would then also be acceptable. As a first step, in this work we just constructed the Unruh-de Sitter state and analyzed its implications for the evolution of de Sitter space.

An important observation is that the state constructed is asymmetric and can be viewed as a combination of the Bunch-Davies vacuum for incoming modes and the static vacuum for outgoing modes. As a consequence it breaks homogeneity, but only mildly so, allowing for an approximate analysis of the backreacted evolution in terms of slightly corrected FLRW equations. The presence of the static vacuum for outgoing modes implies a singularity at the past horizon, but it is well-defined across the future horizon on a (large) planar patch, and explains the presence of negative energy density as compared to the Bunch-Davies state. By removing the outgoing thermal flux that would be present in the Bunch-Davies state, as seen by a static observer, the system is no longer in thermal equilibrium. In fact, the negative energy density in the static outgoing vacuum depends on the de Sitter radius, becoming more negative as the Hubble radius decreases, explaining the direction of the instability towards smaller de Sitter radius and equivalently higher de Sitter temperature. Note also that by generalizing to a class of isotropic, but inhomogeneous, initial states, the Bunch-Davies choice is very special, fine-tuned and in fact unstable to small perturbations.

If we would be interested in using the Unruh state for cosmological purposes, one might worry that the introduction of an inhomogeneous initial state might be in direct conflict with observation. Inflationary states should produce a large, flat and very homogeneous observable universe. However, selecting the Unruh-de Sitter vacuum does not change this conclusion. The state just identifies an initial center, or 'patch', that exponentially expands and can accommodate the entire observable universe after roughly 60 e-folds of inflation. The inherent instability of the Unruh-de Sitter state only becomes apparent after S_{dS} e-folds, which is typically much larger than 60 e-folds. We therefore expect a suppression of corrections to inflationary correlation functions computed in the Bunch-Davies state. In other words, the properties of the Unruh-de Sitter state are expected to be indistinguishable from the Bunch-Davies vacuum, as long as the number of e-folds is much smaller than the de Sitter entropy. We hope to explicitly confirm these expectations in future work.

Another point of concern with the Unruh-de Sitter state is how the decrease of the de Sitter horizon can be consistent with thermodynamics, as the decrease of the horizon radius naively seems to violate the second law. Of course, for black holes we know how this works. By taking into account the entropy of the emitted radiation, the total generalized entropy still increases. It is therefore possible that in de Sitter space, something similar happens.

Indeed, the change in entropy is given by the change in horizon area, which in the Unruh state decreases

$$\dot{S}_{horizon} = \frac{\dot{A}}{4\,G_N} = -\frac{16\pi^2\,M_p^2}{H^3}\dot{H} < 0 \ . \tag{3.45}$$

However, the system is not in complete equilibrium and the energy flux through the horizon also carries an entropy which can be calculated using the first law. We will consider the entropy in a fixed proper volume V and, as before, we will assume the radiation to be distributed in a homogeneous and isotropic manner. From the first law, the entropy change is then given by

$$dS_{matter} = \frac{dE}{T} \ . \tag{3.46}$$

By using the continuity equation

$$\dot{\rho} = -3H(\rho + p) \ , \tag{3.47}$$

where ρ is the energy density and p the pressure we can express this as

$$\dot{S}_{matter} = -\frac{3H}{T}(\rho + p)V \ . \tag{3.48}$$

Any contribution from a cosmological constant ($\rho = -p$) would drop out, as expected. By combining this with the result for the change in horizon area and by using the Friedmann equations, the total entropy change can be written as

$$\dot{S}_{total} = \dot{S}_{horizon} + \dot{S}_{matter} = \left(\frac{8\pi^2}{H^3} - \frac{3HV}{T}\right)(\rho + p) \ . \tag{3.49}$$

Now we take the temperature of the matter to be at the de Sitter temperature $T_{dS} = \frac{H}{2\pi}$ and the volume to be a Hubble volume: $V = \frac{4\pi}{3H^3}$. We then find that both entropy contributions cancel exactly such that

$$\dot{S}_{total} = 0 \ , \tag{3.50}$$

which saturates the second law. This was already observed in [108], showing that a decrease of the horizon area is not necessarily in conflict with the second law.

The fact that the generalized entropy vanishes in a static patch is also supported by Bousso's N-bound [109], which roughly states that the entropy in a causal diamond in a spacetime with a positive cosmological constant should be bounded from above by the de Sitter entropy. Because the static patch is the largest causal diamond in de Sitter, it better be that $S_{total} \leq S_{dS}$. At the same time, if we assume the initial entropy to be given by the de Sitter entropy the generalized second law implies $S_{total} \geq S_{dS}$. Therefore, the only way how these two bounds can be consistent with each other is if the total entropy in a static patch remains constant.

For volumes larger than the proper volume of a static patch, the N-bound no longer applies and the matter entropy term in (3.49) at fixed temperature dominates over the horizon contribution, which implies that

$$\dot{S}_{tot} > 0 \ , \tag{3.51}$$

only when $(\rho + p) < 0$. Notably, this is the case in the Unruh-de Sitter state. We therefore see that also in the Unruh state in de Sitter space, a decrease of the horizon area is not directly in conflict with the second law of thermodynamics, at least not when we approximate the energy as being homogeneously spread out over a single Hubble volume.

3.5 Conclusions

In this chapter, we considered quantum field theory of a massless minimally coupled scalar field on a de Sitter background. We discussed and constructed different vacuum states that are the de Sitter analogue of the Hartle-Hawking, Boulware and Unruh vacua for black holes. In particular, focussing on two-dimensional de Sitter space, we computed the energy-momentum tensor in these different states and related the two-dimensional energy-momentum tensor to the near-horizon energy-momentum tensor in four dimensions in the s-wave sector. We argued that the Unruh-de Sitter state, which is only singular at the past horizon, might be a viable and natural alternative to the de Sitter invariant Bunch-Davies state which is typically assumed in inflationary scenarios. Modifications to the standard predictions from 60 e-folds of slow-roll inflation assuming the Bunch-Davies state are likely negligible, but we would like to investigate this further in future work. If the instability we found in the Unruh state persists, it puts a fundamental bound on the maximum number of e-folds possible given by the de Sitter entropy.

Such a bound might be relevant in the context of the recent (refined) de Sitter swampland conjectures [87, 88]. Note however that, first of all, the instability of the Unruh-de Sitter state is much slower than any of the de Sitter swampland conjectures, removing any potential tension with single field slow-roll inflation or dark energy constraints. Secondly, the instability of the Unruh-de Sitter state evolves towards smaller de Sitter radius which, as we explained, has a clear physical interpretation. This is to be contrasted with the de Sitter swampland conjectures, in which the de Sitter radius increases. This different behaviour is not in conflict with our results, as the setup leading to this conclusion is rather different; in the de Sitter swampland conjectures it is assumed that a tower of states becomes light which contributes to the de Sitter entropy, whereas we do not consider such a situation. Our results therefore seem to be more closely related to the so-called 'quantum break time' of de Sitter that was recently revisited in [110], and perhaps to the vacuum state modifications suggested in [111].

The physical origin of the instability is clear. Similar to black holes, one identifies a special static observer that only measures a flux of incoming radiation, which implies that the expectation value of the vacuum energy is negative, producing a backreaction effect that shrinks the horizon slightly. This conclusion is also supported by the complementary point of view in terms of radiation tunneling through the de Sitter horizon [112]. In an effective field theory description of the tunneling process in de Sitter space one indeed requires the identification of a state that is empty for an observer freely falling through the horizon. Just as for black holes [1], this naturally selects the Unruh state. This identifies a special observer, breaking homogeneity. The inhomogeneous nature of the state is crucial, since it introduces unbalanced incoming and outgoing flux leading to an instability. Contrary to the de Sitter invariant Bunch-Davies vacuum, the Unruh-de Sitter state is a non-equilibrium state. In fact, in the class of all isotropic states that are non-singular on the future horizon, the fine-tuned homogeneous Bunch-Davies vacuum is perturbatively unstable.

By allowing inhomogeneous vacuum states that only introduce singularities at the past horizon, the situation in de Sitter becomes analogous to that for black holes. Indeed, for initial Hubble parameters far below the Planck scale, allowing for an effective field theory description, the instability is tiny and measured in terms of the de Sitter entropy. As a consequence we expect these initial state effects to be suppressed in the context of slow-roll inflationary phases that only require $\mathcal{O}(100)$ e-folds. It will certainly be interesting to show this more explicitly in terms of correlation functions to characterize the differences between cosmological observables in the Unruh and Bunch-Davies state more precisely.

Furthermore, we only gave a preliminary estimate of the effect of backreaction in the Unruh state and we relied on an FLRW ansatz for the backreacted geometry. Because the Unruh state (mildy) breaks homogeneity, it would be of interest to go beyond this approximation and calculate backreaction in a complete treatment. This should reveal exactly how de Sitter space evolves and if the Unruh vacuum indeed is a well-behaved state (as we anticipated here) that can be consistently used for cosmological purposes. We hope to come back to some of these interesting questions in future work.

4 Supersymmetry Breaking in de Sitter Space

After having discussed quantum effects in de Sitter space in semi-classical gravity in Chapter 3, we now continue our study of de Sitter space, but in string theory. This chapter is based on [3, 4] and focusses on spontaneous supersymmetry breaking. In particular, we apply the formalism of constrained superfields to understand the effective description of de Sitter vacua in the KKLT scenario [32], which we reviewed in Chapter 1. We show, by including polarization effects, that the anti-D3-branes used in this construction break supersymmetry spontaneously. This allows us to identify the leading corrections of the effective description of KKLT vacua in terms of a single nilpotent goldstino superfield. The energy scale at which this description breaks down is rather low and we comment on the implications for cosmological models based on the KKLT scenario.

4.1 Introduction

De Sitter vacua are at the heart of any cosmological model. However, to fully address the UV-sensitivity of inflation and embed dark energy into a UV-complete framework, it is necessary to understand how de Sitter vacua can be constructed in string theory and supergravity. Unfortunately, this has proven to be a tremendous challenge and one has to face several challenges we reviewed in Chapter 1 (see also the introduction of [35]) to construct solutions of string theory with a positive cosmological constant. Nevertheless, a breakthrough came in 2003 when Kachru, Kallosh, Linde and Trivedi (KKLT) provided a generic mechanism of moduli stabilization in anti-de Sitter space and a subsequent uplift to a de Sitter vacuum by introducing supersymmetry-breaking anti-D3-branes [32]. Also, after that many other different approaches for de Sitter compactifications have been uncovered and we refer the reader to [24] for an excellent review.

This does not mean however that all details of de Sitter solutions in string theory are understood. One point of concern regarding the KKLT scenario specifically has always been that it was unclear if the anti-D3-branes can be understood to break supersymmetry spontaneously in an effective four-dimensional theory. Only recently, a description of de Sitter vacua that exhibit spontaneous supersymmetry breaking was uncovered in effective four-dimensional $N = 1$ supergravity, by making use of constrained superfields [113–117]. By imposing constraints on a (linearly supersymmetric) multiplet, it becomes possible to describe fields transforming nonlinearly under spontaneously broken supersymmetry and at the same time it eliminates unwanted degrees of freedom. General prescriptions to obtain constrained superfields from linearly transforming ones in a supergravity context were given in [118, 119]. For some reviews of constrained superfields and their applications to cosmology, see [120, 121].

Over the years, also other aspects of KKLT have been scrutinized, such as the possible dangerous backreaction of anti-D3-branes on the internal space [44–58] and antibrane backreaction on four-dimensional moduli [38, 122]. Recently, new doubts have even been cast on general aspects of moduli stabilization, before adding antibranes, in flux compactifications [35]. It is thus fair to say that various details of the KKLT scenario still need to be better understood. For some recent work discussing these and other issues regarding constructions of de Sitter vacua in string theory, we refer the reader to [10, 34, 36, 39–43, 59, 87, 123–138].

In this chapter, we will focus our attention on understanding how anti-D3-branes in flux backgrounds break supersymmetry spontaneously and show how this has a description in term of constrained superfields. For this purpose, it is important to remember that constrained superfields are often only effective descriptions of low-energy excitations. For example, we saw in Chapter 1 that in a vacuum with spontaneously broken supersymmetry the massless goldstino can be packaged in a chiral superfield that satisfies a nilpotent constraint, which is only valid below the mass scale of the scalars that are projected out [28, 139, 140]. As argued in [28], this can be extended to multiple fields and integrating out additional heavy degrees of freedom generally results in extra constraints that describe the universal low-energy dynamics of the theory, see also [118, 140].

It is thus of crucial importance to understand the embedding of constrained superfields in a putative UV-complete description. Can we indeed realize large mass splittings such that the constrained superfields correspond to a good approximation of the relevant low-energy physics? This question is especially important when considering cosmology. As one typically accesses high energy scales during inflation it is necessary to ensure that the fields eliminated by the constraints have large enough masses to be integrated out. Otherwise, a constrained superfield description will be invalid. Such concerns have been raised in [141] and, in a specific model of inflation, it has been shown in [142] that finite mass effects can significantly limit the range of parameters for which an effective nilpotent description is available.

In the context of KKLT, the first effective description of a single anti-D3-brane in a flux background in terms of constrained superfields has been constructed by putting the anti-D3-brane on top of an orientifold plane [143, 144]. This simplifies the problem, as it projects out all bosonic degrees of freedom on the worldvolume. The low-energy spectrum then only consists of a single massless gaugino (originally part of a vector multiplet) that plays the role of goldstino. Unfortunately, by doing so it is no longer possible to restore linear supersymmetry within this theory, making it difficult to estimate the size of corrections to a constrained superfield description. To answer this question, it is necessary to remove the orientifold plane and this study was initiated in [145, 146], where it was found that also the other degrees of freedom on the worldvolume of the anti-D3-brane can be captured by additional constrained superfields.

However, to learn how a constrained superfield description of an anti-D3-brane might emerge from a linearly supersymmetric theory and to study the leading corrections, we will choose to study multiple antibranes probing the Klebanov-Strassler (KS) geometry [147].[1] It has been shown by Kachru, Pearson and Verlinde (KPV) that in this background, anti-D3-branes polarize to form an NS5-brane wrapping an S^2 and settle into a metastable state [148]. The upshot of taking into account polarization is that it shows there is a direct connection between the nonsupersymmetric metastable state and a supersymmetric vacuum in which all antibranes have annihilated. This makes it possible to see the supersymmetry-breaking metastable phase emerge from a linearly supersymmetric theory.

To correctly identify the goldstino after including polarization effects, we reduce the NS5-brane action to four dimensions and truncate to the lowest lying modes of a Kaluza-Klein tower. We then indeed find that there lives a massless gaugino on the NS5-brane worldvolume, which was identified as the goldstino in the case of a single anti-D3-brane. However, one should not jump to the conclusion too soon that this is still the goldstino when polarization effects are taken into account and make sure that it also transform correctly under supersymmetry. By computing the supersymmetry transformations in the metastable minimum we find that the standard non-linear transformations receive corrections that scale with M/p, with p the number of antibranes and M the flux number of the KS background. Hence, a strict decoupling limit where these corrections vanish does not exist, since it is equivalent to sending the dimensionless ratio p/M to infinity, while the metastable state only exists for sufficiently small p/M. This signals a breakdown of the

[1] To study the physics of anti-D3-branes in warped throats, such as in the KKLT scenario, it is standard practice to model the throat by the noncompact KS geometry [147].

effective theory and suggests that the gaugino cannot be identified as the goldstino in the metastable state.

This can be explained by the fact that, due to polarization, the theory probes higher-dimensional physics and ignoring the Kaluza-Klein tower is inconsistent. We illustrate this by constructing a linearly supersymmetric model, in which a metastable state emerges that breaks supersymmetry spontaneously and matches many features of the true metastable vacuum. We explicitly confirm the presence of a goldstino in this model and observe that the goldstino does not coincide with the massless gaugino. We find that restoring linear supersymmetry requires including the full dynamics on the S^2 which implies that the standard effective reduction of KPV to a single (massive) scalar degree of freedom can be misleading and in fact obscures the appearance of a massless goldstino in the metastable vacuum. Only at energy scales below the Kaluza-Klein scale, one can integrate out the massive tower of states on the S^2 after which a description in terms of a single nilpotent goldstino superfield, in addition to the four-dimensional massless vector multiplet, becomes accurate.

This chapter is organized as follows. In section 4.2 we review some of the relevant bosonic details of brane polarization and show that the full polarization dynamics involves a Kaluza-Klein tower. Then, we include fermions in section 4.3 and compute the supersymmetry transformations in section 4.4. Continuing, we argue in section 4.5 that a full understanding of supersymmetry breaking in the metastable state requires including higher-dimensional physics. We then present our supersymmetric model that embeds the metastable minimum into a linearly supersymmetric theory and explicitly confirm the existence of a goldstino. We close this chapter in section 4.6 with the conclusions and an appendix featuring some technical details on fermions.

4.2 Brane polarization: bosonic action

We will start by giving a short recap of some of the main results of KPV [148] and highlight the features that will be important later on. In particular, we will show that the effective potential they derived is a consistent truncation that is useful to determine some qualitative features of the metastable polarized state, but that it is not suitable to be used as a four-dimensional effective action, since it does not take into account all of the dynamics on the S^2.

KPV considered creating a nonsupersymmetric solution by adding a small number of anti-D3-branes to the KS geometry [147], which is a non-compact version of a

GKP solution [33] which we reviewed in Chapter 1. This geometry consists of a long warped throat with a topology $R^{1,3} \times S^3$ at its tip. The throat is supported by putting M units of RR 3-form flux through the A-cycle and K units of NSNS 3-form flux through the B cycle of the throat

$$\frac{1}{4\pi^2}\int_A F_3 = M \ , \qquad \frac{1}{4\pi^2}\int_B H_3 = -K \ . \tag{4.1}$$

At the tip of the throat, the metric is given by

$$ds^2 = e^{2A_0}\eta_{\mu\nu}dx^\mu dx^\nu + e^{-2A_0}(d\psi^2 + \sin^2\psi \ d\Omega_2^2) \ , \tag{4.2}$$

where $e^{2A_0} \simeq (g_s M)^{-1} \ll 1$ is the warp factor at the tip. Since KPV considered $p > 1$ antibranes, the wordvolume description of the anti-D3-branes becomes a non-Abelian gauge theory with gauge group $U(p)$, that contains noncommuting matrix degrees of freedom. This allows for a version of the Myers effect [149] where the anti-D3-branes polarize to collectively form a spherical 5-brane configuration that has a description for $p \gg 1$ in terms of an NS5-brane[2] wrapping an S^2 of the S^3. The bosonic action for the NS5-brane is given by

$$S_{\text{NS5}} = -\frac{\mu_5}{g_s^2}\int d^6\xi\sqrt{-\det(G_{||})\det(G_\perp + 2\pi g_s\mathcal{F}_2)} + \mu_5\int B_6 \ . \tag{4.3}$$

Here, $G_{||}$ is the metric along the worldvolume of the anti-D3-branes and $G_\perp$ the metric along the S^2 and the form fields in the action are

$$2\pi\mathcal{F}_2 = 2\pi F_2 - C_2 \ , \qquad dB_6 = -\frac{1}{g_s}dV_4 \wedge F_3 \ . \tag{4.4}$$

The gauge field A on the worldvolume has field strength $dA = F_2$ and gives the anti-D3 charge p carried by the NS5-brane:

$$\int_{S^2} F_2 = 2\pi p \ , \tag{4.5}$$

and $F_3 = dC_2$. We are specifically interested in the effective dynamics in the angular direction ψ on the S^3, which is transverse to the NS5-brane wrapped on an S^2 inside the S^3. By expanding the kinetic term around $\psi = 0$ we obtain

$$S_{\text{NS5}} = p \ T_{\text{D3}}\int d^4x\left(-\frac{1}{2}\partial_\mu\psi\partial^\mu\psi - \frac{e^{4A_0}}{p}V(\psi)\right) \ . \tag{4.6}$$

Here $T_{\text{D3}} = \mu_3/g_s$ is the tension of an (anti)-D3-brane. The function $V(\psi)$ is an effective potential for the the azimuthal angle ψ on the S^3.[3]

$$V(\psi) = \sqrt{Q(\psi)^2 + P(\psi)^2} - Q(\psi) \ , \tag{4.7}$$

[2] Formally, the NS5-brane action is strongly coupled as it is derived by S-dualizing the D5-brane action. Nevertheless, KPV found reasonable agreement between the NS5-brane perspective and the non-Abelian anti-D3-brane perspective.

[3] Of course, this is not the potential for the canonically normalized field. We will use the canonically normalized scalar field when we derive physical quantities such as the mass.

with $b_0^2 \approx 0.93$. Here we defined the effective 3-brane charge $Q(\psi)$ as

$$Q(\psi) \equiv -\frac{1}{2\pi}\int_{S^2} \mathcal{F}_2 = \frac{M}{\pi}\left(\psi - \frac{1}{2}\sin(2\psi)\right) - p \; . \tag{4.8}$$

and

$$P(\psi) \equiv \frac{1}{4\pi^2 g_s}\int_{S_2} \sqrt{G_\perp} = \frac{b_0^2 M}{\pi}\sin^2\psi \; . \tag{4.9}$$

By expanding the potential (4.7) around $\psi = 0$, it was found by KPV that this potential has a nonsupersymmetric metastable minimum present for $p/M \lesssim 0.08$ at $\psi_{\rm min} = \frac{2p\pi}{Mb_0^4}$. In addition, the full potential also features a global supersymmetric minimum at $\psi = \pi$ where all antibranes have annihilated and linear supersymmetry is restored; see Figure 4.1.

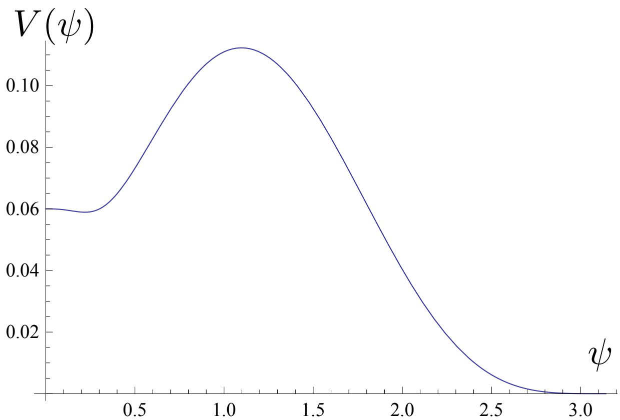

Figure 4.1: *The effective potential* (4.7) *plotted for* $p/M = 0.03$. *This potential contains a supersymmetry-breaking metastable minimum when* $p/M \ll 1$ *and a global supersymmetric minimum at* $\psi = \pi$.

KPV also described the metastable minimum from the perspective of the non-Abelian worldvolume gauge theory of the anti-D3-branes. In that case, the potential expanded around the nonsupersymmetric North Pole[4] ($\psi = 0$) is given as a function of three bosonic $p \times p$ dimensional matrix degrees of freedom $\phi^{i=1,2,3}$

$$V(\phi^i) = 2p + \frac{i}{3}\kappa\, \epsilon_{ijk}{\rm Tr}\left([\phi^i,\phi^j]\phi^k\right) - \frac{\lambda}{4}{\rm Tr}\left([\phi^i,\phi^j]^2\right) \; . \tag{4.10}$$

We will first discuss the features of this potential and after that match its parameters to the effective potential (4.7). Note that up to now we have ignored the fermionic degrees of freedom, but these will be included in the next section.

[4]Contrary to [148] we refer to the $\psi = 0$ pole of the S^3 as 'North Pole' and to the $\psi = \pi$ pole as 'South Pole'. We apologize for any confusion this might cause.

The critical points of (4.10) are given by

$$[\phi^i, \phi^j] = \frac{\kappa}{\lambda} i\epsilon^{ijk}\phi_k \ , \tag{4.11}$$

which corresponds to the commutation relations of the generators of $SU(2)$ upon rescaling $\phi^i = \frac{\kappa}{\lambda} J^i$. We see that any $SU(2)$ representation extremizes the potential, but two representations are worth mentioning specifically. The trivial representation (which has commuting matrices) corresponds to parallel anti-D3-branes and has the highest vacuum energy. The configuration of lowest energy is given by the p-dimensional irreducible representation, corresponding to the minimum where the anti-D3-branes have polarized [148]. The vacuum energy in the metastable minimum, after reinstating the correct units, is given by

$$V_{\rm min} = T_{\rm D3} e^{4A_0} p \left(2 - \frac{\kappa^4}{24\lambda^3}(p^2-1)\right) \ , \tag{4.12}$$

The radius of the S^2 in the minimum is given by

$$R^2_{S^2} = \frac{e^{-2A_0}}{p} {\rm Tr}\left(\phi^i\phi_i\right)\ell_s^2 = e^{-2A_0}\frac{\kappa^2}{4\lambda^2}(p^2-1)\ell_s^2 \ , \tag{4.13}$$

where ℓ_s is the string length.

Comparing (4.12) and (4.13) to the equivalent quantities from the Abelian NS5-brane perspective, we find

$$\kappa^2 = \frac{M^2}{p^2\pi^2}(p^2-1) \ , \quad \lambda = \frac{M^2 b_0^4}{4p^2\pi^2}(p^2-1) \ . \tag{4.14}$$

Expanding the matrices around the metastable vacuum in fluctuations φ^i as

$$\phi^i = \frac{\kappa}{\lambda} J^i + \varphi^i \ , \tag{4.15}$$

one finds a tower of fields that are labeled by $l = 1, \ldots, p-1$.[5] Details on how to explicitly diagonalize the mass matrix can be found for example in [150].

After canonically normalizing the scalar fields, the masses of the different states in the metastable minimum are given by

$$m_\varphi^2 = \frac{\kappa^2}{\lambda} l(l+1) e^{4A_0} m_s^2 = \frac{4}{b_0^4} l(l+1) e^{4A_0} m_s^2 \ , \tag{4.16}$$

with a $2(2l+1)$ multiplicity for the eigenvalues of the mass matrix for each l [151]. Notice that the masses are warped down from the string scale m_s.

[5] The $l = 0$ modes are gauge redundancies and should be omitted.

From the Abelian NS5-brane perspective we can similarly describe fluctuations around the metastable vacuum by performing a Kaluza-Klein reduction of the NS5-brane action on the S^2. One then also finds a tower of states [152] that can be matched with (4.16). This emphasizes that the effective potential derived by KPV should be understood as a (bosonic) truncation of the full NS5-brane theory that keeps only one bosonic degree of freedom. This is the azimuthal angle ψ, whose mass (using canonical normalization) around the metastable minimum is given by

$$m_\psi^2 = \frac{8}{b_0^4} e^{4A_0} m_s^2 \ . \tag{4.17}$$

This mass matches the one calculated from the non-Abelian perspective for $l = 1$, see (4.16). We want to stress once more that this degree of freedom is in fact part of a Kaluza-Klein tower of states, which will turn out to be important later. The fact that this mass does not explicitly feature the radius of the S^2, which is proportional to p/M, can be explained as follows. By expanding the potential in the metastable minimum in fluctuations, we find that the mass of the scalar fluctuation is given by $m_\psi^2 \sim \psi_{\min}^2 / R_{S^2}^2$. The overall p/M dependence therefore drops out in a small ψ expansion around $\psi = 0$.

We thus come to the conclusion that the Kaluza-Klein scale in the four-dimensional effective theory is set by the mass of this fluctuation

$$E_{\mathrm{KK}} = m_\psi \ , \tag{4.18}$$

which is warped down from the string scale. Thus, strictly speaking the effective theory in the metastable minimum can only be considered four-dimensional at energy scales $E \ll m_\psi \ll m_s$. Also note that the supersymmetry breaking scale, set by the value of the KPV potential at the metastable minimum, is of the same order as m_ψ. This strongly suggests that the restoration of linear supersymmetry hinges on the inclusion of all the Kaluza-Klein modes on the S^2, of which the ψ field is just one particular component.

Now that we have introduced all relevant details of the bosonic KPV action, we will include fermions next.

4.3 Brane polarization: fermionic action

If the metastable minimum breaks supersymmetry spontaneously, there should exist an associated massless goldstino. For a single anti-D3-brane on top of an orientifold plane, the goldstino was identified as the four-dimensional gaugino fermion on the worldvolume of the antibrane, which is a singlet under the $SU(3)$ holonomy of the six-dimensional internal space [143,144]. Removing the orientifold plane, we now want to revisit the situation for the polarized NS5-brane. Based on the bosonic truncation of the effective potential to a single bosonic single degree of freedom ψ that we discussed in the previous section, we expect the effective four-dimensional worldvolume description to reduce to the known results for p anti-D3-branes at the North Pole and $(M-p)$ D3-branes at the South Pole, both probing the KS background.

4.3.1 The fermionic action up to second order

Just as for the bosonic action, we formally obtain the fermionic NS5-brane worldvolume action from S-duality of a D5-brane. The action up to quadratic order in fermions is given by [153] (notice that we have a background with a constant dilaton)

$$S_{\rm NS5} = \frac{1}{2}\frac{\mu_5}{g_s^2}\int d^6\xi\sqrt{-\det(g+2\pi g_s\mathcal{F}_2)}\bar{\theta}(1-\Gamma_{\rm NS5})\left[(\tilde{M}^{-1})^{\alpha\beta}\hat{\Gamma}_\beta D_\alpha - \Delta\right]\theta\,, \tag{4.19}$$

where

$$\begin{aligned}\tilde{M}_{\alpha\beta} &= g_{\alpha\beta} + 2\pi g_s\sigma_3\mathcal{F}_{\alpha\beta}\,, \\ D_\alpha &= \nabla_\alpha + W_\alpha\,, \\ W_\alpha &= \frac{1}{8}\left(-F_{\alpha np}\Gamma^{np}\sigma_3 + \frac{1}{3!}g_s^{-1}H_{mnp}\Gamma^{mnp}\hat{\Gamma}_\alpha\sigma_1\right)\,, \\ \Delta &= \frac{1}{24}\left(-F_{mnp}\,\sigma_3 - g_s^{-1}H_{mnp}\,\sigma_1\right)\Gamma^{mnp}\,.\end{aligned} \tag{4.20}$$

We only included terms in the action that are non-zero at the tip of the throat, because we are not interested in dynamics taking us away from the tip (we dropped terms with five-form and one-form field strengths). The indices m, n are ten-dimensional curved indices, α, β indicate worldvolume indices. To avoid confusion with the equations below, we wrote the pullbacks of gamma matrices on the worldvolume with hats: $\hat{\Gamma}_\alpha = \Gamma_{\underline{m}} e_m{}^{\underline{m}}\partial_\alpha x^m$ and we underline tangent space indices $(\underline{m}, \underline{n}\ldots)$. The fermion θ is a doublet of Majorana-Weyl spinors with positive chirality.

We now use the specific embedding of the NS5-brane in the KS throat and use the leg structure of the three-forms to simplify the expressions. The F_3 flux is fully along the S^3 spanned by (θ, ϕ, ψ) while H_3 is orthogonal to F_3 in the internal space. This means we can drop H_3 terms with legs along the worldvolume of the NS5-brane. Also we will drop the terms with $\partial_\alpha\psi$ coming from the pullbacks of gamma matrices, as those do not contribute to the mass matrix. We only highlight the main points of the calculation here. For more general expressions and more detailed information, see appendix 4.A.

The combination in right brackets of (4.19) gives:

$$
\begin{aligned}
&(\tilde{M}^{-1})^{\alpha\beta}\Gamma_\beta D_\alpha - \Delta = \qquad (4.21)\\
&(\tilde{M}^{-1})^{\alpha\beta}\Gamma_\beta \nabla_\alpha - \frac{1}{24}\left(-\cos(2\alpha)F_{mnp}\sigma_3 + (1+\sin^2\alpha)g_s^{-1}H_{mnp}\sigma_1\right)\Gamma^{mnp}\\
&+\cos^2(\alpha)((2\pi g_s\sigma_3\mathcal{F})^{-1})^{\alpha\beta}\left(-\frac{1}{8\cdot 3!}g_s^{-1}H^{npq}\sigma_1\Gamma_{\alpha\beta npq} - \frac{1}{4}F_{\alpha\beta q}\sigma_3\Gamma^q\right),
\end{aligned}
$$

with the position-dependent angle α defined as

$$
\cos(\alpha(\psi)) \equiv -\frac{Q(\psi)}{\sqrt{Q(\psi)^2+P(\psi)^2}}\,, \qquad \sin(\alpha(\psi)) \equiv -\frac{P(\psi)}{\sqrt{Q(\psi)^2+P(\psi)^2}}\,. \qquad (4.22)
$$

It is important for our calculations to note that Γ_{NS5} is off-diagonal. As explained in appendix 4.A, at the tip of the throat, this projector takes on a fairly simple form:

$$
\Gamma_{\mathrm{NS5}} = -\begin{pmatrix} 0 & \beta_- \\ \beta_+ & 0 \end{pmatrix}, \qquad \beta_\pm = \Gamma_{\underline{0123}}(\pm\cos(\alpha) - \Gamma_{\underline{45}}\sin(\alpha))\,. \qquad (4.23)
$$

We still need to gauge fix the κ-symmetry on the brane. We do this by taking the gauge fixing condition on the doublet $\theta = (\theta_1, \theta_2)$

$$
\sigma_3\theta = -\theta \quad \Rightarrow \quad \theta_1 = 0\,. \qquad (4.24)
$$

Now we can express the action in terms of the spinor θ_2 only. This gauge fixing condition is convenient due to its simplicity, but it is not suitable when one also wants to perform an orientifold projection. The calculation for the mass matrix can also be done in a gauge where we set $(1+\Gamma_{\mathrm{NS5}})\theta = 0$, compatible with an orientifold. We show in the appendix that this choice of gauge does not change the mass matrix.

We introduce the notation for the remaining spinor components

$$
\lambda \equiv \theta_2\,. \qquad (4.25)
$$

Taking care of the off-diagonal matrix Γ_{NS5} and using that for a ten-dimensional Majorana-Weyl spinor λ the only fermion bilinears that are non-zero have three or seven gamma matrices, we find the result

$$
S_{\mathrm{NS5}} = T_{\mathrm{D3}}e^{4A_0}p\int d^4x\int\frac{d\Omega_2}{4\pi}\,\bar{\lambda}[(\tilde{M}^{-1})^{\mu\nu}\Gamma_\nu\nabla_\mu + \mathcal{M}]\lambda\,, \qquad (4.26)
$$

with $d\Omega_2$ the volume element on the unit two-sphere.

The only terms that contribute to the mass matrix $\mathcal{M}$ are

$$\mathcal{M} = \frac{1}{24}\left(\cos(2\alpha)F_{mnp} - g_s^{-1}\cos(\alpha)H_{mnp}\Gamma_{\underline{0123}}\right)\Gamma^{mnp}\,. \tag{4.27}$$

This is the mass matrix on the six-dimensional worldvolume. The reduction to four dimensions could also pick up extra mass terms coming from the reduction of the kinetic term [154]. Furthermore, just as for the bosons the reduction over the S^2 gives rise to a Kaluza-Klein tower. For now, we will ignore this tower and just focus on the lowest lying modes, but this will be important later on. To determine if the extra mass terms still allow for a massless fermion, we have to make sure the internal piece of the modified Dirac operator together with the mass matrix $[(\tilde{M}^{-1})^{\alpha\beta}\Gamma_\alpha\nabla_\beta + \mathcal{M}]$ has a zero mode, which we will check next.

4.3.2 Reduction to four dimensions

In the previous section we obtained the action for the worldvolume fermions from the six-dimensional point of view. We now discuss the four-dimensional interpretation. When we perform the reduction to four dimensions, we will write λ in terms of four fermions: a singlet λ^0 and a triplet λ^i under the $SU(3)$ holonomy of the six-dimensional transverse internal space. This decomposition can for instance be found in [144].

Let us first focus on the reduction of the mass matrix. We observe that, up to angles that parameterize the position of the NS5 on the S^3, it is completely determined by the flux of the background, which can be written in terms of the complexified three-form[6]

$$G_3 = F_3 - ig_s^{-1}H_3\,. \tag{4.28}$$

Supersymmetry of the KS background dictates that $\star_6 H_3 = -g_s F_3$ or equivalently that the complex three-form is imaginary self-dual (ISD) $G_3 = i\star_6 G_3$ [148]. This immediately implies that the only relevant structure for the fermionic mass matrix we have to reduce to four dimensions is the real part of the complex three-form:

$$\mathcal{M} = \frac{1}{48}\left(\cos(2\alpha) + \cos(\alpha)\right)(G_3 + \bar{G}_3)_{mnp}\Gamma^{mnp} \tag{4.29}$$

Up to the coordinate-dependent prefactor $(\cos(2\alpha) + \cos(\alpha))$, this is the known mass term for anti-D3-branes in a supersymmetric background with fluxes that carry only D3-brane charges, as reviewed in [144]. By evaluating the prefactor, we

[6]Notice that compared to Chapter 1 we use a different sign convention for the definition of G_3.

find

$$\begin{aligned} \psi = 0: \qquad & \mathcal{M} = \frac{1}{24}\left(G_3 + \bar{G}_3\right)_{mnp}\Gamma^{mnp}, \\ \psi = \pi: \qquad & \mathcal{M} = 0, \\ \psi = \psi_{\rm min}: \qquad & \mathcal{M} = \left(\frac{1}{24} - \frac{5}{6}\frac{\pi^2}{b_0^{12}}\frac{p^2}{M^2}\right)\left(G_3 + \bar{G}_3\right)_{mnp}\Gamma^{mnp}, \end{aligned} \tag{4.30}$$

where we expanded the last expression for small $\alpha(\psi_{\rm min}) = \frac{4\pi}{b_0^6}\frac{p}{M}$. The results at the North and South Pole match earlier results for anti-D3-branes and D3-branes on GKP backgrounds derived in [144, 155] (note that we are working in an S-dual frame compared to those references, so one should take $G_3 \to -ig_s^{-1}G_3$ for comparison to those references.)

The general discussion of our mass terms also carries through directly as in [144]. The background three-form is (2,1) and primitive, and therefore we find that the only non-zero contributions to the mass matrix can come from the triplet:

$$\bar{\lambda}\mathcal{M}\lambda = m_{ij}\bar{\lambda}^i_+\lambda^j_+ + \bar{m}_{\bar{\imath}\bar{\jmath}}\bar{\lambda}^{\bar{\imath}}_-\lambda^{\bar{\jmath}}_-, \tag{4.31}$$

where the m_{ij} are linear in the components of the background flux and $\pm$ subscripts denote four-dimensional Weyl spinors $\lambda_\pm = \frac{1}{2}(1 + i\Gamma_{\underline{0123}})\lambda$. We thus find that the mass matrix only leaves λ^0 massless, similar to a single anti-D3-brane that does not polarize [144].

The kinetic term of the fermions still contains a 'modified Dirac operator'

$$(\tilde{M}^{-1})^{\alpha\beta}\Gamma_\alpha\nabla_\beta\lambda = ((g + 2\pi g_s\sigma_3\mathcal{F})^{-1})^{\alpha\beta}\Gamma_\alpha\nabla_\beta\lambda \tag{4.32}$$

that could contribute to the mass matrix in four dimensions. We can ask whether there is a fermion that remains massless. Here, the global structure of the full KS geometry is very important. If we would just reduce a Dirac operator on an S^2 without flux, this would leave no fermion massless as a two-sphere admits no covariantly constant spinors. However, the KS geometry is an example of a Stenzel metric which is Ricci flat and therefore has a covariantly constant spinor [156]. In the reduction to four dimensions, this spinor can be identified as λ^0, for which the modified Dirac operator and the mass matrix both vanish. We therefore conclude that, irrespective of polarization dynamics, λ^0 remains massless.[7]

However, contrary to expectations for a goldstino, λ^0 is massless *everywhere* and not just in the metastable minimum. We therefore have to make sure that it also transforms in the correct way under the broken supersymmetry, before being able to unambiguously identify the gaugino as goldstino in the metastable minimum. For this reason, we will now examine the supersymmetry transformations.

[7] In [3], a slightly different reasoning was used to argue that λ^0 is massless that is based on the non-Abelian perspective. Here, we presented a complementary point of view.

4.4 Supersymmetry transformations

In the previous section we showed that, even after including polarization effects, there still exists a massless fermionic degree of freedom. If this is the goldstino associated with supersymmetry breaking, it should also transform in a standard non-linear manner under the broken supersymmetry. In this section we will show that, to leading order in an expansion around the North Pole, this massless degree of freedom transforms in the expected way. However, in the metastable minimum, corrections appear that force us to conclude that λ^0 is no longer the goldstino when polarization effects are included.

To begin, we need the expressions for the supersymmetry transformations in non-trivial flux backgrounds, which can be found in short in appendix 4.A, adapted from [153]. Supersymmetry of the background requires that

$$(1 + i\sigma_2\Gamma_{\underline{0123}})\,\epsilon = 0 \quad \Leftrightarrow \quad \epsilon_2 = \Gamma_{\underline{0123}}\epsilon_1\,. \tag{4.33}$$

With a slight abuse of notation, we will write the 32-component Majorana-Weyl spinor again as $\epsilon \equiv -2\epsilon_2$. We have the following supersymmetry transformations:

$$\begin{aligned}
\delta_\epsilon\lambda &= -\frac{1}{2}[\mathbb{1} - \beta]\epsilon + \mathcal{O}(\lambda)^2\,, \qquad (4.34)\\
\delta_\epsilon\psi &= \frac{1}{2}\bar{\lambda}\Gamma^\psi[\mathbb{1} + \beta]\epsilon + \xi^\mu\partial_\mu\psi + \mathcal{O}(\lambda)^3\,,\\
\delta_\epsilon A_\mu &= -\frac{1}{2}\bar{\lambda}(\Gamma_\mu + \Gamma_\psi\partial_\mu\psi)[\mathbb{1} + \beta]\epsilon + \frac{1}{2}C_{\mu m}\bar{\lambda}\Gamma^m[\mathbb{1} + \beta]\epsilon + \xi^\nu F_{\nu\mu} + \mathcal{O}(\lambda)^3\,,
\end{aligned}$$

with $\xi^\mu = -\frac{1}{2}\bar{\lambda}\Gamma^\mu(\mathbb{1} + \beta)\epsilon$ and the operator β defined as $\beta = \Gamma_{\underline{0123}}\beta_+$, see (4.73):

$$\begin{aligned}
\beta = -\Big(\cos(\alpha) - \Gamma_{\underline{45}}\sin(\alpha)\Big)\Big(&1 + \frac{1}{2}F_{\mu\nu}\Gamma^{\mu\nu} + \partial^\mu\psi\Gamma_{\psi\mu} - \frac{1}{2}g_{\psi\psi}\partial_\mu\psi\partial^\mu\psi \qquad (4.35)\\
&- \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{4}F_{\mu\nu}(\star_4 F)^{\mu\nu}\Gamma_{\underline{0123}} + \frac{1}{2}\partial_\rho\psi F_{\mu\nu}\Gamma^{\psi\rho\mu\nu} + \ldots\Big)\,.
\end{aligned}$$

We do not write fermion terms in β, as those result in transformations that take use beyond the quadratic fermion order in the action.

More details on these transformations can be found in appendix 4.A. From here, we can already see the general form of the transformations around the poles, since

$$\begin{aligned}
\text{at } \psi = 0: &\quad \beta = -\mathbb{1} + \ldots\,, \qquad (4.36)\\
\text{at } \psi = \pi: &\quad \beta = +\mathbb{1} + \ldots\,,
\end{aligned}$$

where the ellipses denote terms with field fluctuations. So around $\psi = 0$ we find non-linear transformations and at $\psi = \pi$ linear ones.

To obtain four-dimensional supersymmetry transformations, in the end we always decompose the spinor into the singlet λ^0 and the triplet λ^i under the $SU(3)$ holonomy. Moreover we can focus on just one of the triplet fermions, say $i = 3$, due to the arbitrary orientation of the S^2 inside the transverse S^3, corresponding to the superpartner of the scalar ψ at the south pole where supersymmetry is restored. The other directions come along for the ride and we can ignore them throughout. We are also interested in the supersymmetry transformations with parameter ϵ^0, the $SU(3)$ singlet component of the 32-component Majorana-Weyl spinor ϵ, as this is the supersymmetry preserved by the background.

With all the relevant information in place, we present a summary of the four-dimensional fermionic, scalar and gauge field supersymmetry transformations at the different locations of interest: both poles and most importantly the metastable minimum.

4.4.1 At the South Pole

Let us first analyze the South Pole $\psi = \pi$, where the D3-branes do not break the background $N = 1$ supersymmetry. We obtain to leading order in fluctuations the expression for β:

$$\beta = \mathbb{1} + \frac{1}{2}F_{\mu\nu}\Gamma^{\mu\nu} + \partial_\mu\psi\Gamma_\psi\Gamma^\mu \, , \tag{4.37}$$

and the reduction of the supersymmetry transformations to four dimensions gives

$$\begin{aligned} \delta_\epsilon\lambda^0 &= \frac{1}{4}\gamma^{\mu\nu}F_{\mu\nu}\epsilon^0 \, , \\ \delta_\epsilon\lambda^3 &= \frac{1}{\sqrt{2}}\gamma^\mu\partial_\mu\tilde{\psi}\epsilon^0 \, , \\ \delta_\epsilon\tilde{\psi} &= \frac{1}{\sqrt{2}}\bar{\lambda}^3\epsilon^0 \, , \\ \delta_\epsilon A_\mu &= -\frac{1}{2}\bar{\lambda}^0\gamma_\mu\epsilon^0 \, , \end{aligned} \tag{4.38}$$

where we redefined the scalar as follows.

$$\tilde{\psi} = -e^{\frac{\psi}{\psi}}\psi = -(g_s M b_0^2)^{1/2}\psi \, , \tag{4.39}$$

and rescaled spinors as $\lambda \to \frac{1}{\sqrt{2}}\lambda$, $\epsilon \to \frac{1}{\sqrt{2}}\epsilon$. We conclude that, as expected, at $\psi = \pi$ a linearly realized $N = 1$ supersymmetry exists under which (λ^0, A_μ) form a vector multiplet and (λ^3, ψ) correspond to a chiral multiplet. If we would have included the other two directions on the S^2 that we now have ignored, they would form two additional chiral multiplets. Those correspond to the decomposition of the spectrum of the $N = 4$ supersymmetric Yang-Mills (SYM) multiplet on the D3-brane transforming under the $N = 1$ supersymmetry of the background.

4.4.2 At the North Pole

At the (unstable) North Pole we expect the effective description to formally reduce to the results for a supersymmetry-breaking anti-D3-brane in the KS background. We will write the transformations to at most quadratic order in field fluctuations. Since $\sin(\alpha) = \mathcal{O}(\psi^2), \cos(\alpha) = 1 + \mathcal{O}(\psi^4)$, we set $\cos(\alpha) = 1$, as the subleading terms will come in at higher order in the supersymmetry transformations. Then we indeed reproduce to quadratic order the results of [146].

We will expand the supersymmetry transformations up to the first non-trivial order in the fields. Then we only have to expand the operator β to first order:

$$\beta = -\mathbb{1} - \frac{1}{2} F_{\mu\nu}\Gamma^{\mu\nu} - \partial^\mu\psi\Gamma_{\psi\mu} + \dots \,. \tag{4.40}$$

The supersymmetry transformations around $\psi = 0$ are

$$\begin{aligned}
\delta_\epsilon\lambda &= -\epsilon - \frac{1}{4}F_{\mu\nu}\Gamma^{\mu\nu}\epsilon - \frac{1}{2}(\partial_\mu\psi)\Gamma_\psi\Gamma^\mu\epsilon + \mathcal{O}(\phi^2)\,, \\
\delta_\epsilon\psi &= -\frac{1}{2}(\bar\lambda\Gamma^\mu\epsilon)\partial_\mu\psi - \frac{1}{4}(\bar\lambda\Gamma^{\psi\mu\nu}\epsilon)F_{\mu\nu} + \mathcal{O}(\phi^3)\,, \\
\delta_\epsilon A_\mu &= -\frac{1}{2}(\bar\lambda\Gamma^\rho\epsilon)F_{\rho\mu} - \frac{1}{2}(\bar\lambda\Gamma_\psi\epsilon)\partial_\mu\psi + \frac{1}{4}(\bar\lambda\Gamma_{\mu\rho\sigma}\epsilon)F^{\rho\sigma} + \frac{1}{2}(\bar\lambda\Gamma_{\psi\rho\mu}\epsilon)\partial^\rho\psi + \mathcal{O}(\phi^3)\,,
\end{aligned} \tag{4.41}$$

with ϕ the collection of all fields $\phi = \{\psi, \lambda, A_\mu\}$. We recognize the first terms as the standard non-linear transformations. By requiring the fields to transform non-linearly under the supersymmetry we can perform appropriate field redefinitions of the spinors, scalar and gauge field, that fix the transformations uniquely:

$$\begin{aligned}
\tilde\lambda &= -\lambda + \frac{1}{4}F_{\mu\nu}\Gamma^{\mu\nu}\lambda + \frac{1}{2}(\partial_\mu\psi)\Gamma_\psi\Gamma^\mu\lambda + \mathcal{O}(\phi^3)\,, \\
\tilde\psi &= \psi - \frac{1}{8}(\bar\lambda\Gamma^{\psi\mu\nu}\lambda)F_{\mu\nu} + \mathcal{O}(\phi^4)\,, \\
\tilde A_\mu &= A_\mu - \frac{1}{4}(\bar\lambda\Gamma_\psi\lambda)\partial_\mu\psi + \frac{1}{8}(\bar\lambda\Gamma_{\mu\rho\sigma}\lambda)F^{\rho\sigma} + \frac{1}{4}(\bar\lambda\Gamma_{\psi\rho\mu}\lambda)\partial^\rho\psi + \mathcal{O}(\phi^4)\,,
\end{aligned} \tag{4.42}$$

and we have the standard-looking transformations

$$\begin{aligned}
\delta_\epsilon\tilde\lambda &= \epsilon + \mathcal{O}(\phi^2)\,, \\
\delta_\epsilon\tilde\psi &= \frac{1}{2}(\bar{\tilde\lambda}\Gamma^\mu\epsilon)\partial_\mu\tilde\psi + \mathcal{O}(\phi^3)\,, \\
\delta_\epsilon\tilde A_\mu &= \frac{1}{2}(\bar{\tilde\lambda}\Gamma^\rho\epsilon)\tilde F_{\rho\mu} + \mathcal{O}(\phi^3)\,.
\end{aligned} \tag{4.43}$$

With an additional rescaling of the spinors $\tilde\lambda \to \sqrt{2}\tilde\lambda, \epsilon \to \sqrt{2}\epsilon$, we then find the following supersymmetry transformations in terms of the appropriate four-dimensional fields around $\psi = 0$

$$\begin{aligned}\delta_\epsilon \tilde{\lambda}^0 &= \epsilon^0 + \mathcal{O}(\phi^2) \\ \delta_\epsilon \tilde{\lambda}^3 &= 0 + \mathcal{O}(\phi^2) \\ \delta_\epsilon \tilde{\psi} &= (\bar{\tilde{\lambda}}^0 \gamma^\mu \epsilon^0)\partial_\mu \tilde{\psi} + \mathcal{O}(\phi^3) \\ \delta_\epsilon \tilde{A}_\mu &= (\bar{\tilde{\lambda}}^0 \gamma^\nu \epsilon^0)\tilde{F}_{\nu\mu} + \mathcal{O}(\phi^3)\,.\end{aligned} \tag{4.44}$$

We conclude that this seems to describe an exact non-linear realization of (broken) supersymmetry when adding anti-D3-branes to the KS background and ignoring the (higher-order) dynamics describing the polarization in the transverse S^3 directions. This matches the results for a single anti-D3-brane in the supersymmetric background of [143–146]. Note that this (direct) expansion of the theory around the north pole is only a *formal* result: since the scalar field ψ sits at the maximum of its potential, this is an expansion around an unstable configuration.

4.4.3 At the metastable minimum

Now let us include the polarization dynamics and determine the transformations at the metastable minimum ψ_{min}, which should include corrections due to the dynamics on the S^3. We first expand in ψ and then in the fluctuations around the metastable minimum. The expansion for α around the metastable minimum is then

$$\alpha(\psi_{\text{min}} + \delta\psi) = \frac{4\pi}{b_0^6}\frac{p}{M} + \frac{4}{b_0^2}\delta\psi + \frac{b_0^2}{\pi}\frac{M}{p}\delta\psi^2 + \ldots \tag{4.45}$$

The leading corrections in the expansions of ψ-fluctuations and powers of p/M are then captured by expanding β in powers of α:

$$\beta = \Big(1 - \alpha\Gamma_{\underline{45}} - \frac{1}{2}\alpha^2 + \ldots\Big)\beta|_{\psi=0}\,, \tag{4.46}$$

where $\beta|_{\psi=0}$ is given by (4.40).

We find that after the field redefinition (4.42) and the spinor rescalings the transformations (4.43) are corrected by the α-expansion (or equivalently ψ-expansion):

$$\begin{aligned}\delta_\epsilon \tilde{\lambda} &= \delta_\epsilon \tilde{\lambda}|_{\psi=0} - \frac{1}{2}\alpha\Gamma_{\underline{45}}\epsilon - \frac{1}{4}\alpha^2\epsilon + \ldots\,, \\ \delta_\epsilon \tilde{\psi} &= \delta_\epsilon \tilde{\psi}|_{\psi=0} - \alpha\bar{\tilde{\lambda}}\Gamma^{\underline{45}\psi}\epsilon - \frac{1}{2}\alpha^2\bar{\tilde{\lambda}}\Gamma^{\psi}\epsilon + \ldots\,, \\ \delta_\epsilon \tilde{A}_\mu &= \delta_\epsilon \tilde{A}_\mu|_{\psi=0} + \alpha\bar{\tilde{\lambda}}\Gamma_{\underline{45}}\Gamma_\mu\epsilon + \frac{1}{2}\alpha^2\bar{\tilde{\lambda}}\Gamma_\mu\epsilon + \ldots\,.\end{aligned} \tag{4.47}$$

The transformations in the metastable minimum become

$$
\begin{aligned}
\delta_\epsilon \tilde{\lambda}^0 &= \epsilon^0 - \frac{1}{4}\alpha^2 \epsilon^0 + \dots\,, \\
\delta_\epsilon \tilde{\lambda}^3 &= 0 - \alpha \epsilon^0 + \dots\,, \\
\delta_\epsilon \tilde{\psi} &= (\bar{\tilde{\lambda}}^0 \gamma^\mu \epsilon^0)\partial_\mu \tilde{\psi} - 2\sqrt{2} e^{\psi}_{\underline{\psi}} (\alpha \bar{\tilde{\lambda}}^0 \epsilon^0 + \frac{1}{4}\alpha^2 \bar{\tilde{\lambda}}^3 \epsilon^0) + \dots\,, \\
\delta_\epsilon \tilde{A}_\mu &= (\bar{\tilde{\lambda}}^0 \gamma^\nu \epsilon^0)\tilde{F}_{\nu\mu} + 2\alpha \bar{\tilde{\lambda}}^0 \gamma_\mu \epsilon^0 + \frac{1}{2}\alpha^2 \bar{\tilde{\lambda}}^3 \gamma_\mu \epsilon^0 + \dots
\end{aligned}
\tag{4.48}
$$

The first terms correspond to the standard non-linear transformations, but we find two types of corrections. First of all we observe that there are corrections that vanish in the probe limit $p/M \to 0$. These terms are just proportional to (the square of) $\psi_{\min} \sim p/M$ and reflect the shift towards the metastable minimum. In fact, if we could ignore the field $\delta\psi$ (as well as the spinor λ^3), the probe limit would consistently reproduce a subset of the non-linear supersymmetry transformations at the North Pole. In other words, if the $\delta\psi$ and λ^3 fields were infinitely massive, the probe limit takes you to the North Pole and a constrained superfield description in terms of just λ^0 and A_μ would be adequate.

However, when we keep $\delta\psi$ in the spectrum to capture polarization effects, we find that the corrections to the non-linear supersymmetry transformations cannot seem to be consistently decoupled. The reason for this is that the corrections that are proportional to $(\delta\psi)^2$ are all, except for the gaugino, proportional to M/p suggesting that in the probe limit corrections become large and one should include (all) higher-order terms. This therefore signals that for small p/M, as required for the existence of a metastable minimum, a constrained superfield description truncated to $\delta\lambda^0, \delta A_\mu, \delta\lambda^3$ and $\delta\psi$ breaks down.

This should not come as a complete surprise. As we already saw in section 4.2, a proper reduction to four dimensions over the S^2 involves a Kaluza-Klein tower of fields that we have ignored so far. As we will argue next, such a truncation is inconsistent because higher-dimensional physics is crucial to restore linear supersymmetry and the origin of supersymmetry breaking cannot be understood as being purely four-dimensional. To show this, we will construct a model in which the metastable state is embedded into a linearly supersymmetric theory. This allows us to see explicitly how the metastable state emerges from a supersymmetric theory. Using this model, we are then finally able to conclude that the metastable state indeed breaks supersymmetry spontaneously and that the appearance of a massless goldstino crucially hinges on the presence of the Kaluza-Klein tower.

4.5 Supersymmetric embedding of antibrane polarization

Before we present our linearly supersymmetric embedding, let us briefly explain why the probe actions used so far (the non-Abelian anti-D3-action and the truncated NS5-brane action) are insufficient to find a supersymmetric embedding of the metastable minimum.

4.5.1 Probes cannot restore supersymmetry

To illustrate this point, let us take a step back and consider a single probe anti-D3-brane. The worldvolume theory of an anti-D3-brane (in flat space) has 32 supersymmetries, 16 of which are linearly realized and the other 16 nonlinearly; see for example [157]. The supersymmetries that are preserved by the antibrane are solutions to [143]

$$(1 - \Gamma_{\overline{\mathrm{D3}}})\varepsilon = 0 \ , \tag{4.49}$$

where $\Gamma_{\overline{\mathrm{D3}}}$ is the κ-symmetry projector of an anti-D3-brane. On the other hand, because the background we are interested in contains an O3 orientifold projection, the supersymmetries that are preserved by the background are solutions to [143]

$$(1 - \Gamma_{\mathrm{O3}})\varepsilon = 0 \ , \tag{4.50}$$

where Γ_{O3} is the action of an O3 orientifold projection. This background orientifold is not to be confused with putting the anti-D3-brane on top of an orientifold plane as was done in [143,144], which also projects out all bosonic degrees of freedom on the worldvolume. For an anti-D3-brane probing the KS background, $\Gamma_{\overline{\mathrm{D3}}} = -\Gamma_{\mathrm{O3}}$, which shows that the linear supersymmetries on the anti-D3-brane are projected out by the orientifold and only the non-linear supersymmetries survive. As a result, all degrees of freedom on the antibrane transform in the standard non-linear manner under supersymmetry [145,146], but there is no chance of restoring linear supersymmetry as the linear supersymmetries are projected out.

One might expect that the situation is better from the perspective of the NS5-brane action truncated to four dimensions and ignoring the Kaluza-Klein tower, because the effective potential (4.7) connects the supersymmetry breaking state to the global supersymmetric vacuum. In this picture however, the supersymmetries preserved by the NS5 brane are solutions to

$$(1 - \Gamma_{\mathrm{NS5}}(\psi))\varepsilon = 0 \ . \tag{4.51}$$

As can be seen from (4.23), at the poles of the S^3 $\Gamma_{\rm NS5}(\psi)$ reduces to

$$\Gamma_{\rm NS5} = \begin{cases} \Gamma_{\overline{\rm D3}} & (\psi = 0) \ , \\ \Gamma_{\rm D3} & (\psi = \pi) \ , \end{cases} \tag{4.52}$$

but away from the poles it does not align nicely with the κ-symmetry projector of an (anti-)D3-brane. So in general, away from the poles, the (reduced) NS5-brane projector seems to break supersymmetry explicitly, as it does not transform in the standard way under (non-linear) supersymmetry. We conclude that if a supersymmetric theory exists in which the metastable state is a solution it cannot be described by one of the truncated KPV four-dimensional probe actions used so far. The action of probe anti-D3-branes only preserves non-linear supersymmetry and has no obvious connection to the linearly supersymmetric regime. The NS5-brane action on the other hand is linearly supersymmetric at $\psi = \pi$, but away from this pole breaks supersymmetry explicitly. This suggests to us that both probe descriptions apparently do not contain enough degrees of freedom to allow for the restoration of supersymmetry. To restore linear supersymmetry we will be forced to include additional (massive) degrees of freedom on the S^2, which are absent in the reduced four-dimensional KPV probe descriptions.

We will use that, at the supersymmetric South Pole the physics is that of $(M - p) \gg 1$ D3-branes, and we will expand around that supersymmetric point to get information on the specific deformations that are needed to describe the metastable state. This allows us to go beyond the level of the probe action and propose a supersymmetric model of $(M - p)$ D3-branes in the KS background, obtained by introducing additional degrees of freedom from the S^2 and specific irrelevant deformations of the $N = 4$ SYM theory. Adding these degrees of freedom and deformations the supersymmetric model then indeed features a metastable vacuum state in which supersymmetry is spontaneously broken.

4.5.2 Towards a supersymmetric embedding

We just argued that neither of the probe actions used by KPV preserve linear supersymmetry when placed inside the KS background. So if we think that anti-D3-brane polarization can nevertheless be embedded in a supersymmetric effective field theory model, how are we going to identify that theory?

Of course, any D-brane state is a solution of superstring theory, so it might be the case that an explicit description in terms of spontaneous breaking of supersymmetry is only possible in the full ten-dimensional superstring theory. This is a possibility, but one that is contrary to expectations in this particular KPV setup. Because the supersymmetry breaking scale is warped down from the string scale [158] one would expect that the metastable state can be embedded in a lower-dimensional effective supersymmetric theory. In particular, if backreaction of anti-D3-branes does not destroy the metastable state and a well-defined supergravity solution exists [55], it must be realizable as a state in a supersymmetric field theory that is (holographically) dual to the supergravity solution [159].

For example, before adding anti-D3-branes to the KS geometry, the holographic dual of KS is given by a nonconformal $N = 1$ cascading gauge theory [147]. The effect of antibranes can then be included by adding a particular nonsupersymmetric perturbation to the cascading gauge theory [159] and it was shown in subsequent work [160, 161] that the resulting gauge theory indeed contains a massless fermion, as expected if the supersymmetry breaking was spontaneous. Although this is suggestive, in these approaches the perturbations describing anti-D3-branes still explicitly break supersymmetry. Our goal here is to instead present an effective fully supersymmetric model in which the polarized antibrane state appears as a metastable nonsupersymmetric solution.

Indications for how to construct such a model can be obtained from the works [51, 54, 56] (for related work, see [162]). Here, the authors argued that, close to the nonsupersymmetric $\psi = 0$ pole, anti-D3-branes source an $AdS_5 \times S^5$ throat perturbed by flux that is dual to relevant deformations of $N = 4$ SYM that break all supersymmetry. Hence, the dual gauge theory that describes these antibranes is a nonsupersymmetric version of the $N = 1^\star$ theory of Polchinski and Strassler [163]. Obviously, because we are interested in finding a supersymmetric starting point, we will not try to identify the polarized state in a theory obtained as an expansion around the nonsupersymmetric ($\psi = 0$) pole, but instead we will start from the supersymmetric ($\psi = \pi$) pole. Here, the physics should be that of $(M - p)$ D3-branes that source a supersymmetric $AdS_5 \times S^5$ throat dual to $N = 4$ SYM. To describe polarization, we suggest that one should add irrelevant deformations to $N = 4$ SYM that have the effect of gluing the $AdS_5 \times S^5$ throat back to the KS region in the UV [164]. In fact, an expansion of the effective potential (4.7) around the supersymmetric pole of the S^3 reveals that, as anticipated, the first term in the effective potential is ψ^4 (which corresponds to the SYM term in the worldvolume gauge theory) and the leading corrections are irrelevant operators (in four dimensions) of mass dimension 6 and 7.

We will provide evidence below that a manifestly supersymmetric model that includes matrix degrees of freedom on the S^2 and irrelevant operators of mass dimension 6 and 7 indeed features a nonsupersymmetric metastable vacuum state with the expected properties. Obviously, this model is finetuned in the sense that all other irrelevant operators should be suppressed; that is, it crucially relies on the details of the UV embedding. Nevertheless, if we extrapolate this model away from the supersymmetric pole, we discover a metastable vacuum that has the same properties as the metastable state found by KPV. It breaks supersymmetry spontaneously by a nonzero F-term and as a consequence features a massless goldstino. This fermion is only massless at the metastable minimum, as appropriate for a goldstino, and therefore provides an additional low-energy degree of freedom on top of the massless gaugino residing in the vector multiplet.

4.5.3 Supersymmetric model

The field content of a stack of $\tilde{M} = (M - p)$ D3-branes in four dimensions is given by three matrix-valued chiral multiplets $\Phi^{i=1,2,3}$ and a matrix-valued vector multiplet. We will ignore the vector multiplet and just focus on the three chiral multiplets, which is sufficient for our purposes. The supersymmetric model we propose to describe the metastable state with is defined by the following superpotential

$$\begin{aligned} W(\Phi^i) = &\frac{a}{3!}\epsilon_{ijk}\mathrm{Tr}\left([\Phi^i,\Phi^j]\Phi^k\right) \\ &+ i\frac{b}{5}\epsilon_{ijk}\mathrm{Tr}\left([\Phi^i,\Phi^j][\Phi^k,\Phi^l]\Phi_l\right) \\ &+ \frac{c}{6}\epsilon_{ijk}\epsilon_{lmn}\mathrm{Tr}\left([\Phi^i,\Phi^j][\Phi^l,\Phi^m][\Phi^k,\Phi^n]\right) \ , \end{aligned} \tag{4.53}$$

and a canonical Kähler potential. The first line of (4.53) is the $N = 4$ SYM term and the second and third lines are the irrelevant deformations, that break supersymmetry from $N = 4 \to 1$. Locally, the metric on field space can be approximated by the flat metric δ_{ij}[8] and we take all coefficients to be real.

The scalar potential is given by

$$V_S = \mathrm{Tr}\left(\frac{\partial W}{\partial \Phi^i}\frac{\partial \overline{W}}{\partial \overline{\Phi}_i}\right) \ . \tag{4.54}$$

To find its critical points, we take the following ansatz for the commutation relations of the scalar fields

$$[\phi^i,\phi^j] = i\epsilon^{ijk}\phi_k \ , \tag{4.55}$$

which are the commutation relations of the generators of $SU(2)$. The trivial representation corresponds to parallel D-branes, representing a vacuum with vanishing energy that makes it supersymmetric. In addition, there can be two other types

[8] We believe our results can easily be generalized beyond this approximation, but it will be sufficient for our purposes here.

of vacua that obey (4.55), depending on the choice of parameters. They are given by

$$\text{Type I:} \quad a + 2b - 4c = 0 \; , \tag{4.56}$$
$$\text{Type II:} \quad a + 4b - 10c = 0 \; . \tag{4.57}$$

The vacuum energy in these two vacua after reinstating units is given by

$$\text{Type I:} \quad V_{\text{min}} = 0 \; , \tag{4.58}$$
$$\text{Type II:} \quad V_{\text{min}} = T_{D3} e^{4A_0} \left(\frac{\tilde{M}}{100} (\tilde{M}^2 - 1)(3a + 2b)^2 \right) \; .$$

We see that the type I vacuum has vanishing vacuum energy (and vanishing F-terms) and therefore corresponds to a supersymmetric state, similar to the supersymmetric polarized states in the $N = 1^\star$ theory [163]. The type II vacuum, however, has nonvanishing F-terms, and positive vacuum energy and it necessarily breaks supersymmetry spontaneously. Upon comparing the vacuum energy with (4.12), we find

$$a = -\frac{2}{3} b + \mathcal{O}(\epsilon) \; , \tag{4.59}$$

where $\epsilon = \sqrt{p/M^3} \ll 1$ is a small parameter. By expanding in fluctuations around the type II vacuum as $\phi^i = J^i + \varphi^i$, we again find a tower of scalar fields. The mass of the lightest fluctuation is given by

$$m_\varphi^2 \simeq \frac{24}{5} \left(a\epsilon - \frac{4}{3}\epsilon^2 \right) e^{4A_0} m_s^2 \; , \tag{4.60}$$

where we used (4.59). This result matches the mass of the lightest fluctuation derived from the non-Abelian KPV potential given by (4.17) when we identify

$$a = \frac{5}{3 b_0^4 \epsilon} + \mathcal{O}(\epsilon) \; . \tag{4.61}$$

Thus, all fluctuations have a positive mass squared when $p/M^3 \ll 1$. Notice that this condition is weaker than, but consistent with, the condition $p/M \lesssim 0.08$ derived in [148] for a metastable minimum to exist. We conclude that the superpotential (4.53) reproduces the main features of the effective potential derived by KPV. As already mentioned, we should stress that since this crucially relies on the inclusion of two (supersymmetric) irrelevant corrections, the model depends sensitively on the UV embedding (in string theory).

In addition, because the metastable vacuum breaks supersymmetry spontaneously it should have a massless fermion in its spectrum corresponding to the goldstino of supersymmetry breaking. By computing the determinant of the fermionic mass matrix[9]

$$\det(M_F^{ij}) = \det\left(\frac{\partial^2 W}{\partial\Phi_i\partial\Phi_j}\right) \ , \tag{4.62}$$

we indeed find that this equals zero exactly when (4.57) is satisfied signalling that, precisely in the metastable minimum, a fermion becomes massless as expected for a goldstino.

Obviously, one should also be able to identify this goldstino from the perspective of a non-truncated Abelian NS5-brane description expanded around the supersymmetric pole. There it should correspond to a fermionic zero mode on the S^2 threaded by magnetic flux (the presence of flux twisting the S^2 is crucial to allow for a zero mode [150, 166]), but since the degrees of freedom are organized differently in the Abelian NS5-brane perspective identifying the goldstino is not straightforward and it would be of interest to confirm its presence. Whether such a fermionic zero mode should also be present in the non-Abelian anti-D3-brane description expanded around the nonsupersymmetric pole, as was considered in [158], is not obvious *a priori*, but it should clearly not be associated with the gaugino when polarization effects are taken into account. In the absence of a supersymmetric embedding the appearance of a goldstone fermion should be expected to be difficult at best.

To summarize, the proposed model allows for a supersymmetric description of brane polarization. It describes both a supersymmetric vacuum corresponding to $\tilde{M}$ parallel D3-branes and a metastable vacuum. The metastable vacuum breaks supersymmetry spontaneously and its spectrum of fluctuations contains the four-dimensional vector multiplet, a massless goldstino and a massive tower of Kaluza-Klein states that, when included, allow for full restoration of supersymmetry. The supersymmetry breaking scale $\sqrt{f}$ is thus of the order of the Kaluza-Klein scale

$$\sqrt{f} \simeq E_{\rm KK} = m_\varphi \ , \tag{4.63}$$

as can be seen by comparing the potential energy and the mass of the lightest fluctuation. As a consequence, below the supersymmetry breaking scale one can integrate out the massive fields such that a constrained superfield description in terms of a single nilpotent superfield that contains the goldstino is a good approximation for the dynamics on the S^2 and supersymmetry is realized nonlinearly. At or above this scale, this description will be modified and one has to take the Kaluza-Klein modes on the S^2 into account. Because the Kaluza-Klein scale is

[9] Because we are working with matrix-valued fields, the fermion mass matrix has four indices. To compute the determinant this tensor first needs to be decomposed to obtain a regular $3N^2 \times 3N^2$ matrix of which we can calculate the determinant. The details of this procedure can be found for example in [165].

warped down from the string scale, as noted in section 4.2, the energy at which one needs to include the Kaluza-Klein modes can be very low. Depending on the details realizing a hierarchy between the energy scale one is probing and the warped down Kaluza-Klein scale might therefore be difficult, but not impossible.

4.6 Conclusions

Constrained superfields provide a powerful technique in the context of a universal (UV-insensitive) low-energy description of spontaneously broken supersymmetry. A crucial requirement is a stable and large enough hierarchy between the scale of the fields that are projected out by the constraints and the relevant scale of the low-energy effective theory. In some cases such a hierarchy might not be achievable, precluding the existence of a standard constrained superfield description. In general however the appropriate constrained superfield description is valid up to an energy scale that should be carefully identified. In this chapter, we studied corrections to the nilpotent goldstino superfield description of anti-D3-branes in the KS background arising from polarization effects.

To do so, we used the four-dimensionally reduced effective theory on the NS5-brane wrapping an S^2 inside the transverse S^3 at the tip of the KS throat geometry of [148]. One of our main observations is that the (non-linear) supersymmetry transformations in the metastable vacuum receive corrections that cannot be decoupled and actually become large in the probe limit $p/M \to 0$. While this theory still features a massless fermion (the gaugino) that one might be tempted to identify as the goldstino, we argued that the corrections to the supersymmetry transformations signal a breakdown of the constrained superfield description in terms of the degrees of freedom on the truncated NS5-brane.

We also gave an interpretation of these results. When the source of supersymmetry breaking is intrinsically higher-dimensional, it might not admit any low-energy description in terms of (simple) constrained superfields. Indeed, we argued that the effective four-dimensional potential used by KPV does not include the Kaluza-Klein tower from reducing over the S^2 to four dimensions and the breakdown of the constrained superfield description seems to hint that this truncation is inconsistent.

To confirm that supersymmetry breaking nevertheless is spontaneous when higher-dimensional physics is involved, we set out to identify a linearly supersymmetric field theory in which the metastable vacuum state can be embedded. Our proposal for an effective supersymmetric field theory model indeed exhibits such a metastable supersymmetry-breaking vacuum with all the expected properties. Besides the fact that we ignored gravity by working in the noncompact KS geometry (effectively sending $M_{\text{Pl}} \to \infty$), this construction is also finetuned; that is, it seems to depend sensitively on the precise UV theory in which it is embedded due to the introduction of irrelevant operators. Keeping those limitations in mind the qualitative features of the metastable solution nicely agree with the polarized state in the KPV model. Expanding around the metastable solution we observed that the restoration of supersymmetry crucially relies on taking into account a tower of Kaluza-Klein modes, with masses warped down from the string scale, explaining why it is impossible for the truncated probe actions to restore supersymmetry.

Our results have a number of consequences. First of all, the crucial importance of Kaluza-Klein modes on the S^2 confirms that the KPV effective potential should only be thought of as a truncation of a more complete description. In other words, the additional degrees of freedom related to the S^2 only decouple at the poles and should in general be included in the four-dimensional effective theory. This impacts models derived from KPV's effective potential, such as the inflationary model recently considered in [167,168], where the effective potential (4.7) was used for large field inflation in the regime $p/M \gg 1$. Our results clearly suggest that it is inconsistent to rely on just the scalar mode ψ for the effective dynamics on the S^2 and one should include (a subset of) a Kaluza-Klein tower of states. As a result one would expect the results reported in [167, 168] to be affected. We note that an inflationary model that is described from the perspective of scalar matrix degrees of freedom, albeit in a different context, already has been studied in [169]. In light of the results obtained here it might be of interest to revisit some of these approaches.

Secondly, the decay rate to tunnel from the metastable state to the global supersymmetric vacuum was calculated in [148] by making use of the effective potential (4.7). However, in the presence of additional degrees of freedom this tunneling rate will likely be modified. For example, it is well known that a coupling between a quantum mechanical system and its environment can lead to a significant suppression of the tunneling rate [170, 171]. Furthermore, additional degrees of freedom can also open up new decay channels; see for example [172] and references therein. If and how these effects modify the lifetime of the metastable vacuum is a question that would be interesting to come back to in future work.

Our results also shed light on how the anti-D3-brane uplift procedure in KKLT might be captured by an effective supersymmetric theory. To be specific, we claim that the constrained superfield description proposed in [143,144], with just a single nilpotent superfield containing the goldstino, is certainly a valid description far below the warped down Kaluza-Klein scale. However, to understand this in terms of spontaneous supersymmetry breaking and identify leading corrections one needs to identify the supersymmetric completion of this nonsupersymmetric phase. This requires the introduction of additional degrees of freedom at the warped down Kaluza-Klein scale. The effective field theory model that we introduced to describe the dynamics on the compact S^2 is a good candidate for a linearly supersymmetric theory in which the polarized state appears as a nonsupersymmetric solution. This model does not contain degrees of freedom at the string or Planck scale, so it can be studied as a supersymmetric low-energy effective field theory, but it does depend sensitively on the specific UV embedding (in string theory).

Finally, we would like to comment on how our results are related to the debate in the literature regarding the existence of the KPV metastable state after taking into account the full backreaction of anti-D3-branes. As mentioned, the nonexistence of a fully backreacted supergravity solution would imply that supersymmetry breaking in the holographically dual gauge theory description is necessarily explicit. Since the effective supersymmetric field theory model that we introduced nicely allows for the appearance of a nonsupersymmetric phase with the expected properties to relate it to antibrane polarization, this suggests that a fully backreacted supergravity solution should exist, in line with results of [55, 59], assuming such a holographic duality.

However, even if the supersymmetric model constructed is holographically dual to this fully backreacted supergravity solution, it is clearly not UV complete and requires a specific UV embedding. Whether or not such an embedding is possible in string theory will determine its ultimate fate. So with these results we can certainly not rule out the possibility that the KPV metastable state and therefore the KKLT de Sitter vacua, after including gravity, belong to the swampland, as was recently conjectured [10, 34, 87, 88, 123]. As a consequence, understanding the UV completion of the proposed model is an important direction for further research.

4.A Details on fermions

In this appendix we review and apply the relevant details of the fermionic action of a Dp-brane of [153, 173, 174], its supersymmetry transformations and gauge fixing. We take the results for a D5 brane with worldvolume flux in the S-dual background to Klebanov-Strassler. We follow the conventions of [153]. For easy comparison with the literature on gauge-fixed fermionic D-brane actions, we keep this appendix wholly in that 'D5-frame' and we adapt notation slightly to match as much as possible the related work for Dp-branes in flat space [157] used in the recent literature on non-linear supersymmetries on anti-D3-branes [144–146, 175].

To transform the results of this appendix ('app') to the expressions used in the text, one has to apply the following S-duality rules to the NS5-frame:

$$\begin{aligned} H_3^{\text{app}} &= -F_3^{\text{text}}, & F_3^{\text{app}} &= H_3^{\text{text}}, \\ e^{\Phi^{\text{app}}} &= (g_s^{-1})^{\text{text}}, & \mathcal{F}^{\text{app}} &= 2\pi g_s^{\text{text}} \mathcal{F}^{\text{text}}. \end{aligned} \tag{4.64}$$

4.A.1 Projection matrix

We obtain the matrix Γ_{D5} from [153]:

$$\Gamma_{\text{D5}} = -\begin{pmatrix} 0 & \beta_- \\ \beta_+ & 0 \end{pmatrix}, \tag{4.65}$$

with

$$\beta_\pm = \Gamma_{\text{D5}}^{(0)} \frac{\sqrt{-\det g}}{\sqrt{-\det(g+\mathcal{F})}} \sum_k \frac{(\pm 1)^k}{k! 2^k} \hat{\Gamma}^{\alpha_1 \ldots \alpha_{2k}} (\mathcal{F})_{\alpha_1\alpha_2} \cdots (\mathcal{F})_{\alpha_{2k-1}\alpha_{2k}}. \tag{4.66}$$

We have $\beta_+\beta_- = 1$ and the relation $\beta_-(\mathcal{F}) = \beta_+(-\mathcal{F})$. Note that hats on gamma matrices denote pullbacks on the worldvolume $\hat{\Gamma}_\alpha = \partial_\alpha X^M \Gamma_M$, and

$$\Gamma_{\text{D5}}^{(0)} = \frac{\varepsilon^{\alpha_1 \ldots \alpha_6}}{6! \sqrt{-\det g}} \hat{\Gamma}_{\alpha_1 \ldots \alpha_6}. \tag{4.67}$$

We will split the field and the metric in a four-dimensional part (along the D3 worldvolume) and a transverse part along the two-sphere as:

$$\mathcal{F} = \mathcal{F}^{\|} + \mathcal{F}^{\perp}, \qquad ds^2 = ds_{\|}^2 + ds_{\perp}^2. \tag{4.68}$$

It is not hard to see that the matrix in the projector splits as:

$$\beta_+ = \beta_+^{\perp} \beta_+^{\|}, \tag{4.69}$$

with

$$\begin{aligned} \beta_+^{\perp} =& \frac{\varepsilon^{\alpha\beta}}{2! \sqrt{G_\perp}} \Gamma_{\alpha\beta} \frac{\sqrt{-\det G_\perp}}{\sqrt{-\det(G_\perp + \mathcal{F}^\perp)}} \left(1 + \frac{1}{2} \mathcal{F}^\perp_{\alpha\beta} \Gamma^{\alpha\beta}\right), \\ \beta_+^{\|} =& \frac{\varepsilon^{\mu_1 \ldots \mu_4}}{4! \sqrt{G_\|}} \hat{\Gamma}_{\mu_1 \ldots \mu_4} \frac{\sqrt{-\det G_\|}}{\sqrt{-\det(G_\| + \mathcal{F}^\|)}} \\ & \times \left(1 + \frac{1}{2} \mathcal{F}^\|_{\mu_1\mu_2} \hat{\Gamma}^{\mu_1\mu_2} + \frac{1}{8} \mathcal{F}^\|_{\mu_1\mu_2} \mathcal{F}^\|_{\mu_3\mu_4} \hat{\Gamma}^{\mu_1\mu_2\mu_3\mu_4}\right) \end{aligned} \tag{4.70}$$

where Greek letters still refer to worldvolume indices, but we make a split: the middle of the alphabet to four dimensions ($\mu, \nu \ldots = 0, 1, 2, 3$) and the beginning to the two-sphere ($\alpha, \beta \ldots = 4, 5$).

The calculation of the term $\beta_+^\perp$ follows straightforwardly from the discussion of section 4.2, with

$$\mathcal{F}^\perp = -Q(\psi)\mathrm{vol}_{S^2}\,. \tag{4.71}$$

The four-dimensional part of the projector parallels that of the projector dubbed β in the appendix of [146]. Note that we only consider the bosonic terms, as fermionic terms in β would take us beyond the quadratic fermionic order in the action. The result for $\beta_+^{\|}$ is

$$\begin{aligned}\beta_+^{\|} = \Gamma_{\underline{0123}}\Big(1 &+ \frac{1}{2}F_{\mu\nu}\Gamma^{\mu\nu} + \partial^\mu X^I\Gamma_{I\mu} - \frac{1}{2}g_{IJ}\partial_\mu X^I\partial^\mu X^J - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &+ \frac{1}{4}F_{\mu\nu}(\star_4 F)^{\mu\nu}\Gamma_{\underline{0123}} - \frac{1}{2}\partial^\mu X^I\partial^\nu X^J\Gamma_{IJ\mu\nu} + \frac{1}{2}\partial_\rho X^I F_{\mu\nu}\Gamma^{I\rho\mu\nu} + \ldots\Big)\,.\end{aligned} \tag{4.72}$$

The ellipses indicate terms higher order in fields, indices have been raised and lowered with the metric $G_\|$ and X^I are the transverse coordinates. This is the straightforward covariantization of the κ -symmetry matrix for a D3-brane.

Applied to one non-trivial transverse scalar $X^1 = \psi$, we have

$$\begin{aligned}\beta_+^\perp &= \cos(\alpha) - \Gamma_{\underline{45}}\sin(\alpha) \\ \beta_+^{\|} &= \Gamma_{\underline{0123}}\Big(1 + \frac{1}{2}F_{\mu\nu}\Gamma^{\mu\nu} + \partial^\mu\psi\Gamma_{\psi\mu} - \frac{1}{2}g_{\psi\psi}\partial_\mu\psi\partial^\mu\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &\qquad + \frac{1}{4}F_{\mu\nu}(\star_4 F)^{\mu\nu}\Gamma_{\underline{0123}} + \frac{1}{2}\partial_\rho\psi F_{\mu\nu}\Gamma^{\psi\rho\mu\nu} + \ldots\Big)\,.\end{aligned} \tag{4.73}$$

4.A.2 Fermionic action

We briefly describe how to get the mass terms of the action (4.19). After gauge fixing $\theta_1 = 0$, and writing $\lambda = \theta_2$, the terms not involving derivatives define a mass matrix $\mathcal{M}$ as

$$\bar\lambda\mathcal{M}\lambda \equiv \bar\lambda(1-\Gamma_{\mathrm{D5}})[(\tilde M^{-1})^{\alpha\beta}\Gamma_\alpha W_\beta - \Delta]\lambda. \tag{4.74}$$

We split the terms not involving a covariant derivative along the four dimensions and the two-sphere as

$$(\tilde M^{-1})^{\alpha\beta}\Gamma_\alpha W_\beta - \Delta = M_\| + M_\perp\,, \tag{4.75}$$

with

$$\begin{aligned}M_\| &= [G_\|^{\mu\nu}\Gamma_\nu W_\mu - \Delta]\,, \\ M_\perp &= [(G_\perp + 2\pi g_s\sigma_3\mathcal{F})^{-1}]^{\alpha\beta}\Gamma_\alpha W_\beta\,.\end{aligned} \tag{4.76}$$

We find (using α, β for directions on the two sphere, and μ, ν for four dimensions)

$$
\begin{aligned}
M_{\parallel} &= \frac{1}{8}\left(H_{\mu pq}\sigma_3 + e^{\Phi}F_{\mu pq}\sigma_1\right)\Gamma^{\mu pq} - \frac{1}{24}\left(H_{mnp}\sigma_3 + e^{\Phi}F_{mnp}\sigma_1\right)\Gamma^{mnp} \\
M_{\perp} &= \sin^2(\alpha)\left(\frac{1}{8}(\sigma_1 F^{\alpha np} + \sigma_3 H^{\alpha np})\Gamma_{\alpha np} - \frac{1}{8}\sigma_1 F^{mnp}\Gamma_{mnp}\right) \\
&\quad + \cos^2(\alpha)((\sigma_3\mathcal{F})^{-1})^{\alpha\beta}\Bigg(\frac{1}{8}H_{\alpha pq}\sigma_3\Gamma_{\beta}{}^{pq} - \frac{1}{8\cdot 3!}e^{\Phi}F^{npq}\sigma_1\Gamma_{\alpha\beta npq} \\
&\quad + \frac{1}{8}\left(2H_{\alpha\beta q}\sigma_3 - e^{\Phi}F_{\alpha\beta q}\sigma_1\right)\Gamma^q\Bigg)
\end{aligned}
\tag{4.77}
$$

with

$$
\cos(\alpha) = -\frac{\sqrt{\det\mathcal{F}}}{\sqrt{-\det(G_\perp + \mathcal{F})}}\,, \qquad \sin(\alpha) = -\frac{\sqrt{\det G_\perp}}{\sqrt{-\det(G_\perp + \mathcal{F})}}\,. \tag{4.78}
$$

The signs in these last two equations are chosen for later convenience.

Now we use that the flux H_3 is fully along S^3 and $\mathcal{F}$ is along S^2, while F_3 is orthogonal. So the non-zero terms in $M_{\parallel}, M_{\perp}$ are

$$
\begin{aligned}
M_{\parallel} &= -\frac{1}{24}\left(H_{mnp}\sigma_3 + e^{\Phi}F_{mnp}\sigma_1\right)\Gamma^{mnp}\,, \\
M_{\perp} &= \sin^2(\alpha)\left(\frac{1}{12}(\sigma_3 H^{mnp})\Gamma_{mnp} - \frac{1}{4\cdot 3!}\sigma_1 e^{\Phi}F^{mnp}\Gamma_{mnp}\right)\,, \\
&\quad + \cos^2(\alpha)((\sigma_3\mathcal{F})^{-1})^{\alpha\beta}\left(-\frac{1}{8\cdot 3!}e^{\Phi}F^{npq}\sigma_1\Gamma_{\alpha\beta npq} + \frac{1}{4}H_{\alpha\beta q}\sigma_3\Gamma^q\right)\,,
\end{aligned}
\tag{4.79}
$$

which gives the result (4.21).

From (4.65) and (4.73) we find that for vanishing F and neglecting the derivative terms on ψ (as they are higher order in the action), we get

$$
\Gamma_{\mathrm{D5}} = -\begin{pmatrix} 0 & \beta_- \\ \beta_+ & 0 \end{pmatrix}\,, \qquad \beta_\pm = \Gamma_{\underline{0123}}(\pm\cos\alpha - \sin\alpha\Gamma_{\underline{45}})\,. \tag{4.80}
$$

Now we use that for Majorana-Weyl bilinears only terms with three or seven gamma matrices are non-zero.

$$
\bar{\lambda}\Gamma^{m_1\dots m_n}\lambda = 0 \qquad \text{for } n \notin \{3, 7\} \tag{4.81}
$$

We now see that the last term in $M_\perp$ will not contribute at all and we find

$$
\begin{aligned}
&\bar{\lambda}(1 - \Gamma_{\mathrm{D5}})\mathcal{M}\lambda = \\
&\frac{1}{24}\bar{\lambda}\left[\cos(2\alpha)H_{mnp} + \left(1 + \sin^2(\alpha) - \frac{1}{2}\cos^2(\alpha)(\mathcal{F}^{-1})^{\alpha\beta}\Gamma_{\alpha\beta}\right)e^{\Phi}\beta_+ F_{mnp}\right]\Gamma^{mnp}\lambda\,.
\end{aligned}
\tag{4.82}
$$

With the identity $(\mathcal{F}^{-1})^{\alpha\beta}\Gamma_{\alpha\beta} = 2\tan(\alpha)\Gamma_{\underline{45}}$ and dropping again terms with the wrong number of Γ matrices, we find

$$\mathcal{M} = \frac{1}{24}\bar{\lambda}\left(\cos(2\alpha)H_{mnp} + \cos(\alpha)e^{\Phi}F_{mnp}\Gamma_{\underline{0123}}\right)\Gamma^{mnp}\lambda\,. \tag{4.83}$$

Finally we can use the Majorana-Weyl property $\Gamma_{(10)}\lambda = \lambda$, to write:

$$F_{mnp}\Gamma_{\underline{0123}}\lambda = (\star_6 F)_{mnp}\lambda\,, \tag{4.84}$$

with $\star_6$ the Hodge star operator on the six-dimensional internal manifold. This yields the final result (4.27):

$$\mathcal{M} = \frac{1}{24}\left(\cos(2\alpha)H_{mnp} + e^{\Phi}\cos(\alpha)F_{mnp}\Gamma_{\underline{0123}}\right)\Gamma^{mnp}\,. \tag{4.85}$$

Fermionic action: orientifold compatible gauge choice

For completeness, we show that taking the alternative gauge choice

$$(1+\Gamma_{\text{D5}})\theta = 0 \quad \Leftrightarrow \quad \theta_1 = -\Gamma_{\underline{0123}}(\cos(\alpha) + \sin(\alpha)\Gamma_{\underline{45}})\theta_2\,, \tag{4.86}$$

to fix the κ-symmetry we obtain the same mass matrix. This gauge choice is useful when one also wants to perform an orientifold projection, which has to be compatible with the gauge fixing condition. Using this condition, we can write the terms appearing in $\mathcal{M}_{\|}$ and $\mathcal{M}_{\perp}$ completely in terms of $\lambda \equiv \theta_2$.

$$\begin{aligned} \bar{\theta}H_{mnp}\Gamma^{mnp}\theta &= 0\,, \\ \bar{\theta}H_{mnp}\Gamma^{mnp}\sigma_3\theta &= -2\bar{\lambda}H_{mnp}\Gamma^{mnp}\lambda\,, \\ \bar{\theta}F_{mnp}\Gamma^{mnp}\sigma_1\theta &= -2\cos(\alpha)\bar{\lambda}F_{mnp}\Gamma^{mnp}\Gamma_{\underline{0123}}\lambda\,. \end{aligned} \tag{4.87}$$

We then find after some algebra that

$$\begin{aligned} \bar{\lambda}\mathcal{M}_{\|}\lambda &= \frac{1}{12}\bar{\lambda}\left(H_{mnp} + e^{\Phi}\cos(\alpha)F_{mnp}\right)\Gamma^{mnp}\lambda\,, \\ \bar{\lambda}\mathcal{M}_{\perp}\lambda &= -\frac{1}{6}\sin^2(\alpha)\,\bar{\lambda}H_{mnp}\Gamma^{mnp}\lambda\,. \end{aligned} \tag{4.88}$$

Here, we again used (4.81) to eliminate some terms. The total mass matrix is then given by

$$\bar{\lambda}\mathcal{M}\lambda = \frac{1}{12}\bar{\lambda}\left(\cos(2\alpha)H_{mnp} + e^{\Phi}\cos(\alpha)F_{mnp}\Gamma_{\underline{0123}}\right)\Gamma^{mnp}\lambda\,, \tag{4.89}$$

in agreement with the mass matrix in the gauge where $\theta_1 = 0$ up to a factor of 2.

4.A.3 Supersymmetry transformations

The fields on the brane enjoy a combination of supersymmetry transformations, κ-symmetry with spinorial parameter κ and diffeomorphisms (we leave out the possibility of gauge transformations of the gauge field). To linear order in the fermions θ, these are:

$$\begin{aligned}\delta\theta &= \epsilon + [\mathbb{1} + \Gamma_{\rm D5}]\kappa + \xi^\alpha \partial_\alpha \theta\,, \\ \delta X^m &= -\bar{\theta}\Gamma^m \epsilon + \bar{\theta}\Gamma^m[\mathbb{1} + \Gamma_{\rm D5}]\kappa + \xi^\alpha \partial_\alpha X^m\,, \\ \delta A_\alpha &= -\bar{\theta}\hat{\Gamma}_\alpha \sigma_3 \epsilon - C_{\alpha m}\bar{\theta}\Gamma^m \epsilon + \bar{\theta}\hat{\Gamma}_\alpha \sigma_3 [\mathbb{1} + \Gamma_{\rm D5}]\kappa \\ &\quad + C_{\alpha m}\bar{\theta}\Gamma^m[\mathbb{1} + \Gamma_{\rm D5}]\theta + \xi^\beta F_{\beta\alpha}\,.\end{aligned} \tag{4.90}$$

As explained in [153, 157], we can fix the gauge redundancy in the following way. We fix κ-symmetry by the spinor gauge choice $\theta_1 = 0$ or $(\mathbb{1} + \sigma_3)\theta = 0$ and by requiring that this remains valid under the combined transformation

$$(\mathbb{1} + \sigma_3)\delta\theta = 0\,. \tag{4.91}$$

The diffeomorphism invariance can be fixed by requiring static gauge, such that $\delta X^\alpha = 0$. The background spinor obeys

$$\epsilon_2 = \Gamma_{\underline{0123}}\epsilon_1\,. \tag{4.92}$$

This sets

$$\epsilon^1 + \kappa^1 - \beta_- \kappa^2 = 0\,, \qquad \text{and} \qquad \xi^\alpha = \bar{\lambda}\Gamma^\alpha[\mathbb{1} + \beta]\epsilon_2\,. \tag{4.93}$$

We will denote the transverse scalars by X^I and with slight abuse of notation $\epsilon = -2\epsilon_2$. Then the supersymmetry transformations after fixing the κ-gauge that leave the quadratic action (4.26) invariant are (see also [153])

$$\begin{aligned}\delta_\epsilon \lambda &= -\frac{1}{2}[\mathbb{1} - \beta]\epsilon + \mathcal{O}(\lambda^2)\,, \\ \delta_\epsilon X^I &= \frac{1}{2}\bar{\lambda}\Gamma^I[\mathbb{1} + \beta]\epsilon + \xi^\alpha \partial_\alpha X^I + \mathcal{O}(\lambda^3)\,, \\ \delta_\epsilon A_\alpha &= -\frac{1}{2}\bar{\lambda}(\hat{\Gamma}_\alpha + \Gamma_I \partial_\alpha X^I)[\mathbb{1} + \beta]\epsilon + \frac{1}{2}C_{\alpha m}\bar{\lambda}\Gamma^m[\mathbb{1} + \beta]\epsilon + \xi^\beta F_{\beta\alpha} + \mathcal{O}(\lambda^3),\end{aligned} \tag{4.94}$$

with

$$\beta = \Gamma_{\underline{0123}}\beta_+\,, \qquad \xi^\alpha = -\frac{1}{2}\bar{\lambda}\Gamma^\alpha[\mathbb{1} + \beta]\epsilon\,. \tag{4.95}$$

5 Summary & Outlook

In this final chapter, we summarize the main results presented in this thesis and put them into context. We provide an outlook and mention several interesting directions for future work.

5.1 Instabilities of anti-de Sitter space

One important observation that we made in this thesis is that to study quantum gravity it is not always necessary to directly probe Planckian energies. It turns out that UV effects do not always decouple from IR physics, contrary to expectations from effective field theory. Instead, for an arbitrary effective theory to be consistently coupled to quantum gravity it needs to fulfill certain consistency conditions known as the swampland conjectures. One of these conditions is the Weak Gravity Conjecture (WGC), which is closely related to decay of nonsupersymmetric extremal charged black holes.

In Chapter 2, we studied the WGC in the context of four-dimensional Reissner-Nordström black holes. By focussing on the (thermal) emission of the black hole in the s-wave sector, which is the dominant decay channel, we were able to construct an effective action for a spherically symmetric shell that we used to describe the emission process in terms of quanta tunneling through the horizon. Doing so made it possible to take into account backreaction and the resulting decay rate of the black hole deviates slightly from Hawking's original calculation, which did not take into account backreaction.

Most interestingly, this decay rate is governed by the black hole entropy and remains finite even after taking the extremal limit. In this case, the black hole temperature vanishes and the decay process can be interpreted as Schwinger pair production. However, this decay channel is only allowed when there exists a (super-)extremal state whose charge-to-mass ratio exceeds that of the extremal black hole, as required by the WGC. When this condition is satisfied, the near-horizon anti-de Sitter space then also becomes unstable in line with the conjectures of [60, 61].

Strictly speaking, our results only apply to AdS_2, but based on the universal form of the decay rate in terms of the black hole entropy we expect a generalization to higher-dimensional anti-de Sitter spaces to be possible. Furthermore, our results might also be used to derive constraints on low-energy effective descriptions. Typically, black hole/black brane solutions contain a compact space and by reducing over it we can obtain a lower-dimensional effective theory. It would then be of interest to understand if and how the instability of the 'parent black geometry' carries over to restrict the properties of the effective description. In fact, some ideas along these lines have already been applied to a compactification of the standard model (see for example [61, 176]) leading to non-trivial constraints on beyond the standard model physics.

5.2 Instabilities of cosmological spacetimes

In Chapter 3, we extended our analysis of quantum instabilities of black holes and anti-de Sitter space to cosmological spacetimes. We constructed the analogue of the Unruh state for black holes and analyzed if this state can be used as a physically acceptable alternative to the more commonly used Bunch-Davies state for patches of spacetime that are locally described by de Sitter space. We constructed this state by an asymmetric choice of boundary conditions in the quantization of a massless scalar field: incoming modes are taken to be in the Bunch-Davies vacuum, but for outgoing modes we pick the static vacuum. As a result, a static observer will only measure incoming radiation and the outgoing component of the flux is removed. This behaviour is to be contrasted with the Bunch-Davies vacuum, in which a static observer measures both incoming and outgoing radiation.

Starting from two-dimensional de Sitter space and generalizing the computation to four dimensions, we constructed the energy-momentum tensor in the Unruh state and showed it is only singular on the past horizon and well-defined on an entire planar patch. Therefore, the Unruh state could be a physically acceptable state for cosmological purposes. A striking property of the Unruh state is that it contains a negative energy density at the horizon that violates the null energy condition. This can be understood from the fact that the Unruh state can be viewed as being constructed from the Bunch-Davies state by removing the (positive energy) outgoing flux. Because the Unruh state breaks some of the de Sitter symmetries, pure de Sitter space is no longer a solution when including backreaction effects. We estimated the effect of backreaction by approximating the energy in a static patch to be smeared out homogeneously. Because the negative energy density violates the null energy condition, we find that the Hubble parameter slowly increases with a rate that is Planck suppressed. This puts a fundamental bound on the lifetime of de Sitter space in the Unruh state set by its gravitational entropy.

To fully understand what are the cosmological applications of the Unruh state, one would need to characterize the difference between choosing the Bunch-Davies vacuum and the Unruh state in terms of inflationary correlation functions and derive the modifications of, for example, the power spectrum and bispectrum. Because backreaction effects in the Unruh state are Planck suppressed, we anticipate that any difference is similarly suppressed during a phase of $\mathcal{O}(60)$ e-folds of slow-roll inflation, but it would be interesting to study this in more detail in future work.

In the context of string theory a bound on the lifetime of de Sitter space is not that surprising, as de Sitter solutions in string theory are typically only metastable. In this thesis, we mainly focussed on a particular construction of de Sitter vacua in string theory known as the KKLT scenario. One of the open questions after the original KKLT paper [32], was exactly how the anti-D3-branes used in the KKLT scenario break supersymmetry spontaneously. Under the assumption of putting a single anti-D3-brane on top of an orientifold plane, it was shown in [143, 144] that its contribution to the potential can be captured in a manifestly supersymmetric fashion using a single nilpotent superfield containing the goldstino. However, in Chapter 4 we argued that the situation is much more subtle when this assumption is removed and polarization effects are taken into account.

Due to polarization, supersymmetry breaking can no longer be understood as being purely four-dimensional. Instead, higher-dimensional physics needs to be properly taken into account which reveals the existence of a Kaluza-Klein tower of modes that are crucial to be able to correctly identify the goldstino. In addition, these modes might also affect cosmological models based on the KKLT scenario. In particular, the decay rate of the metastable vacuum derived in [148] was estimated by truncating to the lowest lying Kaluza-Klein mode. It could very well be the case that when taking into account the dynamics of the entire Kaluza-Klein tower, this result is modified and it would be interesting to return to some of these questions in future work.

5.3 Outlook

To close this this thesis, let us remind ourselves of the goal we started out with, which was to understand the (universal) properties of quantum gravity and see how they constrain low-energy physics. One of the lessons we have learned is that vacuum solutions, such as anti-de Sitter space and de Sitter space have a tendency to decay due to quantum effects, unless protected by a symmetry. As mentioned before, the instability of anti-de Sitter space dovetails nicely with expectations from the swampland conjectures. On the other hand, it is unclear how the instability of de Sitter space in the Unruh state connects (if at all) to string theory and in particular to the metastability of de Sitter space in the KKLT scenario. Said differently, how should the boundary conditions defining the Unruh state be interpreted in string theory and especially in the KKLT scenario?

In future work, it would be interesting to explore this connection (if it exists) especially in the light of the recent de Sitter swampland conjectures [87, 88], which claim that metastable de Sitter space belongs to the swampland. According to these conjectures, any solution with a positive energy density is necessarily rolling with a rate that is of order one in Planck units. This seems to be in stark contrast with both the instability of de Sitter space in the Unruh state (which is much slower and of the opposite sign) and the metastability of de Sitter space in the KKLT scenario. However, it should be mentioned that the bounds of [87, 88] are currently only supported by circumstantial evidence and it could be that the KKLT scenario evades these bounds by working in a regime of parameter space where these bounds do not apply. Future work should clarify this and at the moment no conclusive statements can be made about the tension (or absence thereof) between the KKLT scenario and the swampland conjectures.

Another interesting direction for future study is to understand more precisely at what energy scale constraints dictated by the swampland conjecture have to appear. After all, the idea of the swampland conjectures is to constrain low-energy effective theories, such as scalar potentials driving inflation. At the moment, it is not clear however how constraining the swampland conjectures truly are, as illustrated by the different versions of the WGC that are on the market. If only the mild form of the WGC should be satisfied, we could imagine a situation where the state that satisfies the WGC only appears above the cutoff of a low-energy observer. In this case, the WGC might not be that constraining after all. There are indications however that a stronger form must be true.

If we consider the worldsheet description of perturbative string theories for example, modular invariance of the partition function can be used to relate constraints on the heavy sector of the theory to constraints on the light sector and vice versa. This exemplifies once more that in quantum gravity (or at least in string theory) the UV does not always decouple from the IR. We can use this to our benefit by assuming a mild form of the WGC derived from black hole physics [20–22] and strengthen it using modular invariance to constrain the low-energy spectrum of the theory [5].

The hope is that, eventually, we will know precisely what are the requirements that any effective theory should satisfy in order for it to be consistently embedded into quantum gravity. If this program succeeds, it could be used to 'bootstrap' the quantum gravity problem: instead of doing detailed string theory computations, we could explore the properties of quantum gravity simply by imposing all necessary consistency conditions.

Bibliography

[1] L. Aalsma and J. P. van der Schaar, "Extremal Tunneling and Anti-de Sitter Instantons," *JHEP* **03** (2018) 145, arXiv:1801.04930 [hep-th].

[2] L. Aalsma, M. Parikh, and J. P. van der Schaar, "Back(reaction) to the Future in the Unruh-de Sitter State," arXiv:1905.02714 [hep-th].

[3] L. Aalsma, J. P. van der Schaar, and B. Vercnocke, "Constrained superfields on metastable anti-D3-branes," *JHEP* **05** (2017) 089, arXiv:1703.05771 [hep-th].

[4] L. Aalsma, M. Tournoy, J. P. van der Schaar, and B. Vercnocke, "Supersymmetric embedding of antibrane polarization," *Phys. Rev.* **D98** (2018) 086019, arXiv:1807.03303 [hep-th].

[5] L. Aalsma, A. Cole, and G. Shiu, "Weak Gravity Conjecture, Black Hole Entropy, and Modular Invariance," arXiv:1905.06956 [hep-th].

[6] W. Taylor and Y.-N. Wang, "The F-theory geometry with most flux vacua," *JHEP* **12** (2015) 164, arXiv:1511.03209 [hep-th].

[7] W. G. Unruh, "Notes on black hole evaporation," *Phys. Rev.* **D14** (1976) 870.

[8] R. Parentani, "The Energy momentum tensor in Fulling-Rindler vacuum," *Class. Quant. Grav.* **10** (1993) 1409–1416, arXiv:hep-th/9303062 [hep-th].

[9] H. Ooguri and C. Vafa, "On the geometry of the string landscape and the swampland," *Nucl. Phys. B* **766** no. 1-3, (2007) 21–33, arXiv:0605264 [hep-th].

[10] T. D. Brennan, F. Carta, and C. Vafa, "The String Landscape, the Swampland, and the Missing Corner," *PoS* **TASI2017** (2017) 015, arXiv:1711.00864 [hep-th].

[11] E. Palti, "The Swampland: Introduction and Review," 2019. arXiv:1903.06239 [hep-th].

[12] 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, arXiv:hep-th/0601001 [hep-th].

[13] B. Heidenreich, M. Reece, and T. Rudelius, "Evidence for a sublattice weak gravity conjecture," *JHEP* **08** (2017) 025, arXiv:1606.08437 [hep-th].

[14] S. Andriolo, D. Junghans, T. Noumi, and G. Shiu, "A Tower Weak Gravity Conjecture from Infrared Consistency," *Fortsch. Phys.* **66** no. 5, (2018) 1800020, arXiv:1802.04287 [hep-th].

[15] T. Banks and L. J. Dixon, "Constraints on String Vacua with Space-Time Supersymmetry," *Nucl. Phys.* **B307** (1988) 93–108.

[16] T. Banks and N. Seiberg, "Symmetries and Strings in Field Theory and Gravity," *Phys. Rev.* **D83** (2011) 084019, arXiv:1011.5120 [hep-th].

[17] D. Harlow and H. Ooguri, "Constraints on symmetry from holography," arXiv:1810.05337 [hep-th].

[18] D. Harlow and H. Ooguri, "Symmetries in quantum field theory and quantum gravity," arXiv:1810.05338 [hep-th].

[19] L. Susskind, "Trouble for remnants," arXiv:hep-th/9501106 [hep-th].

[20] Y. Hamada, T. Noumi, and G. Shiu, "Weak Gravity Conjecture from Unitarity and Causality," arXiv:1810.03637 [hep-th].

[21] B. Bellazzini, M. Lewandowski, and J. Serra, "Amplitudes' Positivity, Weak Gravity Conjecture, and Modified Gravity," arXiv:1902.03250 [hep-th].

[22] C. Cheung, J. Liu, and G. N. Remmen, "Proof of the Weak Gravity Conjecture from Black Hole Entropy," *JHEP* **10** (2018) 004, arXiv:1801.08546 [hep-th].

[23] M. Montero, "A Holographic Derivation of the Weak Gravity Conjecture," arXiv:1812.03978 [hep-th].

[24] D. Baumann and L. McAllister, *Inflation and String Theory*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2015. arXiv:1404.2601 [hep-th].

[25] F. Denef, "Les Houches Lectures on Constructing String Vacua," *Les Houches* **87** (2008) 483–610, arXiv:0803.1194 [hep-th].

[26] M. Dine and N. Seiberg, "Is the Superstring Weakly Coupled?," *Phys. Lett.* **162B** (1985) 299–302.

[27] S. P. Martin, "A Supersymmetry primer," *Adv. Ser. Direct. High Energy Phys.* (1997) 1–98, `arXiv:hep-ph/9709356 [hep-ph]`.

[28] Z. Komargodski and N. Seiberg, "From Linear SUSY to Constrained Superfields," *JHEP* **09** (2009) 066, `arXiv:0907.2441 [hep-th]`.

[29] J. Wess and J. Bagger, *Supersymmetry and supergravity.* Princeton University Press, 1992.

[30] D. V. Volkov and V. P. Akulov, "Is the neutrino a goldstone particle?," *Phys. Lett. B* **46** no. 1, (1973) 109–110.

[31] J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond.* Cambridge University Press, 1998.

[32] S. Kachru, R. Kallosh, A. D. Linde, and S. P. Trivedi, "De Sitter vacua in string theory," *Phys. Rev.* **D68** (2003) 046005, `arXiv:hep-th/0301240 [hep-th]`.

[33] S. B. Giddings, S. Kachru, and J. Polchinski, "Hierarchies from fluxes in string compactifications," *Phys. Rev.* **D66** (2002) 106006, `arXiv:hep-th/0105097 [hep-th]`.

[34] U. H. Danielsson and T. Van Riet, "What if string theory has no de Sitter vacua?," *Int. J. Mod. Phys.* **D27** no. 12, (2018) 1830007, `arXiv:1804.01120 [hep-th]`.

[35] S. Sethi, "Supersymmetry Breaking by Fluxes," *JHEP* **10** (2018) 022, `arXiv:1709.03554 [hep-th]`.

[36] S. Kachru and S. P. Trivedi, "A comment on effective field theories of flux vacua," *Fortsch. Phys.* **67** no. 1-2, (2019) 1800086, `arXiv:1808.08971 [hep-th]`.

[37] G. L. Pimentel and J. Stout, "Real-Time Corrections to the Effective Potential," `arXiv:1905.00219 [hep-th]`.

[38] J. Moritz, A. Retolaza, and A. Westphal, "Toward de Sitter space from ten dimensions," *Phys. Rev.* **D97** no. 4, (2018) 46010, `arXiv:1707.08678 [hep-th]`.

[39] Y. Hamada, A. Hebecker, G. Shiu, and P. Soler, "Brane gaugino condensate in 10d," `arXiv:1812.06097 [hep-th]`.

[40] R. Kallosh, "Gaugino Condensation and Geometry of the Perfect Square," arXiv:1901.02023 [hep-th].

[41] Y. Hamada, A. Hebecker, G. Shiu, and P. Soler, "Understanding KKLT from a 10d perspective," arXiv:1902.01410 [hep-th].

[42] F. Carta, J. Moritz, and A. Westphal, "Gaugino condensation and small uplifts in KKLT," arXiv:1902.01412 [hep-th].

[43] F. F. Gautason, V. Van Hemelryck, T. Van Riet, and G. Venken, "A 10d view on the KKLT AdS vacuum and uplifting," arXiv:1902.01415 [hep-th].

[44] P. McGuirk, G. Shiu, and Y. Sumitomo, "Non-supersymmetric infrared perturbations to the warped deformed conifold," *Nucl. Phys. B* **842** no. 3, (2011) 383–413, arXiv:0910.4581.

[45] I. Bena, M. Graña, and N. Halmagyi, "On the existence of meta-stable vacua in Klebanov-Strassler," *JHEP* **2010** no. 9, (2010) 87, arXiv:0912.3519.

[46] J. Blåbäck, U. H. Danielsson, and T. Van Riet, "Resolving anti-brane singularities through time-dependence," *JHEP* **2013** no. 2, (2013) 1–13, arXiv:1202.1132.

[47] I. Bena, A. Buchel, and Ó. J. C. Dias, "Horizons cannot save the landscape," *Phys. Rev. D* **87** no. 6, (2013) 63012, arXiv:1212.5162.

[48] F. F. Gautason, D. Junghans, and M. Zagermann, "Cosmological constant, near brane behavior and singularities," *JHEP* **2013** no. 9, (2013) 123, arXiv:1301.5647.

[49] I. Bena, M. Graña, S. Kuperstein, and S. Massai, "Tachyonic anti-M2 branes," *JHEP* **2014** no. 6, (2014) 173, arXiv:1402.2294.

[50] J. Blåbäck, U. H. Danielsson, D. Junghans, T. Van Riet, and S. C. Vargas, "Localised anti-branes in non-compact throats at zero and finite T," *JHEP* **2015** no. 2, (2015) 18, arXiv:1409.0534.

[51] I. Bena, M. Graña, S. Kuperstein, and S. Massai, "Giant Tachyons in the Landscape," *JHEP* **02** (2015) 146, arXiv:1410.7776 [hep-th].

[52] U. H. Danielsson and T. Van Riet, "Fatal attraction: more on decaying anti-branes," *JHEP* **2015** no. 3, (2015) 1–24, arXiv:1410.8476.

[53] B. Michel, E. Mintun, J. Polchinski, A. Puhm, and P. Saad, "Remarks on brane and antibrane dynamics," *JHEP* **09** (2015) 021, arXiv:1412.5702

[hep-th].

[54] I. Bena and S. Kuperstein, "Brane polarization is no cure for tachyons," *JHEP* **2015** no. 9, (2015) 112, arXiv:1504.00656.

[55] D. Cohen-Maldonado, J. Diaz, T. van Riet, and B. Vercnocke, "Observations on fluxes near anti-branes," *JHEP* **01** (2016) 126, arXiv:1507.01022 [hep-th].

[56] I. Bena, J. Blåbäck, and D. Turton, "Loop corrections to the antibrane potential," *JHEP* **07** (2016) 132, arXiv:1602.05959 [hep-th].

[57] D. Cohen-Maldonado, J. Diaz, and F. F. Gautason, "Polarised antibranes from Smarr relations," *JHEP* **2016** no. 5, (2016) 175, arXiv:1603.05678.

[58] U. H. Danielsson, F. F. Gautason, and T. Van Riet, "Unstoppable brane-flux decay of $\overline{D}6$-branes," *JHEP* **2017** no. 3, (2017) 141, arXiv:1609.06529 [hep-th].

[59] J. Armas, N. Nguyen, V. Niarchos, N. A. Obers, and T. Van Riet, "Meta-stable non-extremal anti-branes," arXiv:1812.01067 [hep-th].

[60] B. Freivogel and M. Kleban, "Vacua Morghulis," arXiv:1610.04564 [hep-th].

[61] H. Ooguri and C. Vafa, "Non-supersymmetric AdS and the Swampland," *Adv. Theor. Math. Phys.* **21** (2017) 1787–1801, arXiv:1610.01533 [hep-th].

[62] P. Kraus and F. Wilczek, "Selfinteraction correction to black hole radiance," *Nucl. Phys.* **B433** (1995) 403–420, arXiv:gr-qc/9408003 [gr-qc].

[63] P. Kraus and F. Wilczek, "Effect of selfinteraction on charged black hole radiance," *Nucl. Phys.* **B437** (1995) 231–242, arXiv:hep-th/9411219 [hep-th].

[64] M. K. Parikh and F. Wilczek, "Hawking radiation as tunneling," *Phys. Rev. Lett.* **85** (2000) 5042–5045, arXiv:hep-th/9907001 [hep-th].

[65] S. Massar and R. Parentani, "Gravitational instanton for black hole radiation," *Phys. Rev. Lett.* **78** (1997) 3810–3813, arXiv:gr-qc/9701015 [gr-qc].

[66] J. D. Bekenstein, "Statistical black-hole thermodynamics," *Phys. Rev. D* **12** (1975) 3077–3085.

[67] E. Keski-Vakkuri and P. Kraus, "Microcanonical D-branes and back reaction," *Nucl. Phys.* **B491** (1997) 249–262, `arXiv:hep-th/9610045 [hep-th]`.

[68] G. W. Gibbons, "Vacuum Polarization and the Spontaneous Loss of Charge by Black Holes," *Commun. Math. Phys.* **44** (1975) 245–264.

[69] J. M. Maldacena, J. Michelson, and A. Strominger, "Anti-de Sitter fragmentation," *JHEP* **02** (1999) 011, `arXiv:hep-th/9812073 [hep-th]`.

[70] D. Brill, "Splitting of an extremal Reissner-Nordström throat via quantum tunneling," *Phys. Rev.* **D46** (1992) 1560–1565, `arXiv:hep-th/9202037 [hep-th]`.

[71] P. Painlevé, "La mécanique classique et la théorie de la relativité," *Comptes Rendus Academie des Sciences* **173** (1921) 677–680.

[72] S. W. Hawking and G. T. Horowitz, "The Gravitational Hamiltonian, action, entropy and surface terms," *Class. Quant. Grav.* **13** (1996) 1487–1498, `arXiv:gr-qc/9501014 [gr-qc]`.

[73] S. Massar and R. Parentani, "On the gravitational back reaction to Hawking radiation," `arXiv:gr-qc/9801043 [gr-qc]`.

[74] A. Hansen and F. Ravndal, "Klein's Paradox and Its Resolution," *Physica Scripta* **23** no. 6, (1981) 1036.

[75] R. Brito, V. Cardoso, and P. Pani, "Superradiance," *Lect. Notes Phys.* **906** (2015) pp.1–237, `arXiv:1501.06570 [gr-qc]`.

[76] J. D. Bekenstein and A. Meisels, "Einstein A and B coefficients for a black hole," *Phys. Rev. D* **15** (May, 1977) 2775–2781.

[77] S. M. Carroll, M. C. Johnson, and L. Randall, "Extremal limits and black hole entropy," *JHEP* **11** (2009) 109, `arXiv:0901.0931 [hep-th]`.

[78] W. Israel, "Singular hypersurfaces and thin shells in general relativity," *Nuovo Cim.* **B44S10** (1966) 1.

[79] M. Cvetic and H. H. Soleng, "Supergravity domain walls," *Phys. Rept.* **282** (1997) 159–223, `arXiv:hep-th/9604090 [hep-th]`.

[80] H. Chung, "Tunneling between single and multi-centered black hole configurations," *Phys. Rev.* **D86** (2012) 064036, `arXiv:1201.3028 [gr-qc]`.

[81] S. Ng and M. Perry, "Brane splitting via quantum tunneling," *Nucl. Phys.*

B634 (2002) 209–229, `arXiv:hep-th/0106008 [hep-th]`.

[82] B. Pioline and J. Troost, "Schwinger pair production in AdS_2," *JHEP* **03** (2005) 043, `arXiv:hep-th/0501169 [hep-th]`.

[83] U. H. Danielsson, G. Dibitetto, and S. C. Vargas, "A swamp of non-SUSY vacua," *JHEP* **11** (2017) 152, `arXiv:1708.03293 [hep-th]`.

[84] S. Aretakis, "Stability and Instability of Extreme Reissner-Nordström Black Hole Spacetimes for Linear Scalar Perturbations I," *Commun. Math. Phys.* **307** (2011) 17–63, `arXiv:1110.2007 [gr-qc]`.

[85] S. Aretakis, "Stability and Instability of Extreme Reissner-Nordström Black Hole Spacetimes for Linear Scalar Perturbations II," *Annales Henri Poincare* **12** (2011) 1491–1538, `arXiv:1110.2009 [gr-qc]`.

[86] P. Zimmerman, "Horizon instability of extremal Reissner-Nordström black holes to charged perturbations," *Phys. Rev.* **D95** no. 12, (2017) 124032, `arXiv:1612.03172 [gr-qc]`.

[87] G. Obied, H. Ooguri, L. Spodyneiko, and C. Vafa, "De Sitter Space and the Swampland," `arXiv:1806.08362 [hep-th]`.

[88] H. Ooguri, E. Palti, G. Shiu, and C. Vafa, "Distance and de Sitter Conjectures on the Swampland," *Phys. Lett.* **B788** (2019) 180–184, `arXiv:1810.05506 [hep-th]`.

[89] J. B. Hartle and S. W. Hawking, "Path Integral Derivation of Black Hole Radiance," *Phys. Rev.* **D13** (1976) 2188–2203.

[90] M. Spradlin, A. Strominger, and A. Volovich, "Les Houches lectures on de Sitter space," pp. 423–453. 2001. `arXiv:hep-th/0110007 [hep-th]`.

[91] E. Witten, "Quantum gravity in de Sitter space," 2001. `arXiv:hep-th/0106109 [hep-th]`.

[92] D. Marolf and I. A. Morrison, "The IR stability of de Sitter QFT: results at all orders," *Phys. Rev.* **D84** (2011) 044040, `arXiv:1010.5327 [gr-qc]`.

[93] A. Rajaraman, "de Sitter Space is Unstable in Quantum Gravity," *Phys. Rev.* **D94** no. 12, (2016) 125025, `arXiv:1608.07237 [hep-th]`.

[94] A. Dabholkar, "Quantum Weyl Invariance and Cosmology," *Phys. Lett.* **B760** (2016) 31–35, `arXiv:1511.05342 [hep-th]`.

[95] P. R. Anderson and E. Mottola, "Quantum vacuum instability of 'eternal' de Sitter space," *Phys. Rev.* **D89** (2014) 104039, `arXiv:1310.1963`

[gr-qc].

[96] T. Markkanen, "De Sitter Stability and Coarse Graining," *Eur. Phys. J.* **C78** no. 2, (2018) 97, arXiv:1703.06898 [gr-qc].

[97] G. Dvali and C. Gomez, "Quantum Exclusion of Positive Cosmological Constant?," *Annalen Phys.* **528** (2016) 68–73, arXiv:1412.8077 [hep-th].

[98] B. Greene, M. Parikh, and J. P. van der Schaar, "Universal correction to the inflationary vacuum," *JHEP* **04** (2006) 057, arXiv:hep-th/0512243 [hep-th].

[99] P. Candelas, "Vacuum polarization in Schwarzschild spacetime," *Phys. Rev. D* **21** (1980) 2185–2202.

[100] J. Martin, "Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask)," *Comptes Rendus Physique* **13** (2012) 566–665, arXiv:1205.3365 [astro-ph.CO].

[101] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge Univ. Press, 1984.

[102] S. M. Christensen and S. A. Fulling, "Trace Anomalies and the Hawking Effect," *Phys. Rev.* **D15** (1977) 2088–2104.

[103] P. Candelas and J. S. Dowker, "Field Theories On Conformally Related Space-times: Some Global Considerations," *Phys. Rev.* **D19** (1979) 2902.

[104] G. W. Gibbons and S. W. Hawking, "Cosmological event horizons, thermodynamics, and particle creation," *Phys. Rev. D* **15** (1977) 2738–2751.

[105] R. M. Wald and U. Yurtsever, "General proof of the averaged null energy condition for a massless scalar field in two-dimensional curved space-time," *Phys. Rev.* **D44** (1991) 403–416.

[106] N. Arkani-Hamed, S. Dubovsky, A. Nicolis, E. Trincherini, and G. Villadoro, "A Measure of de Sitter entropy and eternal inflation," *JHEP* **05** (2007) 055, arXiv:0704.1814 [hep-th].

[107] S. B. Giddings and M. S. Sloth, "Semiclassical relations and IR effects in de Sitter and slow-roll space-times," *JCAP* **1101** (2011) 023, arXiv:1005.1056 [hep-th].

[108] P. C. W. Davies, "Cosmological Horizons and the Generalized Second Law

of Thermodynamics," *Class. Quant. Grav.* **4** (1987) L225.

[109] R. Bousso, "Positive vacuum energy and the N bound," *JHEP* **11** (2000) 038, arXiv:hep-th/0010252 [hep-th].

[110] G. Dvali, C. Gomez, and S. Zell, "Quantum Breaking Bound on de Sitter and Swampland," arXiv:1810.11002 [hep-th].

[111] U. Danielsson, "The quantum swampland," arXiv:1809.04512 [hep-th].

[112] M. K. Parikh, "New coordinates for de Sitter space and de Sitter radiation," *Phys. Lett.* **B546** (2002) 189–195, arXiv:hep-th/0204107 [hep-th].

[113] I. Antoniadis, E. Dudas, S. Ferrara, and A. Sagnotti, "The Minimal Volkov - Akulov - Starobinsky Supergravity," *Phys. Lett.* **B733** (2014) 32–35, arXiv:1403.3269 [hep-th].

[114] E. Dudas, S. Ferrara, A. Kehagias, and A. Sagnotti, "Properties of Nilpotent Supergravity," *JHEP* **09** (2015) 217, arXiv:1507.07842 [hep-th].

[115] E. A. Bergshoeff, D. Z. Freedman, R. Kallosh, and A. Van Proeyen, "Pure de Sitter Supergravity," *Phys. Rev.* **D92** no. 8, (2015) 085040, arXiv:1507.08264 [hep-th].

[116] F. Hasegawa and Y. Yamada, "Component action of nilpotent multiplet coupled to matter in 4 dimensional $\mathcal{N} = 1$ supergravity," *JHEP* **10** (2015) 106, arXiv:1507.08619 [hep-th].

[117] N. Cribiori, G. Dall'Agata, F. Farakos, and M. Porrati, "Minimal Constrained Supergravity," *Phys. Lett.* **B764** (2017) 228–232, arXiv:1611.01490 [hep-th].

[118] G. Dall'Agata, E. Dudas, and F. Farakos, "On the origin of constrained superfields," *JHEP* **05** (2016) 041, arXiv:1603.03416 [hep-th].

[119] S. Ferrara, R. Kallosh, A. Van Proeyen, and T. Wrase, "Linear Versus Non-linear Supersymmetry, in General," *JHEP* **04** (2016) 065, arXiv:1603.02653 [hep-th].

[120] A. Sagnotti and S. Ferrara, "Supersymmetry and Inflation," *PoS* **PLANCK2015** (2015) 113, arXiv:1509.01500 [hep-th].

[121] S. Ferrara, A. Kehagias, and A. Sagnotti, "Cosmology and Supergravity," *Int. J. Mod. Phys.* **A31** no. 25, (2016) 1630044, arXiv:1605.04791 [hep-th].

[122] J. Moritz and T. Van Riet, "Racing through the swampland: de Sitter uplift vs weak gravity," *JHEP* **09** (2018) 099, `arXiv:1805.00944 [hep-th]`.

[123] P. Agrawal, G. Obied, P. J. Steinhardt, and C. Vafa, "On the Cosmological Implications of the String Swampland," *Phys. Lett.* **B784** (2018) 271–276, `arXiv:1806.09718 [hep-th]`.

[124] G. Dvali and C. Gomez, "On Exclusion of Positive Cosmological Constant," *Fortsch. Phys.* **67** no. 1-2, (2019) 1800092, `arXiv:1806.10877 [hep-th]`.

[125] D. Andriot, "On the de Sitter swampland criterion," *Phys. Lett.* **B785** (2018) 570–573, `arXiv:1806.10999 [hep-th]`.

[126] F. Denef, A. Hebecker, and T. Wrase, "de Sitter swampland conjecture and the Higgs potential," *Phys. Rev.* **D98** no. 8, (2018) 086004, `arXiv:1807.06581 [hep-th]`.

[127] C. Roupec and T. Wrase, "de Sitter extrema and the swampland," *Fortsch. Phys.* **2018** (2018) 1800082, `arXiv:1807.09538 [hep-th]`.

[128] D. Andriot, "New constraints on classical de Sitter: flirting with the swampland," *Fortsch. Phys.* **67** no. 1-2, (2019) 1800103, `arXiv:1807.09698 [hep-th]`.

[129] J. K. Ghosh, E. Kiritsis, F. Nitti, and L. T. Witkowski, "De Sitter and Anti-de Sitter branes in self-tuning models," *JHEP* **11** (2018) 128, `arXiv:1807.09794 [hep-th]`.

[130] J. P. Conlon, "The de Sitter swampland conjecture and supersymmetric AdS vacua," *Int. J. Mod. Phys.* **A33** no. 29, (2018) 1850178, `arXiv:1808.05040 [hep-th]`.

[131] K. Dasgupta, M. Emelin, E. McDonough, and R. Tatar, "Quantum Corrections and the de Sitter Swampland Conjecture," *JHEP* **01** (2019) 145, `arXiv:1808.07498 [hep-th]`.

[132] M. Cicoli, S. De Alwis, A. Maharana, F. Muia, and F. Quevedo, "De Sitter vs Quintessence in String Theory," *Fortsch. Phys.* **2018** (2018) 1800079, `arXiv:1808.08967 [hep-th]`.

[133] Y. Akrami, R. Kallosh, A. Linde, and V. Vardanyan, "The landscape, the swampland and the era of precision cosmology," *Fortsch. Phys.* **2018** (2018) 1800075, `arXiv:1808.09440 [hep-th]`.

[134] R. Kallosh, A. Linde, E. McDonough, and M. Scalisi, "de Sitter Vacua with

a Nilpotent Superfield," *Fortsch. Phys.* **2018** (2018) 1800068, `arXiv:1808.09428 [hep-th]`.

[135] J. Moritz, A. Retolaza, and A. Westphal, "On uplifts by warped anti-D3-branes," *Fortsch. Phys.* **67** no. 1-2, (2019) 1800098, `arXiv:1809.06618 [hep-th]`.

[136] R. Kallosh, A. Linde, E. McDonough, and M. Scalisi, "4d models of dS uplift in KKLT," *Phys. Rev.* **D99** no. 4, (2019) 046006, `arXiv:1809.09018 [hep-th]`.

[137] F. F. Gautason, V. Van Hemelryck, and T. Van Riet, "The tension between 10D supergravity and dS uplifts," *Fortsch. Phys.* **67** no. 1-2, (2019) 1800091, `arXiv:1810.08518 [hep-th]`.

[138] R. Kallosh, A. Linde, E. McDonough, and M. Scalisi, "dS Vacua and the Swampland," `arXiv:1901.02022 [hep-th]`.

[139] R. Kallosh, A. Karlsson, and D. Murli, "From linear to nonlinear supersymmetry via functional integration," *Phys. Rev.* **D93** no. 2, (2016) 025012, `arXiv:1511.07547 [hep-th]`.

[140] R. Kallosh, A. Karlsson, B. Mosk, and D. Murli, "Orthogonal Nilpotent Superfields from Linear Models," *JHEP* **05** (2016) 082, `arXiv:1603.02661 [hep-th]`.

[141] D. M. Ghilencea, "Comments on the nilpotent constraint of the goldstino superfield," *Mod. Phys. Lett.* **A31** no. 12, (2016) 1630011, `arXiv:1512.07484 [hep-th]`.

[142] E. Dudas, L. Heurtier, C. Wieck, and M. W. Winkler, "UV Corrections in Sgoldstino-less Inflation," *Phys. Lett.* **B759** (2016) 121–125, `arXiv:1601.03397 [hep-th]`.

[143] R. Kallosh and T. Wrase, "Emergence of Spontaneously Broken Supersymmetry on an Anti-D3-Brane in KKLT dS Vacua," *JHEP* **12** (2014) 117, `arXiv:1411.1121 [hep-th]`.

[144] E. A. Bergshoeff, K. Dasgupta, R. Kallosh, A. Van Proeyen, and T. Wrase, "$\overline{\mathrm{D3}}$ and dS," *JHEP* **05** (2015) 058, `arXiv:1502.07627 [hep-th]`.

[145] B. Vercnocke and T. Wrase, "Constrained superfields from an anti-D3-brane in KKLT," *JHEP* **08** (2016) 132, `arXiv:1605.03961 [hep-th]`.

[146] R. Kallosh, B. Vercnocke, and T. Wrase, "String Theory Origin of

Constrained Multiplets," *JHEP* **09** (2016) 063, arXiv:1606.09245 [hep-th].

[147] I. R. Klebanov and M. J. Strassler, "Supergravity and a confining gauge theory: Duality cascades and χSB resolution of naked singularities," *JHEP* **08** (2000) 052, arXiv:hep-th/0007191 [hep-th].

[148] S. Kachru, J. Pearson, and H. L. Verlinde, "Brane / flux annihilation and the string dual of a nonsupersymmetric field theory," *JHEP* **06** (2002) 021, arXiv:hep-th/0112197 [hep-th].

[149] R. C. Myers, "Dielectric branes," *JHEP* **12** (1999) 022, arXiv:hep-th/9910053 [hep-th].

[150] R. P. Andrews and N. Dorey, "Deconstruction of the Maldacena-Núñez compactification," *Nucl. Phys. B* **751** no. 1-2, (2006) 304–341, arXiv:hep-th/0601098 [hep-th].

[151] D. P. Jatkar, G. Mandal, S. R. Wadia, and K. P. Yogendran, "Matrix dynamics of fuzzy spheres," *JHEP* **01** (2002) 039, arXiv:hep-th/0110172 [hep-th].

[152] C. Bachas, M. R. Douglas, and C. Schweigert, "Flux stabilization of D-branes," *JHEP* **05** (2000) 048, arXiv:hep-th/0003037 [hep-th].

[153] L. Martucci, J. Rosseel, D. Van den Bleeken, and A. Van Proeyen, "Dirac actions for D-branes on backgrounds with fluxes," *Class. Quant. Grav.* **22** (2005) 2745–2764, arXiv:hep-th/0504041 [hep-th].

[154] K. Dasgupta, M. Emelin, and E. McDonough, "Fermions on the antibrane: Higher order interactions and spontaneously broken supersymmetry," *Phys. Rev.* **D95** no. 2, (2017) 026003, arXiv:1601.03409 [hep-th].

[155] M. Graña, "D3-brane action in a supergravity background: The Fermionic story," *Phys. Rev.* **D66** (2002) 045014, arXiv:hep-th/0202118 [hep-th].

[156] M. Cvetič, G. W. Gibbons, H. Lü, and C. N. Pope, "Ricci-flat metrics, harmonic forms and brane resolutions," *Communications in Mathematical Physics* **232** no. 3, (2003) 457–500, arXiv:0012011 [hep-th].

[157] E. Bergshoeff, F. Coomans, R. Kallosh, C. S. Shahbazi, and A. Van Proeyen, "Dirac-Born-Infeld-Volkov-Akulov and Deformation of Supersymmetry," *JHEP* **08** (2013) 100, arXiv:1303.5662 [hep-th].

[158] P. McGuirk, G. Shiu, and F. Ye, "Soft branes in supersymmetry-breaking backgrounds," *JHEP* **07** (2012) 188, arXiv:1206.0754 [hep-th].

[159] O. Dewolfe, S. Kachru, and M. Mulligan, "Gravity dual of metastable dynamical supersymmetry breaking," *Phys. Rev. D* **77** no. 6, (2008) 65011, `arXiv:0801.1520`.

[160] M. Bertolini, D. Musso, I. Papadimitriou, and H. Raj, "A goldstino at the bottom of the cascade," *JHEP* **11** (2015) 184, `arXiv:1509.03594 [hep-th]`.

[161] C. Krishnan, H. Raj, and P. N. Bala Subramanian, "On the KKLT Goldstino," *JHEP* **06** (2018) 092, `arXiv:1803.04905 [hep-th]`.

[162] G. S. Hartnett, "Localised anti-branes in flux backgrounds," *JHEP* **2015** no. 6, (2015) 7, `arXiv:1501.06568`.

[163] J. Polchinski and M. J. Strassler, "The String dual of a confining four-dimensional gauge theory," `arXiv:hep-th/0003136 [hep-th]`.

[164] N. J. Evans, C. V. Johnson, and M. Petrini, "Clearing the throat: Irrelevant operators and finite temperature in large N gauge theory," *JHEP* **05** (2002) 2, `arXiv:hep-th/0112058 [hep-th]`.

[165] P. Palffy-Muhoray and X. Zheng, "Eigenvalue decomposition for tensors of arbitrary rank,".

[166] J. M. Maldacena and C. Nuñez, "Towards the large N limit of pure N=1 superYang-Mills," *Phys. Rev. Lett.* **86** (2001) 588–591, `arXiv:hep-th/0008001 [hep-th]`.

[167] F. F. Gautason, M. Schillo, and T. Van Riet, "Is inflation from unwinding fluxes IIB?," *JHEP* **03** (2017) 037, `arXiv:1611.07037 [hep-th]`.

[168] J. Diaz Dorronsoro and M. Schillo, "Towards an explicit model of large field inflation," *JHEP* **05** (2018) 075, `arXiv:1712.04466 [hep-th]`.

[169] A. Ashoorioon, H. Firouzjahi, and M. M. Sheikh-Jabbari, "M-flation: Inflation from matrix valued scalar fields," *JCAP* **2009** no. 6, (2009) , `arXiv:0903.1481`.

[170] A. J. Bray and M. A. Moore, "Influence of dissipation on quantum coherence," *Physical Review Letters* **49** no. 21, (1982) 1545–1549.

[171] A. O. Caldeira and A. J. Leggett, "Quantum tunnelling in a dissipative system," *Annals of Physics* **149** no. 2, (1983) 374–456.

[172] S. De Alwis, R. Gupta, E. Hatefi, and F. Quevedo, "Stability, tunneling and flux changing de Sitter transitions in the large volume string scenario," *JHEP* **2013** no. 11, (2013) 179, `arXiv:1308.1222`.

[173] D. Marolf, L. Martucci, and P. J. Silva, "Fermions, T duality and effective actions for D-branes in bosonic backgrounds," *JHEP* **04** (2003) 051, `arXiv:hep-th/0303209 [hep-th]`.

[174] D. Marolf, L. Martucci, and P. J. Silva, "Actions and Fermionic symmetries for D-branes in bosonic backgrounds," *JHEP* **07** (2003) 019, `arXiv:hep-th/0306066 [hep-th]`.

[175] R. Kallosh and A. Linde, "Inflation and Uplifting with Nilpotent Superfields," *JCAP* **1501** (2015) 025, `arXiv:1408.5950 [hep-th]`.

[176] L. E. Ibanez, V. Martin-Lozano, and I. Valenzuela, "Constraining Neutrino Masses, the Cosmological Constant and BSM Physics from the Weak Gravity Conjecture," *JHEP* **11** (2017) 066, `arXiv:1706.05392 [hep-th]`.

Samenvatting

Quantumzwaartekracht

Een groot deel van de fysica in de 20$^{\text{e}}$ eeuw stond in het teken van de ontwikkeling van twee theorieën: quantumveldentheorie en de algemene relativiteitstheorie. Deze twee theorieën vormen de basis van de moderne theoretische fysica en beschrijven de vier fundamentele krachten van de natuur. Quantumveldentheorie geeft de wiskundige beschrijving van drie van deze krachten: de elektromagnetische kracht, de zwakke kernkracht en de sterke kernkracht. De vierde kracht, zwaartekracht, past niet in het raamwerk van quantumveldentheorie en wordt in plaats daarvan beschreven door Einsteins algemene relativiteitstheorie.

Deze twee theorieën zijn enorm succesvol gebleken in het beschrijven van een rijk scala aan fenomenen. Recent nog zijn twee belangrijke voorspellingen van beide theorieën bevestigd. In 2012 toonden onderzoekers verbonden aan CERN (het onderzoeksinstituut in Genève dat de Large Hadron Collider herbergt) het bestaan van het Higgsboson aan. Deze meting bevestigde een voorspelling van Brout, Englert en Higgs uit de jaren 60 en was een grote triomf voor quantumveldentheorie. Later, in 2015, werden voor het eerst direct zwaartekrachtsgolven waargenomen door wetenschappers verbonden aan de LIGO/VIRGO experimenten. Het bestaan van deze 'golven in de ruimtetijd' werd al door Einstein afgeleid in 1916. Beide ontdekkingen hebben nogmaals de voorspellingen van quantumveldentheorie en de algemene relativiteitstheorie bevestigd en zijn dan ook bekroond met een Nobelprijs.

Een van de vragen die hiermee niet is beantwoord, is hoe deze twee theorieën samengebracht kunnen worden in een overkoepelende beschrijving. Deze vraag is niet alleen interessant vanuit theoretisch oogpunt, maar ook van groot belang voor de beschrijving van systemen wiens zwaartekracht zeer groot is. In dergelijke situaties, die bijvoorbeeld optreden in zwarte gaten en vlak na het ontstaan van het heelal, volstaat de beschrijving van zwaartekracht vanuit de algemene relativiteitstheorie niet meer. Om een fysisch correcte beschrijving van deze systemen te geven moet zwaartekracht quantummechanisch worden behandeld en is een zogeheten theorie van quantumzwaartekracht nodig.

Een van de best begrepen en meest succesvolle voorstellen voor zo'n theorie is snaartheorie. Volgens snaartheorie bestaan de elementaire puntdeeltjes (zoals bijvoorbeeld elektronen) uit kleine tweedimensionale snaren. Ook bevat de theorie hogerdimensionale membranen, wiens excitaties net als die van snaren corresponderen met verschillende type deeltjes. In principe zijn de effecten van deze snaren en membranen direct te meten, maar in de praktijk zijn ze zo klein dat dit met de huidige technologie nagenoeg onmogelijk is. Het blijkt echter dat in sommige situaties microscopische quantumzwaartekrachteffecten toch macroscopische gevolgen kunnen hebben. Zo zijn er bijvoorbeeld aanwijzingen dat de conventionele beschrijving van zwarte gaten vanuit de algemene relativiteitstheorie al op de schaal van de horizon moet worden aangepast om consistent te zijn met quantumzwaartekracht. Centraal in dit proefschrift staat daarom de vraag wat de macroscopische voorspellingen van quantumzwaartekracht zijn. In hoofdstuk 2 en 3 focussen we hierbij voornamelijk op algemene quantumeffecten van zwaartekracht en in hoofdstuk 4 duiken we wat dieper in de snaartheorie.

Resultaten van dit proefschrift

In hoofdstuk 2 bestuderen we quantumeffecten in zwaartekracht in de context van extremale zwarte gaten. Deze zwarte gaten hebben naast een massa ook een even grote elektrische lading. Het interessante aan extremale zwarte gaten is dat ze dichtbij de horizon worden beschreven door een tweedimensionaal anti-de Sitter vacuum: een oplossing van Einsteins algemene relativiteitstheorie met een negatieve energiedichtheid. Alhoewel deze geometrie niet direct ons eigen universum beschrijft (dat een positieve energiedichtheid heeft) is zij wel uitermate geschikt om quantumzwaartekracht te bestuderen in een eenvoudige achtergrond. In hoofdstuk 2 laten we zien dat quantumeffecten ervoor zorgen dat in deze achtergrond spontaan deeltjes worden gecreëerd die tot een instabiliteit van het vacuum leiden. Deze instabiliteit hangt niet af van de precieze microscopische beschrijving van anti-de Sitter vacua en lijkt dus een universele voorspelling van quantumzwaartekracht te zijn.

In hoofdstuk 3 verplaatsen we onze aandacht naar quantumeffecten in de Sitter vacua. Dit zijn oplossingen van de algemene relativiteitstheorie die een positieve energiedichtheid hebben en daarmee een goede beschrijving geven van ons uitdijende heelal. We analyseren de verschillende quantumtoestanden die gekozen kunnen worden in een de Sitter vacuum. De keuze voor een bepaalde quantumtoestand kan worden gezien als het opleggen van specifieke randvoorwaarden die de evolutie van de geometrie bepalen. We identificeren en construeren een nieuwe quantumtoestand die we de Unruh-de Sitter toestand noemen en bepalen zijn karakteristieke eigenschappen. Net als de meer conventioneel gekozen Bunch-Davies toestand lijkt deze toestand een goede beschrijving te geven van het vroege heelal. Een belangrijk verschil met de Bunch-Davies toestand is dat de Unruh-de Sitter toestand minder symmetrieën bewaart en dat de Sitter vacua hierdoor een instabiliteit ontwikkelen. Dit plaatst een fundamentele restrictie op de levensduur van een uitdijend heelal. We beargumenteren dat de Unruh-de Sitter toestand wellicht een meer accurate beschrijving geeft van het vroege universum dan de Bunch-Davies toestand. Verder onderzoek moet uitwijzen of deze toestand ook mogelijk waarneembare eigenschappen heeft waarmee deze kan worden onderscheiden van de Bunch-Davies toestand.

We vervolgen onze discussie in hoofdstuk 4 door de Sitter vacua te bestuderen in snaartheorie. We focussen specifiek op een snaartheoretische constructie van de Sitter vacua die bekend staat als het KKLT scenario. Een van de karakteristieke eigenschappen van deze constructie is dat supersymmetrie (een symmetrie van snaartheorie tussen bosonen en fermionen) wordt gebroken door gebruik te maken van een stapel anti-D3-branen: een van de bouwstenen van snaartheorie. Het breken van supersymmetrie is noodzakelijk aangezien de Sitter vacua door hun positieve energiedichtheid deze symmetrie niet kunnen bewaren. Om wel consistent te zijn met de onderliggende supersymmetrie van snaartheorie moet supersymmetrie gecontroleerd (de technische benaming is ‘spontaan’) worden gebroken in het KKLT scenario. Tot recent was het niet duidelijk of anti-D3-branen supersymmetrie inderdaad spontaan breken, maar in hoofdstuk 4 wordt, voortbouwend op eerder werk, overtuigend aangetoond dat dit wel het geval is. Een universele voorspelling van spontaan gebroken supersymmetrie is de aanwezigheid van een massaloos fermion. Door een matrixmodel op te stellen dat de vrijheidsgraden van de stapel anti-D3-branen beschrijft is dit massaloze fermion expliciet geïdentificeerd. Dit resultaat ondersteunt dus de consistentie van het KKLT scenario wat betreft de breking van supersymmetrie.

Toekomstig onderzoek

De resultaten van dit proefschrift zijn slechts een kleine stap in de goede richting om de eigenschappen van quantumzwaartekracht te ontrafelen. In het bijzonder staat ons begrip van quantumzwaartekracht in de Sitter vacua nog in de kinderschoenen. Momenteel is er een levendige discussie gaande in de literatuur over de consistentie van de verschillende constructies van de Sitter vacua in snaartheorie, zoals het KKLT scenario, en of snaartheorie überhaupt wel stabiele de Sitter vacua toelaat. Deze discussie is ook zeer relevant in de context van de resultaten gepresenteerd in hoofdstuk 3 waar we een quantumtoestand hebben geïdentificeerd waarin de Sitter vacua een instabiliteit ontwikkelen. Het zou interessant zijn om te begrijpen hoe deze toestand zich manifesteert in snaartheorie en of dit een mogelijke verklaring is voor de spanning die er lijkt te bestaan tussen kosmologische (uitdijende) geometrieën en snaartheorie. Meer algemeen suggereren de resultaten van dit proefschrift dat microscopische quantumzwaartekracht effecten macroscopische gevolgen kunnen hebben. Ondanks het feit dat deze effecten typisch zeer klein zijn biedt dit de hoop dat we in sommige situaties toch universele voorspellingen kunnen afleiden die mogelijk te testen zijn met nauwkeurige observaties van zwarte gaten en het uitdijende heelal.

Dankwoord

Anders dan wat je wellicht zou denken is theoretische fysica een echte teamsport. Dit proefschrift zou dan ook niet mogelijk zijn geweest zonder de vele mensen die de afgelopen jaren bewust dan wel onbewust hun steentje hebben bijgedragen.

Allereerst wil ik graag mijn promotor bedanken. Jan Pieter, jouw onverminderde enthousiasme is een bron van inspiratie en heeft ervoor gezorgd dat ik altijd met veel plezier discussies met je heb gevoerd. Ik bewonder je pragmatische aanpak van de fysica en hoop ook in de toekomst nog met je van gedachten te kunnen wisselen over (al dan niet instabiele) de Sitter vacua. Secondly, I would like to thank my copromotor Ben for all his sharp insights and optimism.

Over the years, I have had the opportunity to work on diverse topics with various great people. Alex, Bert, Gary, Jan Pieter, Magnus and Maulik, thank you for always being critical and challenging me.

Furthermore, my PhD would definitely not have been as enjoyable without the amazing people that form the Amsterdam string theory and cosmology groups. To all the staff members, postdocs and PhD students: thank you! I very much enjoyed all the interesting conversations, coffee breaks and cookie times we have had over the years. Ook zou het instituut natuurlijk nooit zo goed draaien zonder de inzet van Anne-Marieke, Astrid, Jirina, Joost en Klaartje, bedankt! Daarnaast wil ik in het bijzonder mijn (voormalige) kantoorgenoten Gerben en Manus bedanken voor de inspirerende discussies, het beantwoorden van mijn vragen en natuurlijk de uitstekende sfeer. Tevens is mijn dank groot aan Vincent en Jorrit, met wie ik ook de bachelor (Vincent) en master (Jorrit en Vincent) heb mogen doorbrengen. Zoals gezegd is fysica een teamsport en ik heb mij geen betere teamleden kunnen voorstellen. Moreover, I am very happy that Greg and Jorrit agreed to be my paranymphs. I am looking forward towards you standing next to me during the defence.

Of course, not only Amsterdam harbours great physicists. I am grateful to Alex and Gary for making me feel so welcome during the three months that I spent in Madison and I am looking forward to the future projects that we will work on.

Naast alle fysici, ben ik minstens zo veel verschuldigd aan mijn familie. Mijn dank is groot aan mijn ouders, die mij de vrijheid hebben gegeven om mijn eigen pad te bewandelen en aan mijn zussen, die altijd geïnteresseerd zijn gebleven in hetgeen waar ik mijn tijd aan spendeer. Ook bedank ik mijn schoonouders Ad en Esther voor het mij zo warm opnemen in de familie. Ik beloof dat ik goed voor jullie dochter zal zorgen in Amerika. Ten slotte bedank ik mijn allerliefste Jolijn. Dankjewel lieverd voor het meeleven en het relativeren van mijn zorgen de afgelopen jaren. Dankzij jouw steun en liefde ben ik mijn PhD redelijk ongeschonden doorgekomen en ik kijk dan ook met ontzettend veel plezier en vertrouwen uit naar onze toekomst.