

RECEIVED: May 10, 2018

REVISED: August 28, 2018

ACCEPTED: September 11, 2018

PUBLISHED: October 1, 2018

# Proof of the weak gravity conjecture from black hole entropy

Clifford Cheung,<sup>a</sup> Junyu Liu<sup>a</sup> and Grant N. Remmen<sup>b,c</sup>

<sup>a</sup>Walter Burke Institute for Theoretical Physics, California Institute of Technology,  
Pasadena, CA 91125, U.S.A.

<sup>b</sup>Center for Theoretical Physics and Department of Physics, University of California,  
Berkeley, CA 94720, U.S.A.

<sup>c</sup>Lawrence Berkeley National Laboratory,  
Berkeley, CA 94720, U.S.A.

E-mail: [clifford.cheung@caltech.edu](mailto:clifford.cheung@caltech.edu), [jliu2@caltech.edu](mailto:jliu2@caltech.edu),  
[grant.remmen@berkeley.edu](mailto:grant.remmen@berkeley.edu)

**ABSTRACT:** We prove that higher-dimension operators contribute positively to the entropy of a thermodynamically stable black hole at fixed mass and charge. Our results apply whenever the dominant corrections originate at tree level from quantum field theoretic dynamics. More generally, positivity of the entropy shift is equivalent to a certain inequality relating the free energies of black holes. These entropy inequalities mandate new positivity bounds on the coefficients of higher-dimension operators. One of these conditions implies that the charge-to-mass ratio of an extremal black hole asymptotes to unity from above for increasing mass. Consequently, large extremal black holes are unstable to decay to smaller extremal black holes and the weak gravity conjecture is automatically satisfied. Our findings generalize to arbitrary spacetime dimension and to the case of multiple gauge fields. The assumptions of this proof are valid across a range of scenarios, including string theory constructions with a dilaton stabilized below the string scale.

**KEYWORDS:** Black Holes, Effective Field Theories

ARXIV EPRINT: [1801.08546](https://arxiv.org/abs/1801.08546)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Proof of <math>\Delta S &gt; 0</math></b>	<b>4</b>
2.1	Assumptions	5
2.2	Positivity argument	5
2.3	Explicit example	8
2.4	Unitarity and monotonicity	9
<b>3</b>	<b>Classical vs. quantum</b>	<b>10</b>
3.1	Leading contributions	10
3.2	Region of interest	11
<b>4</b>	<b>Black hole spacetime</b>	<b>12</b>
4.1	Unperturbed solution	12
4.2	Perturbed solution	13
<b>5</b>	<b>Calculation of entropy</b>	<b>14</b>
5.1	Wald entropy formula	14
5.2	Interaction contribution	15
5.3	Horizon contribution	15
<b>6</b>	<b>New positivity bounds</b>	<b>17</b>
6.1	General bounds	17
6.2	Examples and consistency checks	17
6.3	The weak gravity conjecture	19
6.4	Entropy, area, and extremality	21
<b>7</b>	<b>Discussion and conclusions</b>	<b>22</b>
<b>A</b>	<b>Generalization to arbitrary dimension</b>	<b>22</b>
A.1	Black hole spacetime	22
A.2	Calculation of entropy	24
A.3	New positivity bounds	26
<b>B</b>	<b>Field redefinition invariance</b>	<b>27</b>

---

## 1 Introduction

In this paper we argue that black hole thermodynamics implies new constraints on the coefficients of higher-dimension operators. Our results are based on a certain universal property of entropy. In particular, consider a system  $\mathcal{T}$  prepared in a macrostate whose microstate degeneracy is quantified by entropy  $S$ . Now let us define the system  $\tilde{\mathcal{T}}$  to be a restriction of  $\mathcal{T}$  in which a handful of degrees of freedom have been frozen to a fixed configuration. If  $\tilde{\mathcal{T}}$  is prepared in the same macrostate as  $\mathcal{T}$ , then the corresponding entropy  $\tilde{S}$  will be less than  $S$  because the microstate degeneracy is diminished. Thus, we find that

$$S = \tilde{S} + \Delta S, \quad (1.1)$$

where the entropy shift  $\Delta S$  is strictly positive.

A similar logic applies to black hole entropy. Consider the low-energy effective field theory of photons and gravitons in general spacetime dimension  $D$ , defined at a scale far below the mass gap. The effective Lagrangian is

$$\mathcal{L} = \tilde{\mathcal{L}} + \Delta\mathcal{L}, \quad (1.2)$$

where the first term describes pure Einstein-Maxwell theory,<sup>1</sup>

$$\tilde{\mathcal{L}} = \frac{1}{2\kappa^2} R - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \quad (1.3)$$

and second term encodes corrections from higher-dimension operators,

$$\begin{aligned} \Delta\mathcal{L} = & c_1 R^2 + c_2 R_{\mu\nu} R^{\mu\nu} + c_3 R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} \\ & + c_4 R F_{\mu\nu} F^{\mu\nu} + c_5 R_{\mu\nu} F^{\mu\rho} F^\nu{}_\rho + c_6 R_{\mu\nu\rho\sigma} F^{\mu\nu} F^{\rho\sigma} \\ & + c_7 F_{\mu\nu} F^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma} + c_8 F_{\mu\nu} F^{\nu\rho} F_{\rho\sigma} F^{\sigma\mu}. \end{aligned} \quad (1.4)$$

Without loss of generality, we have dropped all terms involving  $\nabla_\rho F_{\mu\nu}$ , which are equivalent via the Bianchi identities to terms already accounted for or terms involving  $\nabla_\mu F^{\mu\nu}$  [1], which vanish in the absence of charged matter sources, which we assume throughout.

For our analysis, we study a large black hole of fixed mass  $m$  and charge  $q$  as measured in natural units at spatial infinity. In the presence of higher-dimension operators, the metric is

$$g_{\mu\nu} = \tilde{g}_{\mu\nu} + \Delta g_{\mu\nu}. \quad (1.5)$$

Unless otherwise stated, all variables with tildes like  $\tilde{g}_{\mu\nu}$  will denote quantities corresponding to a Reissner-Nordström black hole of the same mass and charge in pure Einstein-Maxwell theory, while variables with deltas like  $\Delta g_{\mu\nu}$  will denote the leading corrections from higher-dimension operators, which are linear in the coefficients  $c_i$ .

It is straightforward to compute the black hole entropy using the Wald formula [2],<sup>2</sup>

$$S = -2\pi \int_\Sigma \frac{\delta\mathcal{L}}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\nu} \epsilon_{\rho\sigma}, \quad (1.6)$$

<sup>1</sup>Throughout, we work in units where  $\kappa^2 = 8\pi G$ , metric signature  $(-, +, \dots, +)$ , and sign conventions  $R_{\mu\nu} = R^\rho{}_{\mu\rho\nu}$  and  $R^\mu{}_{\nu\rho\sigma} = \partial_\rho \Gamma^\mu_{\nu\sigma} - \partial_\sigma \Gamma^\mu_{\nu\rho} + \Gamma^\mu_{\rho\alpha} \Gamma^\alpha_{\nu\sigma} - \Gamma^\mu_{\sigma\alpha} \Gamma^\alpha_{\nu\rho}$ .

<sup>2</sup>A general formula for entanglement entropy has also been proposed [3], but this reduces to the Wald formula for the static Killing horizons relevant to our analysis.

where the integration region  $\Sigma$  is the horizon and  $\epsilon_{\mu\nu}$  is the binormal to the horizon, normalized so that  $\epsilon_{\mu\nu}\epsilon^{\mu\nu} = -2$ . A large portion of our technical analysis will be the explicit evaluation of eq. (1.6) at fixed  $q$  and  $m$ . The entropy shift  $\Delta S$  is then defined according to eq. (1.1), where

$$\tilde{S} = \frac{\tilde{A}}{4G} = \frac{2\pi\tilde{A}}{\kappa^2} \quad (1.7)$$

is the Bekenstein-Hawking entropy [4, 5]. Deviations from eq. (1.7) arise from higher-dimension operator corrections to the Lagrangian,  $\mathcal{L} = \tilde{\mathcal{L}} + \Delta\mathcal{L}$ , and the area of horizon,  $A = \tilde{A} + \Delta A$ .

Using standard thermodynamic identities we show that corrections to the Bekenstein-Hawking entropy of a black hole at fixed mass and charge satisfy

$$\Delta S > 0 \quad (1.8)$$

whenever the free energy of the black hole is less than that of a Reissner-Nordström black hole at the same temperature,

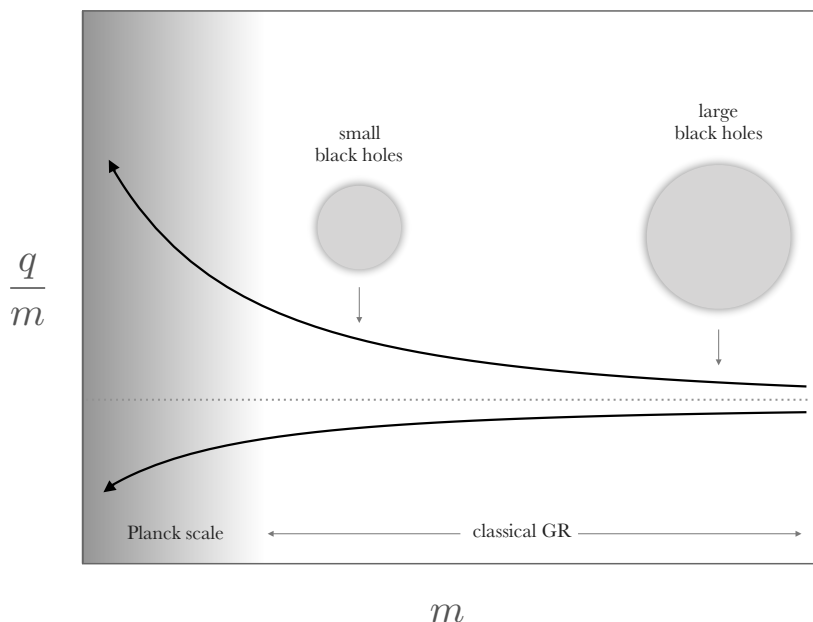
$$F(\beta) < \tilde{F}(\beta). \quad (1.9)$$

Using Euclidean path integral methods we then prove eq. (1.9) for *i*) a thermodynamically stable black hole in *ii*) a theory in which the dominant contributions to higher-dimension operators are generated at tree level by massive quantum fields. The underlying logic of our proof mirrors the parable of  $\mathcal{T}$  and  $\tilde{\mathcal{T}}$  discussed previously. In particular,  $\tilde{\mathcal{L}}$  is obtained directly from the ultraviolet completion of  $\mathcal{L}$  by placing a restriction on massive field fluctuations in the Euclidean path integral. The difference between  $S$  and  $\tilde{S}$  then quantifies the entropy contributions from these modes.

Condition *i*) holds for positive specific heat [6], excluding from consideration the Schwarzschild black hole but permitting Reissner-Nordström black holes over a range of charge-to-mass ratios,  $q/m > \sqrt{2D-5}/(D-2)$ . Condition *ii*) arises naturally in a number of physically-motivated contexts such as string theory, which typically predicts the existence of gravitationally-coupled scalars like the dilaton. If supersymmetry breaking occurs below the string scale, then these states are well described by quantum field theory and can be integrated out at tree level to generate the higher-dimension operators in eq. (1.4). Note that condition *ii*) is perfectly consistent with additional corrections entering at loop level or from intrinsically stringy dynamics, provided these contributions are parametrically smaller than the tree-level component.

While we have only proven the free energy condition in eq. (1.9) under certain assumptions, its connection to positivity of the entropy shift in eq. (1.8) is robust and completely general. Furthermore, a trivial corollary to eq. (1.8) is a positivity condition on the differential entropy generated at each mass threshold encountered while flowing to the infrared.

Remarkably, classical entropy corrections dominate over quantum contributions over a broad range of black hole masses still within the regime of validity of the effective field theory. Within this window, our proof of eq. (1.8) applies. Since the shift in entropy depends on the coefficients of higher-dimension operators, we derive a new class of positivity bounds on the effective field theory. This produces a one-parameter family of constraints



**Figure 1.** Black holes of maximal charge shown as a function of mass  $m$  and charge-to-mass ratio  $q/m$ . Higher-dimension operators induce corrections to the extremality condition. If these corrections are positive, then the WGC is automatically satisfied (upper solid curve) since large black holes are unstable to decay to smaller ones. If these corrections are negative (lower solid curve), then the WGC mandates additional light, superextremal particles to avoid an infinite number of stable extremal black hole remnants.

on the corresponding coefficients  $c_i$  labeled by the charge-to-mass ratio of the black hole from which the bound was derived. Although these conditions are derived from a very particular black hole construction, the resulting positivity bounds constrain the coefficients  $c_i$  in general and are independent of the background.

For the case of highly charged black holes, we obtain a positivity condition on a very specific combination of higher-dimension operator coefficients. Remarkably, it is this exact combination of parameters that also enters into the higher-dimension operator correction to the extremality condition for black holes. In particular, we find that the charge-to-mass ratio for an exactly extremal black hole satisfies

$$\frac{q}{m} - 1 \propto \Delta S, \quad (1.10)$$

where the right-hand side is positive by eq. (1.8). Since higher-dimension operators decouple at long distances, the charge-to-mass ratio asymptotes to unity *from above* as we consider larger and larger extremal black holes. Thus, from charge and energy conservation, it is always possible for an extremal black hole to decay to smaller extremal black holes of marginally higher charge-to-mass ratio, as shown in figure 1. Notably, the existence of such states is precisely mandated by the weak gravity conjecture (WGC), which asserts that an Abelian gauge theory consistently coupled to gravity must contain a state whose

charge exceeds its mass in Planck units [7], so

$$\frac{q}{m} > 1. \quad (1.11)$$

The motivation for the WGC is to forbid the existence of an infinite number of stable states not protected by symmetry. The main result of this paper is that this bound is automatically satisfied by small black holes. Though mysterious at first glance, the connection between entropy and extremality in eq. (1.10) actually follows immediately from the near-horizon properties of the metric for a charged black hole. As we will show, this connection enables us to straightforwardly extend all of our arguments to the multi-charge generalization of the WGC [8] in arbitrary dimension  $D$ .

The WGC is satisfied in numerous concrete examples and is strongly motivated by folk theorems forbidding exact global symmetries that arise in the vanishing-charge limit. Moreover, the WGC is a celebrated avatar of the so-called swampland program [7, 9–11], whose ultimate aim is to systematically delineate regions in effective field theory space consistent with quantum gravitational ultraviolet completion.

Strictly speaking, a bona fide swampland condition distinguishes between low-energy effective field theories that from an infrared standpoint are otherwise entirely consistent. From this perspective it is unclear whether a theory that fails the WGC is in the swampland or *merely pathological* in a sense that can be diagnosed purely from low-energy considerations. For this reason, a related effort has sought to exclude regions of the low-energy parameter space using infrared consistency, e.g., constraints from causality, unitarity, and analyticity of scattering amplitudes [11–15]. While the WGC has previously eluded a formal general proof, the present work demonstrates that it is mathematically equivalent to a certain well-motivated — and in many circumstances provable — property of black hole entropy.

The remainder of this paper is organized as follows. In section 2 we prove, given certain assumptions, that corrections to the Bekenstein-Hawking entropy are positive. Afterwards, in section 3 we discuss how various contributions to the black hole entropy arise and estimate their relative size. In order to constrain the coefficients of higher-dimension operators, we restrict to black holes within a certain mass range. Next, in section 4 we present the perturbative solution for a charged black hole in the presence of higher-dimension operators. We then compute the black hole entropy in section 5 and translate the positivity condition on entropy into a new class of bounds on higher-dimension operator coefficients in section 6. After discussing the implications of these results for the WGC, we conclude in section 7.

## 2 Proof of $\Delta S > 0$

In this section, we study corrections to the Bekenstein-Hawking entropy of a thermodynamically stable black hole at fixed mass and charge. We prove that these contributions are positive whenever they come from higher-dimension operators generated at tree level by quantum fields.

## 2.1 Assumptions

Let us state our assumptions explicitly. First, we assume the existence of quantum fields  $\phi$  at a characteristic mass scale  $m_\phi \ll \Lambda$ , where  $\Lambda$  is the energy scale at which quantum field theory breaks down. This parametric separation is required so that quantum field theory has some regime of validity. By the usual rules of effective field theory, the higher-dimension operators in eq. (1.4) receive small contributions suppressed by the cutoff  $\Lambda$ . In general,  $\Lambda$  can be parametrically smaller than the Planck scale.

Second, we assume that the  $\phi$  degrees of freedom couple to photons and gravitons such that integrating them out generates the higher-dimension operators in eq. (1.4) classically, i.e., at tree level. Since  $m_\phi \ll \Lambda$ , these states induce the dominant contributions to the higher-dimension operators in eq. (1.4). Specifically, the corresponding operator coefficients scale as  $c_i \propto 1/m_\phi^2 \gg 1/\Lambda^2$  since tree-level  $\phi$  exchange is always accompanied by a single factor of  $1/m_\phi^2$  coming from the propagator denominator. Thus, effects arising from the cutoff  $\Lambda$  will be negligible in any context in which quantum field theory is applicable.

As noted previously, states like  $\phi$  are a common feature in string theory, whose low energy spectrum includes particles like the dilaton and moduli, which are massless in the supersymmetric limit. In the presence of supersymmetry breaking, these flat directions are lifted, thus inducing masses  $m_\phi \ll \Lambda$ , where  $\Lambda$  is the string scale.

While our arguments are perfectly consistent with a scale  $\Lambda$  far below the Planck scale, we will frequently refer to pure Einstein-Maxwell theory and the pure Reissner-Nordström solution as a baseline of comparison. We do so entirely out of convenience and not because any component of our argument requires that quantum field theory be applicable up to the Planck scale. Hence, in an abuse of nomenclature, we hereafter refer to the higher-dimension operator contributions of order  $1/m_\phi^2$  as corrections to pure Einstein-Maxwell theory, bearing in mind that we actually mean pure Einstein-Maxwell theory *plus* contributions of order  $1/\Lambda^2$ , which are parametrically smaller than all the contributions of interest.

Third, we focus on black holes that are thermodynamically stable, i.e., have positive specific heat. As we will see, this is necessary for technical reasons so that we can exploit certain properties of the Euclidean path integral.

## 2.2 Positivity argument

Consider a positively charged black hole of mass  $M$  and charge  $Q$  perturbed by higher-dimension operator corrections in general spacetime dimension  $D$ . As we will show in detail in section 4, the perturbed metric  $g_{\mu\nu} = \tilde{g}_{\mu\nu} + \Delta g_{\mu\nu}$  can be computed from the perturbed Lagrangian  $\mathcal{L} = \tilde{\mathcal{L}} + \Delta\mathcal{L}$ , where unless otherwise stated all quantities are expressed as perturbations on a Reissner-Nordström black hole of the same mass  $M$  and charge  $Q$  in pure Einstein-Maxwell theory.

From the perturbed entropy we can define the corresponding inverse temperature  $\beta = \partial_M S$ , which we write as

$$\beta = \tilde{\beta} + \Delta\beta, \quad (2.1)$$

where  $\tilde{\beta} = \partial_M \tilde{S}$  is the inverse temperature of the Reissner-Nordström black hole and  $\Delta\beta = \partial_M \Delta S$  is the shift of the inverse temperature. The latter appears because higher-dimension operators will in general change the temperature of a black hole at a fixed mass and charge.

Next, we compute the free energy  $F(\beta)$  of the perturbed black hole in a canonical ensemble at inverse temperature  $\beta$ . The free energy is calculated from the Euclidean path integral,

$$e^{-\beta F(\beta)} = Z(\beta) = \int d[\hat{g}]d[\hat{A}] e^{-I[\hat{g}, \hat{A}]}, \quad (2.2)$$

where  $\hat{g}$  and  $\hat{A}$  are integration variables running over metric and gauge field configurations. The Euclidean action  $I = \tilde{I} + \Delta I$  is the spacetime integral of the Wick-rotated Lagrangian plus boundary terms appropriate to our choice of the canonical ensemble [16]. Here, the boundary conditions at asymptotic infinity are defined by periodicity  $\beta$  in Euclidean time, with a total electric flux  $Q$  at the boundary [17], though we suppress all dependence on the latter throughout.

By assumption, the higher-dimension operators in the low-energy effective theory are dominated by tree-level contributions from heavy fields. The Euclidean path integral including these ultraviolet modes is

$$\int d[\hat{g}]d[\hat{A}]d[\hat{\phi}] e^{-I_{UV}[\hat{g}, \hat{A}, \hat{\phi}]} = \int d[\hat{g}]d[\hat{A}] e^{-I[\hat{g}, \hat{A}]}, \quad (2.3)$$

where  $\hat{\phi}$  is a collective integration variable running over all configurations of the heavy fields. As a convention, we choose  $\hat{\phi} = 0$  as the boundary condition at asymptotic infinity, thus defining zero as the vacuum expectation value of the field in flat space. In the classical limit, the right-hand side of eq. (2.3) is obtained by solving the equations of motion for the heavy fields and plugging these solutions back into the action.

Now consider an *alternative* field configuration that instead sets all the heavy fields to zero, thus rendering them non-dynamical. This field configuration does *not* satisfy the equations of motion, but this will have no bearing on the following argument. For this configuration the massive fields are decoupled and their contributions to higher-dimension operators are set strictly to zero. It is then a simple mathematical fact that

$$I_{UV}[\hat{g}, \hat{A}, 0] = \tilde{I}[\hat{g}, \hat{A}], \quad (2.4)$$

where the right-hand side is the Euclidean action for pure Einstein-Maxwell theory for any choice of metric and gauge field. The statement of eq. (2.4) encodes our assumption that the dominant contributions to the higher-dimension operators in the effective field theory come from heavy fields.

Putting this all together, we obtain a simple inequality relating the free energy of the perturbed black hole and a Reissner-Nordström black hole at the same temperature,<sup>3</sup>

$$-\log Z(\beta) = I_{UV}[g_\beta, A_\beta, \phi_\beta] < I_{UV}[\tilde{g}_\beta, \tilde{A}_\beta, 0] = \tilde{I}[\tilde{g}_\beta, \tilde{A}_\beta] = -\log \tilde{Z}(\beta). \quad (2.5)$$

---

<sup>3</sup>Strictly speaking, the free energy of a black hole is obtained only after subtracting the free energy contribution from hot flat space or some other reference spacetime [18]. However, since  $\log Z(\beta)$  and  $\log \tilde{Z}(\beta)$  have the same asymptotic boundary conditions, any such reference dependence will cancel from either side of eq. (2.5).



To obtain the first equality we compute  $\log Z(\beta)$  via the saddle-point approximation. Here  $g_\beta$ ,  $A_\beta$ , and  $\phi_\beta$  are the solutions to the classical equations of motion with subscripts to emphasize their consistency with boundary conditions enforcing inverse temperature  $\beta$ . By definition,  $I_{\text{UV}}[g_\beta, A_\beta, \phi_\beta]$  extremizes the Euclidean action. The subsequent inequality then holds if this extremum is *also* a local minimum, in which case off-shell field configurations slightly displaced from the classical solutions will increase the Euclidean action. For this off-shell field configuration we choose the pure Reissner-Nordström metric  $\tilde{g}_\beta$  subject to the same boundary condition dictating inverse temperature  $\beta$ , while pinning all heavy fields to zero. Since this configuration differs only marginally from the true solution of the equations of motion, the displacement from the local minimum will be tiny as required. From eq. (2.4) we see that the resulting expression is formally equal to the Euclidean path integral for pure Einstein-Maxwell theory evaluated on the Reissner-Nordström background. Using the saddle-point approximation once more, we obtain the final equality with  $-\log \tilde{Z}(\beta)$ , which is  $\beta$  times the free energy  $\tilde{F}(\beta)$  for a Reissner-Nordström black hole at inverse temperature  $\beta$ .

Crucially,  $\log \tilde{Z}(\beta)$  does *not* correspond to the free energy of a Reissner-Nordström black hole of mass  $M$ , which has an inverse temperature  $\tilde{\beta}$ . To relate eq. (2.5) to the latter, we plug eq. (2.1) into the right-hand side of eq. (2.5), yielding

$$\log \tilde{Z}(\beta) = \log \tilde{Z}(\tilde{\beta}) - M \partial_M \Delta S, \quad (2.6)$$

where  $\log \tilde{Z}(\tilde{\beta})$  is the free energy of a Reissner-Nordström black hole of mass  $M$  and inverse temperature  $\tilde{\beta}$ . To obtain eq. (2.6), we inserted the thermodynamic relation,  $M = -\partial_{\tilde{\beta}} \log \tilde{Z}(\tilde{\beta})$ , together with the formula for the inverse temperature shift,  $\Delta\beta = \partial_M \Delta S$ . From the definition of the free energy of the canonical ensemble, we then obtain

$$\begin{aligned} \log Z(\beta) &= S - \beta M = (1 - M \partial_M) S \\ \log \tilde{Z}(\tilde{\beta}) &= \tilde{S} - \tilde{\beta} M = (1 - M \partial_M) \tilde{S}, \end{aligned} \quad (2.7)$$

where we have used the fact that the perturbed black hole at inverse temperature  $\beta$  has the same mass as the unperturbed black hole at inverse temperature  $\tilde{\beta}$ . Combining eqs. (2.5), (2.6), and (2.7), we cancel terms to obtain

$$\Delta S > 0, \quad (2.8)$$

establishing our claim. The above argument accords with the natural intuition that constraining microscopic states, i.e., heavy fields, to be non-dynamical will decrease the entropy.

Let us comment briefly on a subtle but important caveat to the above arguments. The inequality in eq. (2.5) crucially assumes that on the black hole solution the Euclidean action is not just an extremum but specifically a local minimum. The latter condition guarantees the stability of the Euclidean action under small off-shell perturbations. As is well known, however, the Euclidean path integral suffers from saddle-point instabilities mediated by conformal perturbations of the metric that are unbounded from below. Fortunately, it was shown in refs. [19, 20] that these particular modes are actually a gauge artifact. For a certain

orthogonal decomposition of the metric, the offending conformal mode can be completely decoupled from the physical degrees of freedom. With an appropriate contour of integration it is then possible to path integrate over this mode to yield a convergent final expression.

Later on, an analysis of the Euclidean Schwarzschild solution [21] revealed a bona fide instability coming from a certain non-conformal perturbation about the background solution. This result has been interpreted as evidence that this solution actually describes a tunneling event from a hot background spacetime into a large black hole [22, 23]. Later analyses [24–26] support these claims and moreover show a direct correlation between the existence of negative modes and the thermodynamic instability that arises from negative specific heat. To our knowledge, in all cases considered these saddle-point instabilities disappear when the specific heat is positive, which for example in  $D = 4$  requires a black hole with  $q/m > \sqrt{3}/2$  in natural units.

For the remainder of this paper we restrict to black holes within this thermodynamically stable window of charge-to-mass ratios so that the extremum of the Euclidean action is a local minimum rather than a saddle point and our proof of eq. (2.8) applies. Crucially, this range of parameters includes highly charged black holes, so the results of this section can be used in our discussion of the WGC later on.<sup>4</sup>

We believe that eq. (2.8) is likely true even after relaxing some of the assumptions outlined in section 2.1, specifically those requiring a tree-level quantum field theoretic ultraviolet completion. In particular, from eqs. (2.6) and (2.7) it is obvious that the positivity condition in eq. (2.8) is mathematically equivalent to an inequality of free energies,

$$F(\beta) < \tilde{F}(\beta), \quad (2.9)$$

which says that the free energy of the perturbed black hole is less than that of a Reissner-Nordström black hole at the same temperature. Here we emphasize that the former is computed in the theory defined by  $\mathcal{L}$  and the latter corresponds to  $\tilde{\mathcal{L}}$ . It is quite possible that the free energy condition in eq. (2.9) holds in complete generality, e.g., including quantum corrections.

### 2.3 Explicit example

It is instructive to examine the above arguments for the explicit example of a massive, gravitationally-coupled scalar field. The Euclidean action for the ultraviolet completion is<sup>5</sup>

$$I_{UV}[g, A, \phi] = \int d^D x \sqrt{g} \left[ -\frac{1}{2\kappa^2} R + \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \left( \frac{a_\phi}{\kappa} R + b_\phi \kappa F_{\mu\nu} F^{\mu\nu} \right) \phi + \frac{1}{2} \nabla_\mu \phi \nabla^\mu \phi + \frac{1}{2} m_\phi^2 \phi^2 \right], \quad (2.10)$$

<sup>4</sup>The stability results of refs. [19–26] were obtained in the context of pure gravity and Einstein-Maxwell theory. However, they should also apply in the presence of additional heavy fields, which at low energies only produce small corrections to the leading-order black hole solutions. The precise crossover from positive to negative specific heat may be slightly shifted by the effects of the corresponding higher-dimension operators but this has no impact on the thermodynamic stability of highly charged black holes, which are safely within this window.

<sup>5</sup>Here we ignore boundary terms since we will be interested only in the low-energy corrections generated from integrating out heavy fields. These states produce higher-derivative effective operators whose effects fall off quickly with distance and are thus subdominant to bulk action contributions.

out of notational convenience dropping the hatted convention employed previously. The coupling constants  $a_\phi$  and  $b_\phi$  are dimensionless and have indefinite sign. The classical solution for  $\phi$  is

$$\phi = \frac{1}{\nabla^2 - m_\phi^2} \left( \frac{a_\phi}{\kappa} R + b_\phi \kappa F_{\mu\nu} F^{\mu\nu} \right). \quad (2.11)$$

Plugging back into the Euclidean action, we obtain

$$I[g, A] = \int d^D x \sqrt{g} \left[ -\frac{1}{2\kappa^2} R + \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2m_\phi^2} \left( \frac{a_\phi}{\kappa} R + b_\phi \kappa F_{\mu\nu} F^{\mu\nu} \right)^2 \right], \quad (2.12)$$

where all gradient terms from the  $\phi$  solution are negligible in the low-energy limit. Next, consider the Euclidean action for the full theory, given a field configuration where  $\phi$  is set strictly to zero. The resulting expression is the Euclidean action for pure Einstein-Maxwell theory,

$$I_{\text{UV}}[g, A, 0] = \int d^D x \sqrt{g} \left[ -\frac{1}{2\kappa^2} R + \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \right] = \tilde{I}[g, A]. \quad (2.13)$$

Putting this all together, we learn that

$$I[g, A] < I[\tilde{g}, \tilde{A}] < \tilde{I}[\tilde{g}, \tilde{A}]. \quad (2.14)$$

The first inequality follows because the action is minimized on the solutions to the classical equations of motion for thermodynamically stable black holes. The second inequality follows because eq. (2.12) differs from eq. (2.13) by a negative-definite contribution. This relation between Euclidean actions then implies eq. (2.9) in the saddle-point approximation.

## 2.4 Unitarity and monotonicity

From section 2.3 it is clear that the entropy inequality  $\Delta S > 0$  is very closely related to unitarity. In particular, the relative signs derived in the previous example hinged critically on the absence of tachyons or ghosts in the ultraviolet completion. This is not so surprising, since the presence of such pathologies introduce saddle-point instabilities on a general background, be it flat space or a black hole. It would be interesting to draw a direct connection between our results and previous discussions of unitarity and analyticity [11, 27].

There is also an interesting connection between our results and monotonicity theorems along renormalization group flows [28]. Our proof of  $\Delta S > 0$  was framed in terms of integrating out all heavy fields at once. However, if the spectrum of particles is hierarchical, then this logic can be applied at each mass threshold in sequence. The total entropy shift is then

$$\Delta S = \int_{\text{UV}}^{\text{IR}} dS, \quad (2.15)$$

where the differential entropy  $dS > 0$  contributed by each state is positive. Extrapolating from this classical result, it is reasonable to conjecture that such a positivity condition persists at the quantum level. Indeed, as we will see in section 6.2, the renormalization of pure Einstein-Maxwell theory accords with this expectation.

It is known that the quantum entropy corrections computed from the Euclidean path integral are in close relation with the entanglement entropies of the corresponding modes,

where the horizon is the entangling surface (see refs. [29, 30] and references therein). Since entanglement entropy is intrinsically positive, so too are the quantum entropy corrections, to the extent to which they are equivalent. We will comment on this connection in more detail in section 3.2.

### 3 Classical vs. quantum

Up until now we have focused on classical corrections to the entropy, ignoring all loop effects. As we will see, there exists a regime of black hole masses in which the classical contributions dominate over the quantum. In this case,  $\Delta S > 0$  according to the proof presented in the previous section. In what follows, we estimate and compare the characteristic size of the leading tree-level and loop-level corrections to the black hole entropy.

#### 3.1 Leading contributions

For concreteness, consider a scalar  $\phi$  of mass  $m_\phi$ . As per the assumptions of the previous section, we assume that this field has the usual minimal coupling to gravitons but also direct couplings to the curvature and electromagnetic field strength. Conservatively, we assume that these couplings are at least of gravitational strength, so the interactions go as  $\sim \phi R/\kappa$  and  $\sim \kappa \phi F^2$ . Here it will be convenient to define a set of rescaled higher-dimension operator coefficients,

$$d_{1,2,3} = \kappa^2 c_{1,2,3}, \quad d_{4,5,6} = c_{4,5,6}, \quad d_{7,8} = \kappa^{-2} c_{7,8}, \quad (3.1)$$

which are the dimensionally natural basis in which to express quantities. All of these rescaled coefficients have mass dimension  $[d_i] = -2$ .

**Tree Level.** The dominant contributions coming from tree-level  $\phi$  particle exchange enter as corrections to the  $R^2$ ,  $RF^2$ , and  $F^4$  operators of size

$$\delta(d_i) \sim \frac{1}{m_\phi^2} \quad (\text{tree}). \quad (3.2)$$

Here each contribution scales with a factor of  $1/m_\phi^2$  coming from the  $\phi$  propagator denominator.

**Loop Level.** First, let us consider loop corrections involving purely gravitational interactions. At one loop, the leading contributions enter through the renormalization of the gravitational constant,

$$\delta(\kappa^{-2}) \sim m_\phi^{D-2} \quad (\text{loop}), \quad (3.3)$$

which follows straightforwardly from dimensional analysis. At loop level, gravitational interactions also yield contributions to higher-dimension operators,

$$\delta(d_i) \sim \kappa^2 m_\phi^{D-4} \quad (\text{loop}), \quad (3.4)$$

which are always subdominant to eq. (3.2).

Loops involving gauge interactions will similarly renormalize the gauge coupling as well as the higher-dimension operator coefficients, yielding contributions that scale as eq. (3.3) and eq. (3.4) but with enhancement factors proportional to the charge-to-mass ratios of fundamental charged particles. In principle, these contributions can dominate. For example, in the standard model, the leading contributions to the Euler-Heisenberg Lagrangian come from loops of electrons. However, as shown in ref. [12], this only happens when there are fundamental charged particles with large charge-to-mass ratios. In this case there is no claim to prove, since WGC is already satisfied. For this reason, we restrict our consideration to theories where all fundamental charged particles fail or are near the WGC bound without satisfying it. In this limit gauge interactions are of the same strength as gravitational interactions so the leading tree-level and loop-level corrections from both scale as in eq. (3.2) and eq. (3.3).

### 3.2 Region of interest

From eqs. (3.2), (3.3), and (3.4), we can estimate the corresponding corrections to the black hole entropy, which takes the schematic form

$$S \sim \frac{\rho^{D-2}}{\kappa^2} + \rho^{D-2} m_\phi^{D-2} + \rho^{D-4} m_\phi^{D-4} + \frac{\rho^{D-4}}{\kappa^2 m_\phi^2} + \dots, \quad (3.5)$$

where  $\rho$  is the radius of the black hole. The first term is the Bekenstein-Hawking entropy, the second term is the quantum correction to the gravitational constant, and the third and fourth terms are the quantum and classical corrections to the higher dimension operators, respectively. Demanding that classical entropy corrections dominate over all quantum corrections requires that

$$\rho \ll \frac{1}{\kappa m_\phi^{D/2}}. \quad (3.6)$$

Crucially, since  $m_\phi$  is much smaller than the Planck scale, this constraint is consistent with  $\rho \gg 1/m_\phi$ , which is necessary to remain within the regime of validity of the effective field theory. For the remainder of this paper we focus on this regime of black hole masses.

Before moving on, let us comment briefly on the expectation of positivity for the quantum entropy corrections. While our results only rely on positivity of the classical contribution, it is reasonable to conjecture that the same might apply to quantum corrections. It is known, however, that the quantum contributions in eq. (3.4) have indefinite sign and in  $D = 4$  these correspond to well-studied logarithmic corrections to black hole entropy [31–34]. Nevertheless, these signs do not matter because we have already shown that these corrections are parametrically subdominant to the contributions from eq. (3.3) related to the renormalization of the gravitational constant.

Meanwhile, corrections of the latter type have also been computed via heat kernel methods and found to be positive for minimally-coupled spin 0 and 1/2 particles but negative for spin 0 particles with non-minimal couplings [35] as well as spin 1 and spin 2 particles [36, 37]. There is, however, a longstanding debate over the physical meaning of these negative corrections. They indicate a naive mismatch with calculations of quantum

field theoretic entanglement entropy, which is manifestly positive. While these contributions have been understood as the entanglement of certain edge modes [38, 39], the sign of the leading power-law divergence was also shown to be regulator-dependent. In general, these negative corrections are formally power-law divergent and scheme-dependent and such quantities should be at least partly absorbed into the renormalized gravitational constant [31, 37, 40, 41]. For a consistent ultraviolet completion, all divergences will disappear and the residual corrections will be finite. On physical grounds, it is expected that, if properly regulated, these quantum entropy corrections will be manifestly positive as expected from the manifest positivity of entanglement entropy.

## 4 Black hole spacetime

We now turn to the study of a spherically symmetric, positively charged black hole of mass  $M$  and charge  $Q$  in the presence of low-energy corrections to pure Einstein-Maxwell theory. For simplicity, we restrict to  $D = 4$  dimensions for the remainder of the body of this paper, but all of our results generalize to arbitrary spacetime dimension  $D \geq 4$ , as shown in appendix A.

Our aim is to derive new bounds on the higher-dimension operator coefficients  $c_i$ . As noted previously, this restricts our consideration to black holes large enough that the effective field theory is valid but small enough to satisfy eq. (3.6), so  $1/m_\phi \ll \rho \ll 1/\kappa m_\phi^2$ , where  $m_\phi$  is the mass scale of the new states. This range always exists provided there is a parametric separation between  $m_\phi$  and the Planck scale. Furthermore, we consider the thermodynamically stable regime where  $q/m > \sqrt{3}/2$  so the specific heat is positive.

Note that the mass and charge are defined at spatial infinity. Since we will only consider static spacetimes, the ADM and Komar formulations of these quantities are equivalent. Explicitly, the Komar mass and charge are<sup>6</sup>

$$M \sim \frac{1}{\kappa^2} \int_{i^0} n_\mu \sigma_\nu \nabla^\mu K^\nu \quad \text{and} \quad Q \sim \int_{i^0} n_\mu \nabla_\nu F^{\mu\nu}, \quad (4.1)$$

where the integral region  $i^0$  is spatial infinity,  $n$  is the unit timelike normal vector,  $\sigma$  is the unit spatial outward-pointing normal vector, and  $K$  is the timelike Killing vector. Since the integral is evaluated at spatial infinity, only the leading behavior at large  $r$  contributes to these expressions. Because higher-dimension operators correct the metric and gauge field at subleading order in  $r$ , they do not affect the definition of the asymptotic mass and charge.

### 4.1 Unperturbed solution

The unperturbed theory is described by the Lagrangian for Einstein-Maxwell theory in eq. (1.3). For notational convenience we will sometimes describe the mass and charge in

---

<sup>6</sup>Here we take all black holes to be of definite mass even though in practice they have a small width given by their inverse lifetime. For instance, as classically stable objects, black holes can only decay quantum mechanically via Hawking emission and Schwinger pair production processes, which are suppressed by additional factors of the gravitational constant. Moreover, decays into smaller black holes will proceed via non-perturbative gravitational effects, which are exponentially suppressed.

natural units of the gravitational constant,<sup>7</sup>

$$m = \frac{\kappa^2 M}{8\pi} \quad \text{and} \quad q = \frac{\kappa Q}{4\sqrt{2}\pi}. \quad (4.2)$$

We also define the extremality parameter

$$\xi = \sqrt{1 - \frac{q^2}{m^2}}, \quad (4.3)$$

where a neutral black hole corresponds to  $\xi = 1$  and an extremal black hole corresponds to  $\xi = 0$ . As noted previously, we will consider black holes with positive specific heat, corresponding to  $q/m > \sqrt{3}/2$ , or equivalently,  $\xi < 1/2$ .

The solution is the Reissner-Nordström black hole, whose metric takes the static and spherically symmetric form

$$ds^2 = \tilde{g}_{\mu\nu} dx^\mu dx^\nu = -\tilde{f}(r) dt^2 + \frac{1}{\tilde{g}(r)} dr^2 + r^2 d\Omega^2, \quad (4.4)$$

where the unperturbed metric components are

$$\tilde{f}(r) = \tilde{g}(r) = 1 - \frac{2m}{r} + \frac{q^2}{r^2} \quad (4.5)$$

and the unperturbed electromagnetic field strength is

$$\tilde{F}_{\mu\nu} dx^\mu \wedge dx^\nu = \frac{Q}{4\pi r^2} dt \wedge dr. \quad (4.6)$$

The unperturbed event horizon is the outer horizon of the Reissner-Nordström black hole and is located at the radius  $r = \tilde{\rho}$ , where

$$\tilde{\rho} = m + \sqrt{m^2 - q^2} = m(1 + \xi). \quad (4.7)$$

The absence of a naked singularity implies that the charge is bounded by the inequality

$$\frac{q}{m} \leq 1, \quad (4.8)$$

which is saturated in the case of an extremal black hole.

## 4.2 Perturbed solution

In the presence of the higher-dimension operators in eq. (1.4), the perturbed metric takes the form

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -f(r) dt^2 + \frac{1}{g(r)} dr^2 + r^2 d\Omega^2, \quad (4.9)$$

where the metric components are complicated functions of the coefficients  $c_i$ . However, as shown in refs. [42, 43], it is straightforward to compute corrections to the Reissner-Nordström solution order-by-order in  $c_i$ . Following the prescription in ref. [42] applied

---

<sup>7</sup>In  $D$  dimensions, the mass dimensions of various quantities are  $[\kappa^2] = 2 - D$ ,  $[R] = 2$ ,  $[F] = D/2$ ,  $[M] = 1$ ,  $[Q] = 2 - D/2$ ,  $[m] = [q] = 3 - D$ ,  $[c_{1,2,3}] = D - 4$ ,  $[c_{4,5,6}] = -2$ ,  $[c_{7,8}] = -D$ , and  $[d_i] = -2$ .

to the higher-dimension operators in eq. (1.4), we find that at first order in  $c_i$  the radial component of the metric is

$$g(r) = 1 - \frac{2m}{r} + \frac{q^2}{r^2} - \frac{q^2}{r^6} \left\{ \begin{aligned} &\frac{4}{5}(d_2 + 4d_3)(6q^2 - 15mr + 10r^2) \\ &+ 8d_4(3q^2 - 7mr + 4r^2) + \frac{4}{5}d_5(11q^2 - 25mr + 15r^2) \\ &+ \frac{4}{5}d_6(16q^2 - 35mr + 20r^2) + \frac{8}{5}(2d_7 + d_8)q^2 \end{aligned} \right\}, \quad (4.10)$$

where the coefficients  $d_i$  are defined in terms of  $c_i$  in eq. (3.1).

## 5 Calculation of entropy

### 5.1 Wald entropy formula

We now compute the entropy corrections to black holes of size much greater than the Compton wavelengths of the heavy modes. A major advantage of this approach is that the effects of all short-distance degrees of freedom are encoded purely by higher-dimension operators. Moreover, even though these states are absent from the low-energy theory, their contributions to the entropy are fully accounted for by the Wald formula in eq. (1.6).

For a spherically symmetric spacetime, the integral in eq. (1.6) is trivial, yielding

$$S = -2\pi A \frac{\delta \mathcal{L}}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\nu} \epsilon_{\rho\sigma} \Big|_{g_{\mu\nu}, \rho}, \quad (5.1)$$

where all quantities are evaluated for the perturbed metric and perturbed horizon radius,  $\rho = \tilde{\rho} + \Delta\rho$ . The perturbed horizon area is  $A = 4\pi\rho^2$  and the binormal is

$$\epsilon_{\mu\nu}(r) = \sqrt{\frac{f(r)}{g(r)}} (\delta_\mu^t \delta_\nu^r - \delta_\mu^r \delta_\nu^t). \quad (5.2)$$

Expanding the area  $A = \tilde{A} + \Delta A$  and the Lagrangian  $\mathcal{L} = \tilde{\mathcal{L}} + \Delta\mathcal{L}$  in perturbations, we obtain

$$S = -2\pi \left( \tilde{A} \frac{\delta \tilde{\mathcal{L}}}{\delta R_{\mu\nu\rho\sigma}} + \tilde{A} \frac{\delta \Delta\mathcal{L}}{\delta R_{\mu\nu\rho\sigma}} + \Delta A \frac{\delta \tilde{\mathcal{L}}}{\delta R_{\mu\nu\rho\sigma}} + \dots \right) \epsilon_{\mu\nu} \epsilon_{\rho\sigma} \Big|_{g_{\mu\nu}, \rho}, \quad (5.3)$$

where  $\tilde{A} = 4\pi\tilde{\rho}^2$  and the ellipses denote terms that are higher order in perturbations. The first term is straightforwardly obtained by differentiating eq. (1.3) with respect to the Riemann tensor,

$$\frac{\delta \tilde{\mathcal{L}}}{\delta R_{\mu\nu\rho\sigma}} = \frac{1}{2\kappa^2} g^{\mu\rho} g^{\nu\sigma}, \quad (5.4)$$

where the proper (anti)symmetrization of indices on the right-hand side is implicit. Since the binormal is normalized as  $\epsilon_{\mu\nu} \epsilon^{\mu\nu} = -2$ , the first term in eq. (5.3) is simply the unperturbed black hole entropy  $\tilde{S}$  defined in eq. (1.7). Moving this term to the left-hand side of eq. (5.3), we obtain an expression for the difference in entropies,

$$\Delta S = \Delta S_I + \Delta S_H, \quad (5.5)$$



split into an “interaction” and “horizon” contribution. Because we are working at first order in perturbations, both of these terms should be evaluated on the unperturbed metric and horizon radius. The interaction contribution,

$$\Delta S_I = -2\pi \tilde{A} \frac{\delta \Delta \mathcal{L}}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\nu} \epsilon_{\rho\sigma} \Big|_{\tilde{g}_{\mu\nu}, \tilde{\rho}}, \quad (5.6)$$

appears because the interactions of photons and gravitons are modified at low energies. Meanwhile, the horizon contribution,

$$\Delta S_H = -2\pi \Delta A \frac{\delta \tilde{\mathcal{L}}}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\nu} \epsilon_{\rho\sigma} \Big|_{\tilde{g}_{\mu\nu}, \tilde{\rho}} = \frac{2\pi}{\kappa^2} \Delta A, \quad (5.7)$$

is present because higher-dimension operators modify the black hole background, thus shifting the location of the horizon. Here we have substituted in eq. (5.4) to write the right-hand side of this expression as simply the shift of the horizon area.

## 5.2 Interaction contribution

To obtain the interaction contribution to the entropy shift we compute

$$\begin{aligned} \frac{\delta \Delta \mathcal{L}}{\delta R_{\mu\nu\rho\sigma}} = & 2c_1 R g^{\mu\rho} g^{\nu\sigma} + 2c_2 R^{\mu\rho} g^{\nu\sigma} + 2c_3 R^{\mu\nu\rho\sigma} \\ & + c_4 F_{\alpha\beta} F^{\alpha\beta} g^{\mu\rho} g^{\nu\sigma} + c_5 F^\mu{}_\alpha F^{\rho\alpha} g^{\nu\sigma} + c_6 F^{\mu\nu} F^{\rho\sigma}, \end{aligned} \quad (5.8)$$

where proper (anti)symmetrization on indices is left implicit as before. Substituting the unperturbed black hole background into eq. (5.8) and evaluating eq. (5.6), we obtain

$$\Delta S_I = \tilde{S} \times \frac{2}{m^2(1+\xi)^3} [8d_3 - 2(1-\xi)(d_2 + 6d_3 + 2d_4 + d_5 + 2d_6)], \quad (5.9)$$

written in terms of the coefficients defined in eq. (3.1). Setting  $\xi = 1$  in our expression for  $\Delta S_I$  in eq. (5.9) agrees with the expressions in ref. [44] as well as their generalization to arbitrary dimension in ref. [45].

## 5.3 Horizon contribution

The horizon contribution to the entropy shift depends on the location of the perturbed horizon. Since the spacetime is static, the horizon is determined by zeros of the metric components  $f(r)$  and  $g(r)$  defined in eq. (4.9). On general grounds,  $f(r)$  and  $g(r)$  have coincident zeros since otherwise the spacetime would contain a region with non-Lorentzian signature. Moreover, we have verified by explicit calculation that  $f(r)$  and  $g(r)$  share the same zeros at first order in perturbations.

The perturbed horizon is located at radius  $\rho = \tilde{\rho} + \Delta\rho$ . To compute  $\rho$ , we expand  $g(r) = \tilde{g}(r) + \Delta g(r)$  at first order in perturbations, as defined in eq. (4.10). The perturbed horizon radius then satisfies the equation

$$0 = g(\rho) = \tilde{g}(\tilde{\rho}) + \Delta g(\tilde{\rho}) + \Delta\rho \partial_{\tilde{\rho}} \tilde{g}(\tilde{\rho}). \quad (5.10)$$

The first term on the right-hand side vanishes by the definition of the unperturbed horizon radius. Solving for the horizon shift, we find

$$\Delta\rho = -\frac{\Delta g(\tilde{\rho})}{\partial_{\tilde{\rho}}\tilde{g}(\tilde{\rho})}. \quad (5.11)$$

At first order, the perturbed horizon area is then given by

$$\Delta A = A - \tilde{A} = 8\pi\tilde{\rho}\Delta\rho = -\frac{8\pi\tilde{\rho}\Delta g(\tilde{\rho})}{\partial_{\tilde{\rho}}\tilde{g}(\tilde{\rho})}. \quad (5.12)$$

Inputting the perturbed metric in eq. (4.10) and evaluating eq. (5.7), we obtain

$$\Delta S_H = \tilde{S} \times \frac{4(1-\xi)}{5m^2\xi(1+\xi)^3}[(1+4\xi)(d_2+4d_3+d_5+d_6)+10\xi d_4+2(1-\xi)(2d_7+d_8)]. \quad (5.13)$$

Note that the horizon contribution to the entropy shift is divergent in the  $\xi \rightarrow 0$  limit corresponding to an extremal black hole. Physically, this occurs because the inner and outer horizons become degenerate, so  $\partial_{\tilde{\rho}}\tilde{g}(\tilde{\rho}) \rightarrow 0$ . In this case, eq. (5.10) implies that for some fixed contribution  $\Delta g(\tilde{\rho})$  from higher-dimension operators, the horizon must shift by a parametrically large amount  $\Delta\rho$  in order to maintain the horizon condition.

Of course, the strict  $\xi \rightarrow 0$  limit is pathological since this produces an infinite entropy shift, signaling a breakdown of perturbation theory. Demanding that the shift in entropy be much smaller than the unperturbed entropy, we obtain the constraint

$$\xi \gg \frac{|d_i|}{m^2}. \quad (5.14)$$

As shown in eq. (3.6), the classical effects of higher-dimension operators are only dominant over the quantum for black hole radii smaller than a certain value,  $\rho \ll 1/\kappa m_\phi^2$ . Using that  $d_i \sim 1/m_\phi^2$  for a tree-level ultraviolet completion, eq. (5.14) becomes  $\xi \gg \kappa^2 m_\phi^2$ .

Perturbativity also requires that the shift in the inverse temperature  $\Delta\beta$  be subdominant to the background inverse temperature  $\tilde{\beta}$  of the unperturbed Reissner-Nordström black hole. As a consistency check, we have verified that  $\beta = \partial_M S$  agrees with the surface gravity of the perturbed black hole metric. For highly charged black holes, the unperturbed inverse temperature goes as  $\tilde{\beta} \sim m/\xi$  while  $\Delta\beta \sim d_i/m\xi^3$ . Demanding that the correction be smaller than the leading contribution,  $\Delta\beta \ll \tilde{\beta}$ , we obtain the stronger condition

$$\xi \gg \frac{|d_i|^{1/2}}{m}. \quad (5.15)$$

Combining this with the upper bound on  $\rho$  from eq. (3.6) and the scaling of  $d_i$ , we obtain

$$\xi \gg \kappa m_\phi. \quad (5.16)$$

Thus, it is always possible to take the limit of a highly charged black hole,  $\xi \ll 1$ , provided the mass scale  $m_\phi$  of the heavy fields is far below the Planck scale as we have assumed throughout.

## 6 New positivity bounds

### 6.1 General bounds

The total entropy shift  $\Delta S = \Delta S_I + \Delta S_H$  is obtained by adding eq. (5.9) and eq. (5.13), yielding

$$\Delta S = \tilde{S} \times \frac{4}{5m^2\xi(1+\xi)^3} \times \left[ (1-\xi)^2(d_2 + d_5) + 2(2 + \xi + 7\xi^2)d_3 + (1-\xi)(1-6\xi)d_6 + 2(1-\xi)^2(2d_7 + d_8) \right]. \quad (6.1)$$

As proven in section 2, the entropy shift is positive under the assumptions we have stated. Combining eq. (1.8) and eq. (6.1), we obtain a family of positivity bounds,

$$(1-\xi)^2 d_0 + 20\xi d_3 - 5\xi(1-\xi)(2d_3 + d_6) > 0, \quad (6.2)$$

where we have defined the parameter

$$d_0 = d_2 + 4d_3 + d_5 + d_6 + 4d_7 + 2d_8. \quad (6.3)$$

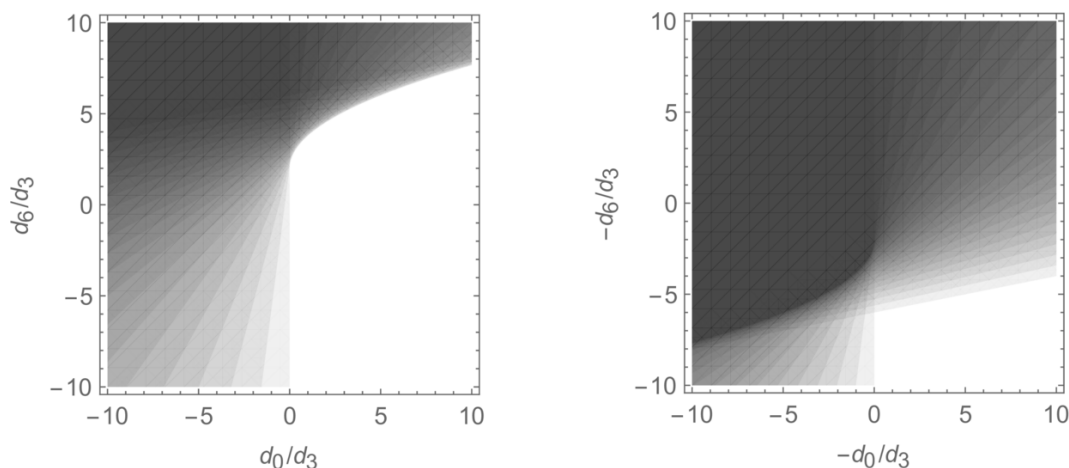
The bound in eq. (6.2) is the main result of this work: a consistency condition on the coefficients of higher-dimension operator corrections to Einstein-Maxwell theory following from the positivity of corrections to the black hole entropy. The generalizations of eqs. (6.1), (6.2), and (6.3) to arbitrary dimension  $D$  are derived and presented in eqs. (A.18), (A.19), and (A.20) in appendix A.

As discussed in section 2, our proof of  $\Delta S > 0$  applies to thermodynamically stable black holes, restricting consideration to the window  $\xi \in (0, 1/2)$ . The full space of bounds over this range defines a convex region in the space of coefficients  $d_i$ . Thus the full space of positivity constraints, depicted in figure 2 for  $D = 4$ , is stronger than those implied by any finite set of choices for  $\xi$ .

We have derived these positivity conditions from a particular physical setup: a black hole of a given charge-to-mass ratio corresponding to  $\xi \in (0, 1/2)$  and mass consistent with eq. (3.6) so that our proof of  $\Delta S > 0$  applies. Despite this specificity, we emphasize that the resulting bounds in eq. (6.2) are consistency conditions on the effective action and thus hold *independently of the background*.

### 6.2 Examples and consistency checks

Our bound in eq. (6.2) and its higher-dimensional generalization in eq. (A.19) pass a number of explicit checks. To begin, we emphasize that these inequalities are invariant under a change of the field basis, which by reparameterization theorems should leave physical observables unchanged. As discussed in appendix B, metric field redefinitions shift the higher-dimension operator coefficients  $d_i$  in a way that renders individual coefficients operationally meaningless. By studying these transformations, one can build a basis of field redefinition invariant combinations of coefficients, of which  $d_0, d_3, d_6$  are a subset. Remarkably, while the separate contributions to the entropy from  $\Delta S_I$  and  $\Delta S_H$  are *not* invariant under field redefinitions, their sum  $\Delta S$  is invariant for all  $\xi$  in arbitrary dimension  $D$ , as shown in appendix B. Field redefinition invariance of the inequality  $\Delta S > 0$  is a prerequisite for this bound to have physical meaning.



**Figure 2.** Constraints on higher-dimension operator coefficients derived from black hole entropy. The shaded regions are excluded, with the gradations corresponding to incremental values of  $\xi \in (0, 1/2)$ . The left and right panels correspond to  $d_3 > 0$  and  $d_3 < 0$ , respectively. In either case,  $d_0 < 0$  is forbidden so  $d_0 > 0$  and the WGC is automatically satisfied.

On physical grounds, it is natural to expect that any positivity condition is preserved under renormalization group flow into the infrared. This is true because a consistent theory should continue to be consistent at arbitrarily long distances. Interestingly, this expectation agrees with known results on the one-loop divergences of pure Einstein-Maxwell theory in  $D = 4$ , which enter solely through the  $R_{\mu\nu}R^{\mu\nu}$  operator [1]. The sign of this divergence is consistent with a negative beta function, indicating that the coefficient of this operator indeed increases in the infrared, consistent with eq. (6.2).

We can also study eq. (A.19) in simple concrete examples. First, consider any theory in general dimension  $D$  in which the strength of gravity is negligible relative to gauge interactions. In this limit all higher-dimension operators are vanishing except for  $d_7$  and  $d_8$ , which control the leading contributions to photon self-interactions in the Euler-Heisenberg effective action. Applying the logic of ref. [11], we have computed the four-photon scattering amplitude at low energies and found that dispersion relations imply the positivity conditions  $2d_7 + d_8 > 0$  and  $d_8 > 0$ , corresponding to different choices of external polarizations. In this case, the former inequality exactly implies that  $d_0 > 0$  in general dimension  $D$ , thus providing a consistency check of eqs. (6.2) and (A.19).

Second, we examine the scalar model described in section 2.3. Translating from Euclidean to Lorentzian signature, we obtain the higher-dimension operator coefficients

$$d_i = \frac{1}{2m_\phi^2} \times \{a_\phi^2, 0, 0, 2a_\phi b_\phi, 0, 0, b_\phi^2, 0\}, \quad (6.4)$$

which, written in terms of  $d_0$  using eq. (A.20), give

$$d_0 = \frac{D-3}{8m_\phi^2} [(D-4)a_\phi + 2(D-2)b_\phi]^2, \quad (6.5)$$

which is a perfect square, so the bound in eq. (A.19) is again satisfied.

Third, we study the low-energy description of the heterotic string, for which the higher-dimension operators have coefficients as given in refs. [42, 46],

$$d_i = \frac{\alpha'}{64} \times \{4, -16, 4, 0, 0, 0, -3, 12\}, \quad (6.6)$$

where we have absorbed the dependence on the dilaton expectation value into  $\alpha'$ . Plugging these parameters into eq. (A.19) yields  $(6D^2 - 30D + 37)\xi^2 + 2(D - 2)\xi + 2D - 5 > 0$ , which holds for all  $\xi \in (0, 1)$  and  $D > 3$ . Thus, we find that eq. (A.19) is actually satisfied even beyond the range of thermodynamic stability and the critical dimension of the string.

### 6.3 The weak gravity conjecture

In the limit of a highly charged black hole we take  $\xi \ll 1$  and our general bound in eq. (6.2) becomes

$$d_0 > 0, \quad (6.7)$$

with  $d_0$  defined in terms of the other  $d_i$  in eq. (6.3).

As it turns out, this inequality is intimately connected with the extremality condition for a black hole. To see why, consider the unperturbed Reissner-Nordström solution, for which the extremal charge-to-mass ratio is  $\tilde{z} = q/m = 1$ . In general, quantum corrections to the gravitational constant and electric charge will renormalize the right-hand side of this condition. Since these contributions affect black holes of all masses universally, their effects can simply be absorbed into the definitions of mass and charge. Meanwhile, higher-dimension operators also shift the maximum charge-to-mass ratio permitted for a physical black hole, i.e., a black hole free from naked singularities. In contrast, these corrections are mass-dependent, so they induce a physical shift of the extremality condition to  $z = \tilde{z} + \Delta z$  [42].

To compute this shift, we analyze the metric component  $g(r, z)$ , interpreted as a function of both the radius and the charge-to-mass ratio. The shifted horizon is defined by the condition  $g(\rho, z) = 0$ . At linear order in perturbations, this condition becomes

$$0 = g(\rho, z) = \tilde{g}(\tilde{\rho}, \tilde{z}) + \Delta g(\tilde{\rho}, \tilde{z}) + \Delta \rho \partial_{\tilde{\rho}} \tilde{g}(\tilde{\rho}, \tilde{z}) + \Delta z \partial_{\tilde{z}} \tilde{g}(\tilde{\rho}, \tilde{z}). \quad (6.8)$$

Since the unperturbed black hole is extremal, the first and third terms on the right-hand side are zero. Solving for the shift in the charge-to-mass ratio yields

$$\Delta z = -\frac{\Delta g(\tilde{\rho}, \tilde{z})}{\partial_{\tilde{z}} \tilde{g}(\tilde{\rho}, \tilde{z})}. \quad (6.9)$$

By explicit calculation, the charge-to-mass ratio shift is

$$\Delta z = \frac{2d_0}{5m^2} > 0, \quad (6.10)$$

which is positive according to eq. (6.7).

As shown in ref. [42], if  $\Delta z$  is positive then small black holes *automatically* satisfy the WGC [7], which posits that an Abelian gauge theory consistently coupled to quantum

gravity must contain a state with charge-to-mass ratio greater than unity. In its original formulation [7], the WGC was presented with several compelling justifications. This included overwhelming circumstantial evidence from a long list of explicit examples in quantum field theory and string theory. A more direct argument was also presented in the form of an elegant thought experiment involving stable black hole remnants [47–51]. In particular, due to mass and charge conservation, a charged black hole is stable unless there exist lighter states with a higher charge-to-mass ratio, into which the black hole can decay. In the infinite-mass limit, the charge-to-mass ratio of an extremal black hole is dictated by the Reissner-Nordström solution and approaches unity. Violation of the WGC would then imply the existence of an infinite tower of stable remnants labeled by the extremal black hole mass, asserted as pathological in ref. [7]. While this thought experiment offers some crucial physical intuition for the WGC, it falls short since there do not exist formal proofs that stable black hole remnants are actually inconsistent with more established principles like the covariant entropy bound. In many cases the number of black hole remnants is very large but still finite; furthermore, the states are labeled by distinct charges and are thus in principle distinguishable [52].

On the other hand, if the WGC is satisfied then extremal black holes are unstable to decay. We have shown here that considerations of black hole entropy imply that the charge-to-mass ratio of an extremal black hole increases with decreasing size. In particular, higher-dimension operators induce a positive shift of the extremality bound, but these corrections decouple for large black hole masses. The upshot is then that an extremal black hole of a given mass can always decay into smaller extremal black holes of a greater charge-to-mass ratio, following the upper curve in figure 1. Our bound in eq. (6.7)—and thus our proof of the WGC — generalizes to  $D$  spacetime dimensions, as shown in appendix A.

Let us comment on the relation between our results and previous work connecting black holes and the WGC. First, while the argument in this paper makes critical use of extremal black holes, our reasoning is completely different from the original proposal of ref. [7], hinging instead on the thermodynamic entropy of black holes rather than their stability. More recently, the WGC has also been linked to the cosmic censorship conjecture [53]. We leave an analysis of this and its relationship to black hole entropy for future work.

In other recent studies [52, 54], the WGC has also been evaluated in the context of black hole entropy using methodologies that differ substantially from our own. Both ref. [52] and ref. [54] examine the leading logarithmic corrections to black hole entropy due to the quantum effects of light matter particles. Such effects are relevant for black holes of size parametrically smaller than the Compton wavelength of the matter. We, in contrast, consider the opposite regime, which effectively corresponds to a gapped spectrum.

Furthermore, refs. [52, 54] argue for the inconsistency of WGC violation through quite different means: ref. [52] makes the argument through the appearance of a low cutoff, while ref. [54] employs the second law of thermodynamics. It is crucial to note that our assertion of a positive entropy shift is logically distinct from the second law of thermodynamics, which applies to the difference in entropy before and after a physical process but within the same physical system. Our construction is instead based on the positivity of classical entropy corrections proven in section 2.

Finally, refs. [52, 54] and also another interesting approach [55] all consider concrete models with explicit spectra of charged and neutral scalars and fermions. For this reason, these works at best show that certain WGC-violating theories are inconsistent. This leaves the logical possibility that more complicated theories that violate the WGC might still be judged valid by their analyses. In comparison, our work applies to large black holes in a general low-energy effective theory, which is insensitive to the precise details of the spectrum and hence constitutes a model-independent argument for the WGC.

## 6.4 Entropy, area, and extremality

A priori, it is somewhat miraculous that the entropy constraint in eq. (6.7) is literally equivalent to the extremality condition in eq. (6.10). To briefly summarize, we have shown that

$$\Delta S \sim \Delta \rho \sim \Delta z > 0, \quad (6.11)$$

so the low-energy corrections to the near-extremal black hole entropy, area, and extremality condition are all proportional to each other and all positive.

Why does the same combination of coefficients  $d_0$  appear in all of these inequalities? As it turns out, this connection is not so mysterious once one considers the perturbed metric component  $g(\rho, z)$  in eq. (6.8) as a function of the shift in horizon radius  $\Delta \rho$  and the shift in the charge-to-mass ratio  $\Delta z$ . For a near-extremal black hole of fixed charge and mass, we set  $\Delta z = 0$  and thus  $\Delta \rho = -\Delta g / \partial_{\tilde{\rho}} \tilde{g}$ . On the other hand, if the charge and mass are free but the unperturbed system is exactly extremal, then the  $\Delta \rho$  term drops out and the charge-to-mass ratio shift is  $\Delta z = -\Delta g / \partial_{\tilde{z}} \tilde{g}$ . At the same time, the radial component of the metric  $\tilde{g}$  is by definition spacelike outside the horizon, so  $\partial_{\tilde{\rho}} \tilde{g} > 0$ . Moreover, since  $\tilde{g}$  dictates the gravitational potential at long distances, it decreases with  $m$  and thus increases with the charge-to-mass ratio, so  $\partial_{\tilde{z}} \tilde{g} > 0$ . This logic implies that  $\Delta \rho$  and  $\Delta z$  have the same sign. Since the entropy shift for a near-extremal black hole is dominated by the shift in the horizon,  $\Delta S \sim \Delta \rho$ , we discover that  $\Delta S > 0$ ,  $\Delta \rho > 0$ , and  $\Delta z > 0$  are equivalent bounds.

Conveniently, the above logic immediately extends to the multi-charge generalization of the WGC proposed in ref. [8]. For a theory with multiple Abelian factors, the charge-to-mass ratio defines a vector  $\mathbf{z}$  in charge space. The WGC then mandates that the unit ball representing all possible large black holes be contained within the convex hull spanned by the set of all  $\mathbf{z}$  for the lighter states in the theory. Crucially, for a multi-charged black hole, the perturbed metric only depends on the magnitude of its charge and not the direction. Hence, eq. (6.8) still applies, provided we define  $\tilde{z} = |\tilde{\mathbf{z}}|$  as the magnitude of the charge-to-mass ratio vector of the black hole and  $\Delta z = \Delta \mathbf{z} \cdot \tilde{\mathbf{z}} / |\tilde{\mathbf{z}}|$  as its shift. Repeating exactly the argument of the previous paragraph, we learn that  $\Delta S > 0$ ,  $\Delta \rho > 0$ , and  $\Delta \mathbf{z} \cdot \tilde{\mathbf{z}} > 0$  are all equivalent. The last inequality implies that the extremality condition for a multi-charged extremal black hole is perturbed so that the corresponding unit ball is expanded outward in every direction, thus proving the multi-charge version of the WGC given by the convex hull condition in ref. [8].



## 7 Discussion and conclusions

In this paper, we derived a positivity condition on classical corrections to the Bekenstein-Hawking entropy. For near-extremal black holes this enforces positivity of a certain linear combination of coefficients of higher-dimension operators. This very same combination of couplings corrects the extremality condition for black holes so that their charge-to-mass ratios approach unity from above for increasing size. Large extremal black holes are thus unstable to decay to smaller extremal black holes. Since the latter have charge-to-mass ratios greater than unity, they automatically fulfill the requirement of the WGC.

Our findings leave a number of avenues for future work. First, it would be interesting to determine if entropy considerations have any additional implications for the swampland program, for example by introducing new operators in extended theories like Einstein-dilaton gravity or by considering black holes embedded in asymptotically AdS or dS space, rotating black holes, or black holes of different topologies. Second, one would ideally like to understand the relationship, if any, between these entropy bounds and other constraints on low-energy dynamics coming from causality, analyticity, and unitarity. Indeed, the positivity of entropy shifts discussed in this paper stems from state counting in the ultraviolet, which is highly reminiscent of dispersion relation bounds utilizing the positivity of forward cross-sections [11–14, 27, 56–58] and amplitudes approaches based on the positivity of spectral representations [15, 59, 60].

## Acknowledgments

We thank Nima Arkani-Hamed, Raphael Bousso, Sean Carroll, William Cottrell, Daniel Harlow, Tom Hartman, Yasunori Nomura, Hirosi Ooguri, Matthew Reece, Prashant Saraswat, Gary Shiu, Leo Stein, and Aron Wall for helpful discussions. C.C. and J.L. are supported by a Sloan Research Fellowship and DOE Early Career Award under grant no. DE-SC0010255. G.N.R. is supported by the Miller Institute for Basic Research in Science at the University of California, Berkeley.

## A Generalization to arbitrary dimension

### A.1 Black hole spacetime

In this appendix we generalize all of our results to arbitrary spacetime dimension  $D \geq 4$ . See footnote 7 for the mass dimensions of various quantities. To begin, we consider the Reissner-Nordström metric in  $D$  dimensions,

$$ds^2 = -\tilde{f}(r)dt^2 + \frac{1}{\tilde{g}(r)}dr^2 + r^2 d\Omega_{D-2}^2, \quad (\text{A.1})$$



where  $d\Omega_{D-2}^2$  is line element on the unit  $(D-2)$ -sphere and

$$\tilde{f}(r) = \tilde{g}(r) = 1 - \frac{2\kappa^2 M}{(D-2)\Omega_{D-2}r^{D-3}} + \frac{Q^2\kappa^2}{(D-2)(D-3)\Omega_{D-2}^2r^{2(D-3)}}, \quad (\text{A.2})$$

as before denoting  $\kappa^2 = 8\pi G$ . The electromagnetic field strength is

$$\tilde{F}_{\mu\nu}dx^\mu \wedge dx^\nu = \frac{Q}{\Omega_{D-2}r^{D-2}}dt \wedge dr, \quad (\text{A.3})$$

where the  $(D-2)$ -dimensional area of the unit codimension-two sphere is

$$\Omega_{D-2} = \frac{2\pi^{\frac{D-1}{2}}}{\Gamma(\frac{D-1}{2})}. \quad (\text{A.4})$$

Next, let us define new variables for mass and charge in units of the Planck scale,

$$m = \frac{\kappa^2 M}{(D-2)\Omega_{D-2}}, \quad q = \frac{\kappa Q}{\sqrt{(D-2)(D-3)\Omega_{D-2}}}, \quad (\text{A.5})$$

along with a rescaled radial coordinate,

$$x = r^{D-3}, \quad (\text{A.6})$$

in terms of which the metric component can be written simply as

$$\tilde{g}(r) = 1 - \frac{2m}{x} + \frac{q^2}{x^2}. \quad (\text{A.7})$$

The outer horizon is located at  $x = \tilde{\chi} = \tilde{\rho}^{D-3}$ , where

$$\tilde{\chi} = m + \sqrt{m^2 - q^2} = m(1 + \xi) \quad (\text{A.8})$$

and  $\xi$  is defined as in eq. (4.3). The extremality condition for the background Reissner-Nordström spacetime as before requires  $q/m \leq 1$ . The requirement of thermodynamic stability restricts our consideration to black holes with  $\xi < \frac{D-3}{D-2}$ , for which the specific heat is positive.

Following the perturbative methods of refs. [42, 43], we can compute the metric components at first order in perturbations, finding

$$g(r) = 1 - \frac{2m}{x} + \frac{q^2}{x^2} - \frac{q^2}{x^{\frac{2(2D-5)}{D-3}}} \sum_{i=1}^8 \alpha_i c_i, \quad (\text{A.9})$$

where the coefficients are

$$\begin{aligned}
 \alpha_1 &= \frac{(D-3)(D-4)}{D-2} \left[ 2 \frac{13D^2 - 47D + 40}{3D-7} q^2 - 8(3D-5)mx + 16(D-2)x^2 \right] \\
 \alpha_2 &= 2 \frac{D-3}{D-2} \left[ \frac{8D^3 - 55D^2 + 117D - 76}{3D-7} q^2 - 4(2D^2 - 10D + 11)mx \right. \\
 &\quad \left. + 2(3D-10)(D-2)x^2 \right] \\
 \alpha_3 &= 4 \frac{D-3}{D-2} \left[ \frac{8D^3 - 48D^2 + 87D - 44}{3D-7} q^2 - 2(4D^2 - 17D + 16)mx \right. \\
 &\quad \left. + 8(D-2)(D-3)x^2 - 2(D-2)(D-4) \frac{m^2 x^2}{q^2} \right] \\
 \alpha_4 &= 4(D-3) \left[ \frac{(7D-13)(D-2)}{3D-7} q^2 - 2(3D-5)mx + 4(D-2)x^2 \right] \\
 \alpha_5 &= 2(D-3) \left[ \frac{(5D-9)(D-2)}{3D-7} q^2 - 2(2D-3)mx + 3(D-2)x^2 \right] \\
 \alpha_6 &= 4(D-3) \left[ 4 \frac{(D-2)^2}{3D-7} q^2 - (3D-5)mx + 2(D-2)x^2 \right] \\
 \alpha_7 &= 8 \frac{(D-2)(D-3)^2}{3D-7} q^2 \\
 \alpha_8 &= 4 \frac{(D-2)(D-3)^2}{3D-7} q^2.
 \end{aligned} \tag{A.10}$$

## A.2 Calculation of entropy

As before, the total entropy shift is  $\Delta S = \Delta S_I + \Delta S_H$ , where  $\Delta S_I$ , defined in eq. (5.6), arises from modifications of the low-energy graviton interactions and  $\Delta S_H$ , defined in eq. (5.7), is induced by the shift of the black hole horizon.

To compute the entropy contribution from interactions, we substitute the unperturbed black hole background from section A.1 into eq. (5.8), yielding

$$\begin{aligned}
 \Delta S_I &= \tilde{S} \times \frac{2(D-3)}{m^{\frac{2}{D-3}}(1+\xi)^{\frac{D-1}{D-3}}} \left\{ 4(D-2)d_3 \right. \\
 &\quad \left. - 2(1-\xi) \left[ (D-4)d_1 + (D-3)d_2 + 2(2D-5)d_3 + (D-2) \left( d_4 + \frac{1}{2}d_5 + d_6 \right) \right] \right\}.
 \end{aligned} \tag{A.11}$$

To obtain the entropy contribution from the shift in the horizon, we apply eq. (5.11). The shift in the horizon area is then

$$\Delta A = A - \tilde{A} = (D-2)\Omega_{D-2}\tilde{\rho}^{D-3}\Delta\rho = -\frac{(D-2)\Omega_{D-2}\tilde{\chi}\Delta g(\tilde{\rho})}{\partial_{\tilde{\rho}}\tilde{g}(\tilde{\rho})}, \tag{A.12}$$

where the unperturbed area is  $\tilde{A} = \Omega_{D-2}\tilde{\rho}^{D-2}$ . Inserting the perturbed metric in eqs. (A.9)

and (A.10), we then obtain

$$\begin{aligned}
 \Delta S_H = \tilde{S} \times \frac{1}{(3D-7)m^{\frac{2}{D-3}}\xi(1+\xi)^{\frac{D-1}{D-3}}} \times \\
 \times \{ d_1(1-\xi)(D-3)(D-4)[(11D-24)\xi + D-4] \\
 + d_2(1-\xi)(D-3)[(10D^2-53D+68)\xi + 2D^2-11D+16] \\
 + 2d_3[-(16D^3-128D^2+337D-292)(1-\xi)^2 \\
 + 2(3D-7)(4D^2-23D+32)(1-\xi) \\
 - 2(D-2)(D-4)(3D-7)] \\
 + 2d_4(1-\xi)(D-2)(D-3)[5(D-2)\xi + D-4] \\
 + 2(d_5+d_6)(D-2)(D-3)(1-\xi)[2(D-2)\xi + D-3] \\
 + 2(2d_7+d_8)(D-2)^2(D-3)(1-\xi)^2 \}.
 \end{aligned} \tag{A.13}$$

As before, we can consider a near-extremal limit in which  $\xi \ll 1$  but  $\Delta S \ll \tilde{S}$  so that perturbation theory still applies. This requires that

$$\xi \gg \frac{|d_i|}{m^{\frac{2}{D-3}}}, \tag{A.14}$$

which permits arbitrarily small  $\xi$  for a sufficiently large black hole. However, for the classical higher-dimension operators to dominate, we again require  $\rho \ll 1/\kappa m_\phi^{D/2}$  according to eq. (3.6). Hence, for a tree-level ultraviolet completion with  $d_i \sim 1/m_\phi^2$ , eq. (A.14) becomes  $\xi \gg \kappa^2 m_\phi^{D-2}$ . Additionally, our argument in section 2 imposes a further perturbativity criterion on the inverse temperature shift,  $\Delta\beta \ll \tilde{\beta}$ . The background inverse temperature is

$$\tilde{\beta} = \frac{2\pi m^{\frac{1}{D-3}}(1+\xi)^{\frac{D-2}{D-3}}}{(D-3)\xi}, \tag{A.15}$$

while in the near-extremal limit the inverse temperature shift goes as  $\Delta\beta \sim d_i/\xi^3 m^{1/(D-3)}$ . Hence, requiring  $\Delta\beta \ll \tilde{\beta}$ , we have

$$\xi \gg \frac{|d_i|^{1/2}}{m^{\frac{1}{D-3}}}. \tag{A.16}$$

Along with the scaling of the  $d_i$  and the bound on  $\rho$  imposed by eq. (3.6), this becomes just the requirement that

$$\xi \gg \kappa m_\phi^{(D-2)/2}, \tag{A.17}$$

so  $\xi$  can still be made parametrically small provided the heavy states are sub-Planckian.

### A.3 New positivity bounds

Combining eqs. (A.11) and (A.13), we obtain the total shift in entropy in  $D$  dimensions,

$$\begin{aligned}
 \Delta S = \tilde{S} \times & \frac{1}{(3D-7)m^{\frac{2}{D-3}}\xi(1+\xi)^{\frac{D-1}{D-3}}} \times \\
 & \times \left\{ d_1(D-3)(D-4)^2(1-\xi)^2 \right. \\
 & + d_2(D-3)(2D^2-11D+16)(1-\xi)^2 \\
 & + 2d_3[(8D^3-60D^2+151D-128)(1-\xi)^2 \\
 & \quad - 2(D-2)(2D-5)(3D-7)(1-\xi) \\
 & \quad + 2(D-2)^2(3D-7)] \\
 & + 2d_4(D-2)(D-3)(D-4)(1-\xi)^2 \\
 & + 2d_5(D-2)(D-3)^2(1-\xi)^2 \\
 & + 2d_6(D-2)(D-3)(1-\xi)[-2(2D-5)\xi + D-3] \\
 & + 4d_7(D-2)^2(D-3)(1-\xi)^2 \\
 & \left. + 2d_8(D-2)^2(D-3)(1-\xi)^2 \right\}. \tag{A.18}
 \end{aligned}$$

Positivity of this entropy shift for all  $\xi \in \left(0, \frac{D-3}{D-2}\right)$  then implies a family of new constraints on the higher-dimension operator coefficients, which generalizes eq. (6.2),

$$(1-\xi)^2 d_0 + (D-2)^2(3D-7)\xi d_3 - \frac{1}{2}(D-2)(D-3)(3D-7)\xi(1-\xi)(2d_3+d_6) > 0, \tag{A.19}$$

where in analogy with eq. (6.3) we have defined

$$\begin{aligned}
 d_0 = & \frac{1}{4}(D-3)(D-4)^2 d_1 + \frac{1}{4}(D-3)(2D^2-11D+16)d_2 \\
 & + \frac{1}{2}(2D^3-16D^2+45D-44)d_3 + \frac{1}{2}(D-2)(D-3)(D-4)d_4 \\
 & + \frac{1}{2}(D-2)(D-3)^2(d_5+d_6) + (D-2)^2(D-3)\left(d_7 + \frac{1}{2}d_8\right). \tag{A.20}
 \end{aligned}$$

As before, the bound in eq. (A.19) is stronger than any finite set of bounds obtained for fixed values of  $\xi$ , i.e., each  $\xi$  yields a linearly independent bound. As shown in appendix B, the bound in eq. (A.19) is field redefinition invariant for all values of  $\xi$ .

In the near-extremal limit,  $\xi \ll 1$ , the bound in eq. (A.19) becomes

$$d_0 > 0. \tag{A.21}$$

The above inequality is closely related to the perturbation of the extremality condition discussed in section 6.3. Applying the same reasoning to general dimension  $D$ , we find that the extremality condition for the perturbed black hole is shifted by

$$\Delta z = \frac{4(D-3)}{(3D-7)(D-2)m^{\frac{2}{D-3}}} d_0, \tag{A.22}$$

where  $d_0$  is exactly the same combination of coefficients defined in eq. (A.20). Thus, the requirement of eq. (1.8) mandating positive entropy shift implies a constraint on the coefficients of higher-dimension operators that increases the charge-to-mass ratio of extremal black holes in the theory. In turn, large black holes can decay to smaller black holes of a higher charge-to-mass-ratio, thus establishing the WGC in general dimension  $D$ .

## B Field redefinition invariance

Any physical observable should be invariant under a reparameterization of the field variables. Let us consider an arbitrary field redefinition,

$$g_{\mu\nu} \rightarrow g_{\mu\nu} + \delta g_{\mu\nu}, \quad (\text{B.1})$$

where the perturbation is second order in derivatives, so

$$\delta g_{\mu\nu} = r_1 R_{\mu\nu} + r_2 g_{\mu\nu} R + r_3 \kappa^2 F_{\mu\rho} F_{\nu}{}^{\rho} + r_4 \kappa^2 g_{\mu\nu} F_{\rho\sigma} F^{\rho\sigma} \quad (\text{B.2})$$

for a set of four arbitrary constants  $r_i$ . Inserting this field redefinition into the action for Einstein-Maxwell theory induces new terms in the action proportional to the equations of motion,<sup>8</sup>

$$\delta \mathcal{L} = \frac{1}{2\kappa^2} \delta g^{\mu\nu} \left( R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} - \kappa^2 T_{\mu\nu} \right). \quad (\text{B.3})$$

This has the net effect of shifting the higher-dimension operator coefficients in the action by

$$\begin{aligned} d_1 &\rightarrow d_1 - \frac{1}{4} r_1 - \frac{D-2}{4} r_2 \\ d_2 &\rightarrow d_2 + \frac{1}{2} r_1 \\ d_3 &\rightarrow d_3 \\ d_4 &\rightarrow d_4 + \frac{1}{8} r_1 + \frac{D-4}{8} r_2 - \frac{1}{4} r_3 - \frac{D-2}{4} r_4 \\ d_5 &\rightarrow d_5 - \frac{1}{2} r_1 + \frac{1}{2} r_3 \\ d_6 &\rightarrow d_6 \\ d_7 &\rightarrow d_7 + \frac{1}{8} r_3 + \frac{D-4}{8} r_4 \\ d_8 &\rightarrow d_8 - \frac{1}{2} r_3. \end{aligned} \quad (\text{B.4})$$

Because the field redefinition depends on four arbitrary constants, this reduces the naive basis of eight higher-dimension operator coefficients down to a set of four combinations that are automatically field redefinition invariant:

$$d_0, d_3, d_6, d_9, \quad (\text{B.5})$$

where  $d_0$  is defined in eq. (A.20) and  $d_9 = d_2 + d_5 + d_8$ . All physical quantities, like the bounds in eqs. (6.2) and (A.19), depend only on these combinations of coefficients.

<sup>8</sup>The particular field redefinition in which the pure Einstein-Maxwell equations of motion are substituted into the higher-dimension operators is a special case of the transformation in eq. (B.2).

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] S. Deser and P. van Nieuwenhuizen, *One Loop Divergences of Quantized Einstein-Maxwell Fields*, *Phys. Rev. D* **10** (1974) 401 [[INSPIRE](#)].
- [2] R.M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev. D* **48** (1993) R3427 [[gr-qc/9307038](#)] [[INSPIRE](#)].
- [3] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044 [[arXiv:1310.5713](#)] [[INSPIRE](#)].
- [4] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333 [[INSPIRE](#)].
- [5] S.W. Hawking, *Black Holes and Thermodynamics*, *Phys. Rev. D* **13** (1976) 191 [[INSPIRE](#)].
- [6] P.C.W. Davies, *Thermodynamics of Black Holes*, *Proc. Roy. Soc. Lond. A* **353** (1977) 499 [[INSPIRE](#)].
- [7] N. Arkani-Hamed, L. Motl, A. Nicolis and C. Vafa, *The string landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060 [[hep-th/0601001](#)] [[INSPIRE](#)].
- [8] C. Cheung and G.N. Remmen, *Naturalness and the Weak Gravity Conjecture*, *Phys. Rev. Lett.* **113** (2014) 051601 [[arXiv:1402.2287](#)] [[INSPIRE](#)].
- [9] C. Vafa, *The string landscape and the swampland*, [hep-th/0509212](#) [[INSPIRE](#)].
- [10] H. Ooguri and C. Vafa, *On the Geometry of the String Landscape and the Swampland*, *Nucl. Phys. B* **766** (2007) 21 [[hep-th/0605264](#)] [[INSPIRE](#)].
- [11] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis and R. Rattazzi, *Causality, analyticity and an IR obstruction to UV completion*, *JHEP* **10** (2006) 014 [[hep-th/0602178](#)] [[INSPIRE](#)].
- [12] C. Cheung and G.N. Remmen, *Infrared Consistency and the Weak Gravity Conjecture*, *JHEP* **12** (2014) 087 [[arXiv:1407.7865](#)] [[INSPIRE](#)].
- [13] B. Bellazzini, C. Cheung and G.N. Remmen, *Quantum Gravity Constraints from Unitarity and Analyticity*, *Phys. Rev. D* **93** (2016) 064076 [[arXiv:1509.00851](#)] [[INSPIRE](#)].
- [14] C. Cheung and G.N. Remmen, *Positive Signs in Massive Gravity*, *JHEP* **04** (2016) 002 [[arXiv:1601.04068](#)] [[INSPIRE](#)].
- [15] C. Cheung and G.N. Remmen, *Positivity of Curvature-Squared Corrections in Gravity*, *Phys. Rev. Lett.* **118** (2017) 051601 [[arXiv:1608.02942](#)] [[INSPIRE](#)].
- [16] G.W. Gibbons and S.W. Hawking, *Action Integrals and Partition Functions in Quantum Gravity*, *Phys. Rev. D* **15** (1977) 2752 [[INSPIRE](#)].
- [17] H.W. Braden, J.D. Brown, B.F. Whiting and J.W. York Jr., *Charged black hole in a grand canonical ensemble*, *Phys. Rev. D* **42** (1990) 3376 [[INSPIRE](#)].
- [18] J.D. Brown and J.W. York Jr., *Quasilocal energy and conserved charges derived from the gravitational action*, *Phys. Rev. D* **47** (1993) 1407 [[gr-qc/9209012](#)] [[INSPIRE](#)].
- [19] G.W. Gibbons, S.W. Hawking and M.J. Perry, *Path Integrals and the Indefiniteness of the Gravitational Action*, *Nucl. Phys. B* **138** (1978) 141 [[INSPIRE](#)].

- [20] G.W. Gibbons and M.J. Perry, *Quantizing Gravitational Instantons*, *Nucl. Phys. B* **146** (1978) 90 [INSPIRE].
- [21] D.J. Gross, M.J. Perry and L.G. Yaffe, *Instability of Flat Space at Finite Temperature*, *Phys. Rev. D* **25** (1982) 330 [INSPIRE].
- [22] J.W. York Jr., *Black hole thermodynamics and the Euclidean Einstein action*, *Phys. Rev. D* **33** (1986) 2092 [INSPIRE].
- [23] S.W. Hawking and D.N. Page, *Thermodynamics of Black Holes in anti-de Sitter Space*, *Commun. Math. Phys.* **87** (1983) 577 [INSPIRE].
- [24] T. Prestidge, *Dynamic and thermodynamic stability and negative modes in Schwarzschild-anti-de Sitter*, *Phys. Rev. D* **61** (2000) 084002 [hep-th/9907163] [INSPIRE].
- [25] H.S. Reall, *Classical and thermodynamic stability of black branes*, *Phys. Rev. D* **64** (2001) 044005 [hep-th/0104071] [INSPIRE].
- [26] R. Monteiro and J.E. Santos, *Negative modes and the thermodynamics of Reissner-Nordström black holes*, *Phys. Rev. D* **79** (2009) 064006 [arXiv:0812.1767] [INSPIRE].
- [27] A. Nicolis, R. Rattazzi and E. Trincherini, *Energy's and amplitudes' positivity*, *JHEP* **05** (2010) 095 [Erratum *ibid.* **11** (2011) 128] [arXiv:0912.4258] [INSPIRE].
- [28] Z. Komargodski and A. Schwimmer, *On Renormalization Group Flows in Four Dimensions*, *JHEP* **12** (2011) 099 [arXiv:1107.3987] [INSPIRE].
- [29] S.N. Solodukhin, *Entanglement entropy of black holes*, *Living Rev. Rel.* **14** (2011) 8 [arXiv:1104.3712] [INSPIRE].
- [30] D. Harlow, *Jerusalem Lectures on Black Holes and Quantum Information*, *Rev. Mod. Phys.* **88** (2016) 015002 [arXiv:1409.1231] [INSPIRE].
- [31] S.N. Solodukhin, *The conical singularity and quantum corrections to entropy of black hole*, *Phys. Rev. D* **51** (1995) 609 [hep-th/9407001] [INSPIRE].
- [32] S.N. Solodukhin, *On 'Nongeometric' contribution to the entropy of black hole due to quantum corrections*, *Phys. Rev. D* **51** (1995) 618 [hep-th/9408068] [INSPIRE].
- [33] D.V. Fursaev, *Temperature and entropy of a quantum black hole and conformal anomaly*, *Phys. Rev. D* **51** (1995) 5352 [hep-th/9412161] [INSPIRE].
- [34] S. Carlip, *Logarithmic corrections to black hole entropy from the Cardy formula*, *Class. Quant. Grav.* **17** (2000) 4175 [gr-qc/0005017] [INSPIRE].
- [35] S.N. Solodukhin, *One loop renormalization of black hole entropy due to nonminimally coupled matter*, *Phys. Rev. D* **52** (1995) 7046 [hep-th/9504022] [INSPIRE].
- [36] D.N. Kabat, *Black hole entropy and entropy of entanglement*, *Nucl. Phys. B* **453** (1995) 281 [hep-th/9503016] [INSPIRE].
- [37] S.N. Solodukhin, *Newton constant, contact terms and entropy*, *Phys. Rev. D* **91** (2015) 084028 [arXiv:1502.03758] [INSPIRE].
- [38] W. Donnelly and A.C. Wall, *Do gauge fields really contribute negatively to black hole entropy?*, *Phys. Rev. D* **86** (2012) 064042 [arXiv:1206.5831] [INSPIRE].
- [39] W. Donnelly and A.C. Wall, *Entanglement entropy of electromagnetic edge modes*, *Phys. Rev. Lett.* **114** (2015) 111603 [arXiv:1412.1895] [INSPIRE].

- [40] L. Susskind and J. Uglum, *Black hole entropy in canonical quantum gravity and superstring theory*, *Phys. Rev. D* **50** (1994) 2700 [[hep-th/9401070](#)] [[INSPIRE](#)].
- [41] J.-G. Demers, R. Lafrance and R.C. Myers, *Black hole entropy without brick walls*, *Phys. Rev. D* **52** (1995) 2245 [[gr-qc/9503003](#)] [[INSPIRE](#)].
- [42] Y. Kats, L. Motl and M. Padi, *Higher-order corrections to mass-charge relation of extremal black holes*, *JHEP* **12** (2007) 068 [[hep-th/0606100](#)] [[INSPIRE](#)].
- [43] M. Campanelli, C.O. Lousto and J. Audretsch, *A perturbative method to solve fourth order gravity field equations*, *Phys. Rev. D* **49** (1994) 5188 [[gr-qc/9401013](#)] [[INSPIRE](#)].
- [44] G. Goon, *Heavy Fields and Gravity*, *JHEP* **01** (2017) 045 [Erratum *ibid.* **03** (2017) 161] [[arXiv:1611.02705](#)] [[INSPIRE](#)].
- [45] T. Clunan, S.F. Ross and D.J. Smith, *On Gauss-Bonnet black hole entropy*, *Class. Quant. Grav.* **21** (2004) 3447 [[gr-qc/0402044](#)] [[INSPIRE](#)].
- [46] D.J. Gross and J.H. Sloan, *The Quartic Effective Action for the Heterotic String*, *Nucl. Phys. B* **291** (1987) 41 [[INSPIRE](#)].
- [47] L. Susskind, *Trouble for remnants*, [hep-th/9501106](#) [[INSPIRE](#)].
- [48] S.B. Giddings, *Black holes and massive remnants*, *Phys. Rev. D* **46** (1992) 1347 [[hep-th/9203059](#)] [[INSPIRE](#)].
- [49] G. 't Hooft, *Dimensional reduction in quantum gravity*, *Conf. Proc. C* **930308** (1993) 284 [[gr-qc/9310026](#)] [[INSPIRE](#)].
- [50] R. Bousso, *The holographic principle*, *Rev. Mod. Phys.* **74** (2002) 825 [[hep-th/0203101](#)] [[INSPIRE](#)].
- [51] T. Banks, M. Johnson and A. Shomer, *A Note on Gauge Theories Coupled to Gravity*, *JHEP* **09** (2006) 049 [[hep-th/0606277](#)] [[INSPIRE](#)].
- [52] G. Shiu, P. Soler and W. Cottrell, *Weak Gravity Conjecture and Extremal Black Hole*, [arXiv:1611.06270](#) [[INSPIRE](#)].
- [53] G.T. Horowitz, J.E. Santos and B. Way, *Evidence for an Electrifying Violation of Cosmic Censorship*, *Class. Quant. Grav.* **33** (2016) 195007 [[arXiv:1604.06465](#)] [[INSPIRE](#)].
- [54] Z. Fisher and C.J. Moggi, *A Semiclassical, Entropic Proof of a Weak Gravity Conjecture*, [arXiv:1706.08257](#) [[INSPIRE](#)].
- [55] S. Hod, *A proof of the weak gravity conjecture*, *Int. J. Mod. Phys. D* **26** (2017) 1742004 [[arXiv:1705.06287](#)] [[INSPIRE](#)].
- [56] B. Bellazzini, *Softness and amplitudes' positivity for spinning particles*, *JHEP* **02** (2017) 034 [[arXiv:1605.06111](#)] [[INSPIRE](#)].
- [57] C. de Rham, S. Melville, A.J. Tolley and S.-Y. Zhou, *UV complete me: Positivity Bounds for Particles with Spin*, *JHEP* **03** (2018) 011 [[arXiv:1706.02712](#)] [[INSPIRE](#)].
- [58] B. Bellazzini, F. Riva, J. Serra and F. Sgarlata, *Beyond Positivity Bounds and the Fate of Massive Gravity*, *Phys. Rev. Lett.* **120** (2018) 161101 [[arXiv:1710.02539](#)] [[INSPIRE](#)].
- [59] G. Dvali, A. Franca and C. Gomez, *Road Signs for UV-Completion*, [arXiv:1204.6388](#) [[INSPIRE](#)].
- [60] A. Jenkins and D. O'Connell, *The story of O: Positivity constraints in effective field theories*, [hep-th/0609159](#) [[INSPIRE](#)].