# On-Shell Methods and Effective Field Theory

by

Callum R. T. Jones

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in the University of Michigan
2020

Doctoral Committee:

Professor Henriette Elvang, Chair
Professor Ratindranath Akhoury
Associate Professor David Baker
Professor Finn Larsen
Assistant Professor Joshua Spitz

Callum R. T. Jones

jonescal@umich.edu

ORCID ID: 0000-0002-1325-9244

## ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge all of the incredible people that I been privileged to know over the past five years. My outstanding collaborators: Henriette Elvang, Steve Naculich, Marios Hadjiantonis, Shruti Paranjape, Brian McPeak, Jim Liu, Sera Cremonini and Laura Johnson. With such diverse expertise, together you have helped me learn about and contribute to areas of physics that alone would have seemed impossible. I hope that we can continue to learn and argue and discover new and amazing things for years to come. Henriette Elvang in particular will forever have my gratitude for agreeing to be my PhD advisor. As I took a chance on a new life, in a new country, you took a chance on me. More than just an outstanding researcher and mentor, the conscientiousness with which you approach all aspects of academic life has given me an unparalleled role model, the example of whom I will carry with me for the rest of my professional life. Moving to Ann Arbor, tens of thousands of kilometres from every person I have ever known and loved, was the most difficult thing I have ever experienced. It would have been unbearable I hadnt found the most wonderful group of friends. Michael Viray, Brian McPeak, Shruti Paranjape, Noah Steinberg, Joshua Foster, Steve Novakov, Brandon and Kate Berg, Joe and Alissa Kleinhenz, you will never know how much your friendship means to me. Most of all, I want to acknowledge Rachel Hyneman, I don’t have the words to express how your love and support has impacted my life over these years, so these will have to do, thank you for everything.

# TABLE OF CONTENTS

# LIST OF FIGURES

FIGURE

# LIST OF TABLES

TABLE

# LIST OF APPENDICES

APPENDIX

# ABSTRACT

Effective field theory methods are now widely used, in both formal and phenomenological contexts, to efficiently study universal aspects of low-energy physics. In many cases, the computational complexity associated with constructing appropriate Wilsonian effective actions and calculating observables using the traditional Feynman diagram expansion, produces a barrier to what is practically calculable. In this thesis I use a variety of modern quantum field theory approaches, including on-shell methods, to efficiently calculate physical observables in EFTs in a variety of physical contexts. Results include: 1) A systematic analysis of soft theorems for photons and gravitons incorporating the effects of generic effective operators. Consistency with spacetime locality is used to prove that the recently discovered subleading soft graviton theorem is universal in generic EFTs. 2) The development of the numerical soft bootstrap algorithm incorporating Goldstone modes with spin and linearly realized supersymmetry. 3) The use of generalized unitarity methods to calculate two infinite classes of electromagnetic duality violating one-loop amplitudes in Born-Infeld electrodynamics. It is explicitly demonstrated that the duality violation can be removed by the addition of finite local counterterms, providing strong evidence that duality is unbroken by quantization. 4) The extension of the black hole Weak Gravity Conjecture to low-energy EFTs of quantum gravity with asymptotically flat boundary conditions and arbitrary numbers of $U(1)$ gauge fields. Using on-shell methods we give a novel proof of a one-loop non-renormalization theorem in Einstein-Maxwell and use it to extend a recently given renormalization group argument for the WGC. 5) A systematic analysis of the leading higher-derivative corrections to the thermodynamic properties of charged black holes with asymptotically $AdS$ boundary conditions in arbitrary dimensions. We generalize a recent conjecture for the positivity of the leading correction to the microcanonical entropy of thermodynamically stable black holes and demonstrate that this implies the positivity of $c - a$ in a dual CFT.

# CHAPTER 1

# Introduction

## 1.1 On-Shell Methods and Feynman Diagrams

One of the most important and successful experimental avenues to investigate physics at sub-atomic scales is the study of particle scattering. Too small for real-time collection of experimental data, often the best we can do is prepare sub-atomic particles into a known initial state at some macroscopically early time, collide them together, and then measure the resulting final state at some macroscopically late time. What happens in between is not measured directly, but by repeating the experiment over-and-over a coherent picture of interactions at sub-atomic scales may emerge.

Since interactions at this scale are intrinsically quantum mechanical, formally the above collision process is described as the probabilistic transition between an initial quantum state $|i\rangle$ at time $t_i$ and a final quantum state $\langle f|$ at time $t_f$. The probability density assigned to any particular final state is given by the square of the absolute value of the *transition amplitude*

$$\mathcal{P}(i \to f) = |\langle f|U(t_f, t_i)|i\rangle|^2, \tag{1.1.1}$$

where $U(t_f, t_i)$ is the unitary time-evolution operator. The difficulty of describing real-time dynamics is overcome in practice by the approximation of macroscopically early/late times as the infinite past/future. The formal transition amplitude describing the scattering of such *asymptotic* states is the so-called *S-matrix* or *scattering amplitude*

$$\mathcal{A}(i \to f) \equiv \langle f|U(\infty, -\infty)|i\rangle. \tag{1.1.2}$$

Furthermore, due to the presence of massless elementary particles such as the photon, or the desire to describe scattering of particles with kinetic energies which are large compared to their rest mass, sub-atomic scattering is intrinsically relativistic. The dual requirements of quantum mechanics and special relativity have a unique synthesis in the framework of *relativistic quantum* *field theory*. The task of the phenomenologically inclined theoretical high-energy physicist, in the context of collider physics experiments, is then twofold: *first* to explore and classify the rich landscape of models of quantum field theory, and *second* to use the knowledge of this landscape to engineer models and calculate observables that may then be compared with real-world experimental data. By more deeply exploring the landscape of models and adjusting their constructions to agree more closely with ever more precise experimental measurements, the project of describing sub-atomic particles as a quantum field theory has grown into one of the most mature and quantitatively successful branches in all of science.

One of the great triumphs of theoretical high-energy physics, that made this progress possible in the second half of the twentieth century, was the development of a systematic perturbative expansion for calculating physical observables in quantum field theory including scattering amplitudes. Key technical innovations during this period of activity included: the development of the *Feynman diagrammatic expansion* [2], the method of *dimensional regularization*, and the understanding of a self-consistent approach to *perturbative renormalization of gauge theories* [3]. In parallel with these developments was a growing capacity for the construction of elaborate models of elementary particle physics. By 1973 [4–6], the Standard Model of Particle Physics (a relatively complicated, spontaneously broken, non-Abelian, chiral gauge theory) could be consistently formulated in the language of Lagrangian quantum field theory. At least in the perturbative regime, physically observable scattering cross-sections and decay rates could *in principle* be calculated to arbitrary accuracy[1].

Over the following decades it became clear that various technical obstacles were preventing many such calculations from being carried out *in practice*. Such problems included an enormous proliferation of Feynman diagrams and inefficient algorithms for reducing high-rank tensor integrals. This theoretical deficiency had the potential for a serious disruption of future experimental high-energy physics programs[2]. The seriousness of the problem for the calculation of background processes for the coming experimental program at the Large Hadron Collider was recognized at the program "Les Houches Physics at TeV Colliders 2005", with the informal development of a prioritized *wish list* of NLO (next-to-leading order) calculations for processes in perturbative QCD [9]. This focusing of attention and energies during the so-called *NLO revolution*, brought forward many new ideas and approaches to calculations in perturbative quantum field theory such as the use of spinor-helicity variables [10], on-shell recursion as a tool for calculating high-multiplicity tree-level parton amplitudes [11], supersymmetric decomposition of one-loop am-

[1] There is the well-known caveat to this statement, that the perturbative Dyson series is only asymptotic and will cease to give an improving approximation at some finite but very high order in the expansion, indicating the need to include non-perturbative effects [7]. At energy scales for which the gauge couplings are sufficiently small this happens at such a high order as to practically irrelevant.

[2] For a clear discussion of the problems associated with the relatively few completed NLO calculations circa 1996, a good reference is the introduction to the TASI lecture notes *Calculating scattering amplitudes efficiently* delivered by Lance Dixon [8].

plitudes in QCD [12], the widespread use of generalized unitarity methods for the construction of loop integrands [13], and the development of efficient numerical integrand reduction algorithms [14]. This phenomenological impetus, combined with purely formal developments at the same time, in particular the discovery by Witten [15], building on similar ideas of Nair [16], of a deep connection between perturbative gauge theory amplitudes and models of twistor-strings, and has developed into a vibrant and self-sustaining research subject of *scattering amplitudes*.

A primary motivation of this type of research has been the search for more efficient approaches to traditionally difficult, or even intractable, calculations in perturbative quantum field theory. *A priori* there is no obvious reason to think that the traditional methods are particularly inefficient; it would seem reasonable to assume that calculations are difficult and complicated because nature is described by quantum field theories that are difficult and complicated. If this were the case, then only marginal improvements to efficiency would be possible and real progress can only come from automation and increasingly powerful computational resources. A strong clue that the traditional approaches are genuinely inefficient is given by the apparent discrepancy between the degree of *complexity* of the explicit symbolic expressions for tree-level scattering amplitudes in non-Abelian gauge theory as calculated using Feynman diagrams and the final expressions obtained after simplification. As an illustrative example consider the color-ordered 5-point gluon scattering amplitude as calculated using planar Feynman diagrams [17]

$$\mathcal{A}_5\left[1^-,2^+,3^-,4^+,5^+\right] = \frac{N_1[1^-,2^+,3^-,4^+,5^+]}{s_{12}s_{45}} + \frac{N_2[1^-,2^+,3^-,4^+,5^+]}{s_{12}} + \mathcal{C}(1,2,3,4,5), \tag{1.1.3}$$

where $+\mathcal{C}(...)$ indicates a sum over the 4 additional cyclic permutations for a total of 10 terms, each corresponding to a single planar color-ordered Feynman diagram, and the Mandelstam invariants are defined as $s_{ij} \equiv (p_i+p_j)^2$. The numerators are local functions of the external momenta and polarization vectors, explicitly

$$\begin{aligned}
N_1[1^-,2^+,3^-,4^+,5^+] &= 2\sqrt{2}\left((\epsilon_1^-\cdot\epsilon_2^+)p_{1\mu_1} + (\epsilon_1^-\cdot p_2)\epsilon^+_{2\mu_1} - (\epsilon_2^+\cdot p_1)\epsilon^-_{1\mu_1}\right)\\
&\quad\times\left((\epsilon_3^-\cdot p_{45})g^{\mu_1\mu_2} + \epsilon_3^{-\mu_1}p_{12}^{\mu_2} + p_3^{\mu_1}\epsilon_3^{-\mu_2}\right)\\
&\quad\times\left(-(\epsilon_5^+\cdot p_4)\epsilon^+_{4\mu_2} + (\epsilon_4^+\cdot\epsilon_5^+)p_{4\mu_2} + (\epsilon_4^+\cdot p_5)\epsilon^+_{5\mu_2}\right),\\
N_2[1^-,2^+,3^-,4^+,5^+] &=\\
&-\sqrt{2}\left((\epsilon_4^+\cdot p_1)(\epsilon_1^-\cdot\epsilon_2^+) + (\epsilon_1^-\cdot p_2)(\epsilon_4^+\cdot\epsilon_2^+) - (\epsilon_2^+\cdot p_1)(\epsilon_1^-\cdot\epsilon_4^+)\right)(\epsilon_3^-\cdot\epsilon_5^+).
\end{aligned} \tag{1.1.4}$$

As first suggested by Xu, Zhang and Chang (XZC) [10], the definite helicity polarization vectors of a massless spin-1 state can be re-expressed in terms of definite helicity external wavefunctions of a massless spin-1/2 state. Defining $u_\pm(p)$ as the helicity $\pm 1/2$ solution of the massless Dirac equation in momentum space

$$\not{p}(1 \pm \gamma_5)u_\pm(p) = 0, \tag{1.1.5}$$

XZC showed that the spin-1 polarization vectors could be expressed as[3]

$$\epsilon^\mu_\pm(p;q) = \mp \frac{\bar{u}_\mp(p)\gamma^\mu u_\mp(q)}{\sqrt{2}\bar{u}_\pm(p)u_\mp(q)}, \qquad \bar{u}_\pm(p) \equiv u^\dagger_\mp(p)\gamma^0, \tag{1.1.6}$$

where $q^\mu$ is any null-vector not proportional to $p^\mu$. At first sight it might seem perverse to rewrite an amplitude for the scattering of bosons, in terms of the external wavefunctions of fermions. One of the earliest and most deceptively simple lessons learned in the modern amplitudes revolution is that all such amplitudes can be reduced, by standard gamma matrix identities, to purely rational functions of two fundamental builing blocks, the *helicity spinor brackets*[4]

$$\langle ij \rangle \equiv \bar{u}_+(p_i)u_-(p_j), \qquad [ij] \equiv \bar{u}_-(p_i)u_+(p_j). \tag{1.1.7}$$

Moreover, such rational functions can often be dramatically simplified. Indeed, the expression (1.1.3) reduces to [18]

$$\mathcal{A}_5[1^-,2^+,3^-,4^+,5^+] = \frac{\langle 13\rangle^4[12][23][34][45][51]}{s_{12}s_{23}s_{34}s_{45}s_{51}}. \tag{1.1.8}$$

This remarkably compact expression is a single instance of the more general, all-multiplicity Parke-Taylor formula for tree-level MHV gluon scattering [19]

$$\mathcal{A}_n\left[1^+,...,i^-,...,j^-,...,n^+\right] = \frac{\langle ij\rangle^4[12][23]...[n-1,n][n1]}{s_{12}s_{23}...s_{n-1,n}s_{n1}}. \tag{1.1.9}$$

For high-multiplicity, this formula represents represents the sum of an enormous number of planar Feynman diagrams [20]:

| Multiplicity ($n$) | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| Planar Feynman diagrams | 1 | 2 | 10 | 38 | 154 | 654 | 2871 | 12975 |

Certainly (1.1.9) is more compact than the Feynman diagrammatic expressions, but does this mean it is less complex?

[3]In the context of a physical scattering amplitude the dependence on $q^\mu$ drops out as one can show that $\epsilon^\mu(p,q) - \epsilon^\mu(p,q') \propto p^\mu$, the difference vanishing when contracted into an on-shell amplitude by the Ward identity. Any particular choice is equivalent to a choice of gauge fixing condition, the overall independence is then an expression of the gauge invariance of physical observables.

[4]We can also express Mandelstam invariants in terms of spinor brackets $(p_i + p_j)^2 = s_{ij} = [ij]\langle ij\rangle$. Doing so in (1.1.8) leads to the more familiar expressions for the Parke-Taylor amplitudes. However, in order to make the discussion of locality below as transparent as possible we have opted to retain the more familiar "propagator" form of the kinematic singularities.

To attempt to answer this question we will adopt the definition of complexity familiar from computer science: the complexity of a task is measured by the minimal number of computational operations needed to complete it. With regard to the two different explicit symbolic representations of the same scattering amplitude (1.1.3) and (1.1.8), we cannot apply this definition directly since an expression is not a task. Rather there are two different classes of task relevant for the present discussion: the *complexity of production* defined as the complexity of producing the expression, and the *complexity of evaluation* defined as the complexity of evaluating or transforming the expression into some other useful form. For applications to collider physics, it is usually sufficient to generate a numerical value for an amplitude which may serve as the input for numerical phase space integration in the calculation of a cross-section. In the language above, the expression (1.1.8) has a lower complexity of evaluation into a purely numerical form than (1.1.3) because of its relative compactness. Moreover the compactness of the expression suggests that the workflow described above (this was the original approach used by Mangano, Parke and Xu to calculate compact expressions for 5- and 6-point parton amplitudes [18]):

**Diagram generation** $\longrightarrow$ **Feynman rules** $\longrightarrow$ **Spinor-helicity variables** $\longrightarrow$ **Algebraic simplification**

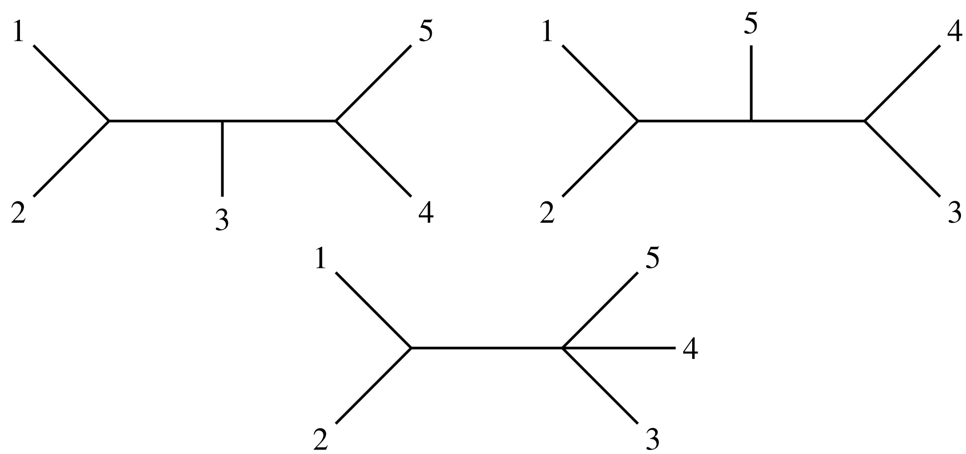
might be replaced with an alternative procedure which generates (1.1.8) *directly*, and therefore also has a smaller complexity of production. We now know that this is indeed the case, beginning with the discovery of efficient recursive (in multiplicity) algorithms for calculating amplitudes, first by Berends and Giele [21] and later by Britto, Cachazo, Feng and Witten (BCFW) [11], it has been possible to generate similarly compact expressions for tree-level gluon scattering in all helicity sectors. This *proof-of-principle* demonstrates that there exist more efficient non-Feynman-diagram based approaches to calculations, and has motivated an enormous amount of work in the field of scattering amplitudes.

What BCFW and related *on-shell methods* have in common, is that they attempt to construct the scattering amplitudes directly, using certain formal properties abstracted from Feynman diagram calculations, but without explicitly making use of a Lagrangian description. Unfortunately, the case for on-shell methods over Feynman diagrams is not so open-and-shut. We have only given one example, numerical evaluation, in which the use of these modern methods is more efficient. Particularly for *formal* high-energy theory (the subject on which this thesis is written), explicit numerical evaluation may not be a priority. Instead, we may prefer symbolic expressions, and are interested in verifying or testing certain formal algebraic properties. We can think of these tests as functions on the symbolic expressions which return *true* if the expression has property $X$ and *false* if it does not. As we will now illustrate with two examples, depending on $X$ the complexity of evaluation can be lower for the compact BCFW expression (1.1.8) in one case and lower for the Feynman diagram expression (1.1.3) in another.

Our first example is the property of *locality*, which in this context means that the location of kinematic singularities corresponds to what we expect from a model defined by a local Lagrangian density. By construction, an expression for an amplitude generated using Feynman diagrams can have kinematic singularities only at the location of propagators corresponding to the edge of a graph. We must insist that compact expressions like (1.1.8) could in principle be obtained from Feynman diagrams, and so this gives a restriction on what kind of singularities can appear both in isolation and in combination. Consider that the expression (1.1.8) has an apparent singularity at $s_{12} = 0$; there are a few things we can check to see if this could have been obtained from a Feynman diagram. First, the singularity corresponds to an invariant mass and could be generated by the Feynman gauge propagator $ig^{\mu\nu}/(p_1+p_2)^2$. Second, the singularity is a simple pole, this is important since at tree-level a given propagator can only appear once in any diagram. Finally, we consider which poles appear in combination, to make this precise we calculate the residue

$$s_{12}\mathcal{A}_5\left[1^-,2^+,3^-,4^+,5^+\right]\Big|_{s_{12}=0} = \frac{\langle 13\rangle^4[12][23][34][45][51]}{s_{23}s_{34}s_{45}s_{51}}. \tag{1.1.10}$$

This expression has an apparent pole at $s_{23} = 0$. This is *not* what we expect if the expression is calculated from Feynman diagrams. Indeed it is straightforward to enumerate all planar tree-level graphs that contribute to (1.1.3) and contain an $s_{12} = 0$ singularity:




It is clear that no diagram is possible that contains both an $s_{12} = 0$ *and* an $s_{23} = 0$ singularity. If such a combination did appear, then we would conclude that the given expression did not correspond to the tree-level scattering amplitude of a local quantum field theory. In the case at hand, since we know that (1.1.3) and (1.1.8) are the same function, the apparent illegal combination of singularities must be an illusion. Indeed a careful calculation of the *double residue* gives

$$s_{23}\left(s_{12}\mathcal{A}_5\left[1^-,2^+,3^-,4^+,5^+\right]\Big|_{s_{12}=0}\right)\Big|_{s_{23}=0} = 0. \tag{1.1.11}$$

This fact is completely obscured in the compact expression (1.1.8), but obvious in the longer expression (1.1.3). The important general lesson of this example is that Feynman diagram expressions *manifest* the expected singularity structures of a local quantum field theory, while the more compact BCFW-generated expressions do not. In the context of this particular formal property it is completely correct to say that the compact expression (1.1.8) is *more complex* than the Feynman diagram expression (1.1.3), in the sense that it requires more computer operations to verify compliance with spacetime locality.

Our second example concerns the property of the *Lorentz invariance* of the scattering amplitude. Under a general Lorentz transformation, a scattering amplitude of massless states in the helicity basis transforms as [22]

$$\mathcal{A}_n\left(\{\Lambda\cdot p_1\}^{h_1},...,\{\Lambda\cdot p_n\}^{h_n}\right)=\exp\left(i\sum_{j=1}^{n}h_j\theta(p_j,\Lambda)\right)\mathcal{A}_n\left(\{p_1\}^{h_1},...,\{p_n\}^{h_n}\right). \quad (1.1.12)$$

Where the phases $e^{ih_j\theta(p_j,\Lambda)}$ are the $1\times1$ Wigner matrices of the massless little group in $d=4$ [23], an explicit expression for this phase is not necessary for the present discussion. The fact that the amplitude transforms only by an overall phase ensures that the associated transition probability $\mathcal{P}(i\to f)\sim|\mathcal{A}(i\to f)|^2$ is independent of the frame of reference of the observer. To better understand (1.1.12), it is useful to rewrite the amplitude with explicit external wavefunctions or polarization tensors $\chi^h(p)$ contracted into an *M-function*, as we would if we were constructing it from Feynman rules

$$\mathcal{A}_n\left(\{p_1\}^{h_1},\{p_2\}^{h_2},...,\{p_n\}^{h_n}\right)=\chi^{h_1}(p_1)\cdot...\chi^{h_n}(p_n)\cdot M_n\left(p_1,...,p_n\right), \quad (1.1.13)$$

where there is an implicit contraction of spinor/vector indices between each $\chi^h$ and $M_n$. By construction, the M-function transforms covariantly as a Lorentz tensor

$$M_n\left(\Lambda\cdot p_1,...,\Lambda\cdot p_n\right)=U_{|h_1|}(\Lambda)\cdot...U_{|h_n|}(\Lambda)\cdot M_n\left(p_1,...,p_n\right), \quad (1.1.14)$$

where $U_{|h_i|}(\Lambda)$ is the matrix representation of the Lorentz transformation $\Lambda$ for a particle of spin $|h_i|$. For a particle of spin-1/2 the external wavefunctions transform as [22]

$$u_\pm(\Lambda\cdot p)=\exp\left(\pm\frac{i}{2}\theta(p,\Lambda)\right)U^{-1}_{1/2}(\Lambda)\cdot u_\pm(p). \quad (1.1.15)$$

If all of the external particles are of spin-1/2 or spin-0 (which have no external wavefunctions), then we see that (1.1.12) follows trivially. In this sense, the Feynman diagram representation of massless scattering for low-spin states is manifestly Lorentz invariant. For higher spins however the situation is more complicated; for spin-1 the polarization vectors transform under Lorentz transformations as

$$\epsilon_{\pm}^{\mu}(\Lambda \cdot p) = \exp\left(\pm i\theta(p,\Lambda)\right) \Lambda^{\mu}{}_{\nu}\epsilon_{\pm}^{\nu}(p) + \alpha(p,\Lambda)p^{\mu}, \tag{1.1.16}$$

where $\alpha(p,\Lambda)$ is a non-zero Lorentz scalar function, the precise form of which is not important for the present discussion. We see that (1.1.12) is almost true, except for the contributions from the $\alpha$-terms. The only resolution to this problem, meaning the only way to restore Lorentz invariance, is to demand that the M-functions satisfy

$$p_{i\mu}M_{n}^{\mu}(p_1, ..., p_i, ..., p_n) = 0. \tag{1.1.17}$$

This condition, otherwise known as the *on-shell Ward identity*, is equivalent to the requirement that the complete amplitude (1.1.13) is a function on an equivalence class of polarization vectors defined by the identification

$$\epsilon^{\mu}(p) \sim \epsilon^{\mu}(p) + \alpha(p)p^{\mu}. \tag{1.1.18}$$

This is nothing but the momentum space version of the familiar statement of linearized gauge invariance, that is the statement that physical observables are functions of equivalence classes of gauge potentials defined by

$$A_{\mu}(x) \sim A_{\mu}(x) + \partial_{\mu}\beta(x). \tag{1.1.19}$$

In the Feynman diagram representation (1.1.3), the property (1.1.17) is completely hidden. It is not true diagram-by-diagram and requires intricate cancellations between terms. Similar statements are true for even higher spin massless particles, the Lorentz group acts on wavefunctions as an affine transformation, and overall Lorentz invariance requires increasingly intricate Ward identity constraints on the M-functions.

Contrast this with the BCFW representation (1.1.8). Using (1.1.6) we have re-expressed the amplitude as a rational function of spin-1/2 wavefunctions which transform according to the simpler rule (1.1.15). The fundamental spinor-bracket building blocks are almost Lorentz invariant, only picking up Wigner phases

$$[ij] \xrightarrow{\Lambda} e^{\frac{i}{2}(\theta(p_i,\Lambda)+\theta(p_j,\Lambda))}[ij], \qquad \langle ij\rangle \xrightarrow{\Lambda} e^{-\frac{i}{2}(\theta(p_i,\Lambda)+\theta(p_j,\Lambda))}\langle ij\rangle. \tag{1.1.20}$$

To replicate the Lorentz transformation of the amplitude (1.1.12), our only requirement is that our rational function scales homogeneously under a so-called *little-group scaling* with an appropriate weight determined by the helicity of the state. Explicitly we require (separately) for each external particle $j$ with helicity $h_j$ that

$$\mathcal{A}_n\left(..., \{\lambda u_+(p_j), \lambda^{-1}u_-(p_j)\}, ...\right) = \lambda^{2h_i}\mathcal{A}_n\left(..., \{u_+(p_j), u_-(p_j)\}, ...\right), \tag{1.1.21}$$

| | *Model studied* | *Manifest properties* | *Non-manifest properties* |
|---|---|---|---|
| **Chapter 2** | EFTs of electrodynamics, gravity and massless higher spin | • Lorentz invariance<br>• Gauge invariance | • Locality |
| **Chapter 3** | EFTs of Goldstone particles:<br>• Nonlinear sigma models<br>• Galileons | • Lorentz invariance<br>• Gauge invariance<br>• Field redefinition invariance<br>• Linear symmetry<br>• Adler zero | • Locality<br>• Non-linear symmetry |
| **Chapter 4** | Born-Infeld electrodynamics | • Lorentz invariance<br>• Gauge invariance<br>• $d$-dimensional unitarity<br>• Electromagnetic duality (tree-level) | • Locality<br>• Electromagnetic duality (loop-level) |

Table 1.1: Summary of the models studied in Chapters 2, 3 and 4. The physical properties of the models which are manifest or guaranteed by construction in the on-shell approach used are indicated, as are the physical properties which are hidden or non-manifest and must be verified by a detailed calculation.

where $\lambda \in \mathbb{C}^*$. In practice, this condition amounts to simply counting angle and square brackets, and more importantly must hold term-by-term without any delicate cancellation. Of particular importance to the analysis presented in Chapter 2, this condition is identical (we just need a different little group weight) for all spins, and so gives a simple approach to constructing Lorentz invariant scattering amplitudes for massless higher-spin fields, by-passing the difficult problem of constructing Lagrangians with higher-spin gauge invariance. In this sense, the Lorentz invariance of the BCFW representation (1.1.8) is manifest. In the language above, the complexity of evaluating Lorentz invariance is lower in the BCFW representation (1.1.8) than the Feynman diagram representation (1.1.3).

A conclusion to be drawn from these two examples is that there is no context independent measure of complexity of the explicit symbolic representation of scattering amplitudes, and no universally less-complex representation for all purposes. This is related to a common and puzzling phenomenon in formal high-energy physics, that there is often no formalism which exists (or is possible) which simultaneously manifests all of the essential properties of a model (in any form, compact or otherwise). The best choice of formalism or representation or method of calculation, depends on the final intended purpose and often requires evaluating a delicate cost-benefit calculus to find some path forward which is the least complex *on balance*. In this thesis, in particular in Chapters 2, 3 and 4, we make use of on-shell methods to address a collection of problems of formal theoretical interest in perturbative quantum field theory. In principle, all of these problems can be addressed using the standard Feynman diagram approach. As we demonstrate repeatedly throughout this thesis, the use of on-shell methods gives an advantage on-balance, sometimes only minor and in other cases quite dramatic. An illustrative way to demonstrate this advantage, and to organize the following chapters is to given in Table 1.1 of the various models considered, together with which properties are made manifest in the on-shell formalism, and which are obscured.

In the remainder of this introduction we will give a brief introduction to the physics of effective field theories. This will include some discussion of the application of on-shell methods as well as a broader discussion incorporating the non-amplitudes-based analysis of quantum gravity EFTs relevant for Chapters 5 and 6.

## 1.2 The Landscape of Low-Energy Effective Field Theories

The models that we will study in the subsequent chapters are examples of a subclass of relativistic quantum field theories known *low-energy effective field theories* (EFTs). We will begin with a general introduction to the conceptual aspects of the *effective* way of thinking about physics, and then give a more precise field theoretic description.

The idea of an effective theory is broadly applicable across all areas of science, not just in the physics of scattering sub-atomic particles; it addresses the basic problem of how we can calculate anything without knowing everything. One starting point for this discussion is to begin with a mathematical model $M$ which makes quantitatively accurate predictions about some physical system over some domain of validity $D$. In high-energy physics a common example of such a domain is an energy range $D \equiv \{E : E < \Lambda_{\text{UV}}\}$, where $\Lambda_{\text{UV}}$ is referred to as a *UV cutoff*. In more general contexts is could be a range of length scales, speeds, temperatures or more commonly some complicated combination. It is a common occurance that observables defined in some sub-domain $D' \subset D$ only make direct reference to, and depends strongly on, a subset of the degrees-of-freedom of the model $M$. These degrees-of-freedom are *relevant* for the physics in $D'$, while the remaining degrees of freedom are *irrelevant*. Here "depends strongly on", means that the full calculation in the model $M$ of observables in $D'$ is well-approximated by ignoring the irrelevant degrees-of-freedom. The ignoring procedure defines a kind of *effective* model $M'$, defined over the domain $D'$ which contains only the relevant degrees-of-freedom and gives a good quantitative approximation to the full model. Implicitly, all models of physical systems are effective models in this sense, due to the limitations of any realistic measurement, we are always ignorant of the true set of dynamical degrees-of-freedom. This kind of leading-order effective theory reasoning has been used in physics since its earliest days; Newton knew that a good approximation for the orbital dynamics of the Earth can be made by treating the dynamics as a two-body Earth-Sun gravitational system, ignoring the internal structure of both the Earth and the Sun (treating them as point-like) and also ignoring the presence of all other celestial bodies [24].

The power of an effective theory continues when we move beyond this leading-order approximation. To be concrete we will consider the case in which the sub-domain $D'$ is defined by a small dimensionless parameter $\epsilon$, commonly a ratio of dimensionful parameters.[5] In the model $M$, observables in $D'$ can always be calculated as a perturbation expansion in $\epsilon$, with each order giving an incrementally better approximation to the full solution. It is the existence of this small parameter expansion that justifies the approximation of ignoring the irrelevent degrees-of-freedom, corresponding to the leading-order $\mathcal{O}(\epsilon^0)$ term in the series, and additionally quantifies the degree of approximation as being $\mathcal{O}(\epsilon)$. To move beyond the leading-order approximation, the model $M'$ containing only the relevant degrees of freedom must be modified in some way to reproduce the next order in the expansion. In practice this is done in the context of some general framework for constructing mathematical models of physical systems (such as classical Hamiltonian mechanics or Lagrangian quantum field theory) and the modification that must be made is to add additional non-linear interactions among the relevant degrees of freedom. Such an $\mathcal{O}(\epsilon)$ improved effective model will give predictions for observables in $D'$ with errors of $\mathcal{O}(\epsilon^2)$. This process can be continued order-by-order adding further interactions to $M'$ and *matching* with the series expansion of the exact solution of the model $M$ until an effective model is constructed which is sufficiently accurate for some purpose.

There are various reasons to prefer the use of effective models over complete ones whenever they are possible to construct. For one, they are perturbative by construction, even if the underlying model is not. An important example in the context of relativistic quantum field theory is *Chiral Perturbation Theory* ($\chi$PT), an effective model which describes the low-energy dynamics of pions [25]. The full model in this case is QCD, which is strongly coupled at low-energies and for which reliable quantitative predictions are only possible by resorting to computationally intensive lattice simulations or other non-perturbative methods . The pions, however, being the pseudo-Nambu-Goldstone modes of spontaneously broken, approximate, chiral symmetry ($SU(2)_L \times SU(2)_R \to SU(2)_V$) are the lightest hadronic resonances and the only on-shell degrees of freedom kinematically accessible in low-energy scattering. In $\chi$PT we construct an effective model for this scattering, which takes the form of an ordinary quantum field theory for an $SU(2)_V$ triplet of pseudo-scalars. This is an example of a *low-energy effective field theory*, in which the relevant small dimensionless parameter is the ratio of the scattering energy $E$ to the

[5] In the solar system example, the small dimensionless ratio that justifies the two-body approximation is the ratio of gravitational forces on the Earth from the other celestial bodies. A slightly better model is a three-body Earth-Sun-Jupiter system; the relevant ratio of gravitational forces is

$$\frac{F_{\text{Jupiter}-\text{Earth}}}{F_{\text{Sun}-\text{Earth}}} = \left(\frac{M_{\text{Jupiter}}}{M_{\text{Sun}}}\right)\left(\frac{r_{\text{Sun-Earth}}}{r_{\text{Jupiter-Earth}}}\right)^2 \approx 10^{-3}.$$

Calculations in the three-body model can be made perturbatively in this small dimensionless ratio with the leading-order term corresponding to the effective two-body approximation.

scale of chiral symmetry breaking $4\pi F_\pi \sim 1$ GeV. This means that the low-energy scattering amplitudes can be systematically approximated by an expansion of the schematic form

$$\mathcal{A}(\pi\,\pi \;\to\; \pi\,\pi) \sim a_2 \left(\frac{E}{4\pi F_\pi}\right)^2 + a_4 \left(\frac{E}{4\pi F_\pi}\right)^4 + a_6 \left(\frac{E}{4\pi F_\pi}\right)^6 + \ldots \tag{1.2.1}$$

The explicit steps for constructing a low-energy effective model of this kind, in the framework of Lagrangian quantum field theory, were first described in exactly this context by Weinberg [26]:

1. Identify all degrees-of-freedom $\Phi_i$, kinematically accessible at energies $E < \Lambda$.
2. Identify all symmetries, both linear and non-linear, and how they act on the low-energy degrees of freedom.
3. Construct the *most general* possible effective action for $\Phi_i$ that realizes the assumed symmetries,
$$S_{\text{eff}}[\Phi_i] = \int \mathrm{d}^4x \left[\mathcal{L}_{\text{kinetic}}(\Phi_i) + \sum_j \frac{c_j}{\Lambda^{\Delta_j - 4}} \mathcal{O}_j\right], \tag{1.2.2}$$
where $\Delta_j$ is the engineering dimension of the operator $\mathcal{O}_j$.
4. Use this effective action to calculate physical observables, such as low-energy scattering amplitudes (1.2.1), order-by-order in the ratio $E/\Lambda$ until the desired accuracy is attained.

Step 3 cannot be completed literally, since in general there will be an infinite number of possible operators we can write down. However, once a finite accuracy is specified, for example errors of size $(E/\Lambda)^N$ are acceptable, it becomes clear that the calculation of observables will depend only on operators with $\Delta_j \leq N - 1$; such observables will then depend on only a finite subset of *Wilson coefficients* $c_j$. For EFTs like $\chi$PT, where a more general underlying model is known, there is a further step to the construction:

5. Determine the values of the needed Wilson coefficients $c_j$ by *matching* observables calculated from the effective action (1.2.2) to the same observables calculated in the underlying model.

Since we only need finitely many coefficients once an accuracy of approximation is specified, only finitely many matchings are needed. Even if, as in the example of QCD, calculating observables in the underlying model is difficult and non-perturbative, once the matching is complete the EFT can be used to calculate *infinitely* many distinct observables within its low-energy domain of validity. In this sense the EFT is predictive, and moreover it is often an extremely efficient approach to calculating observables.

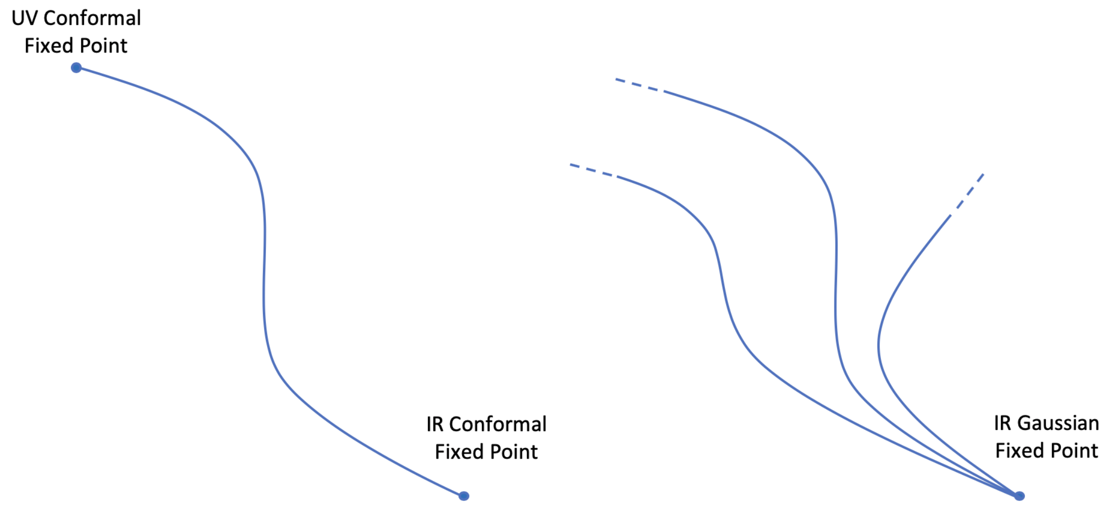



Figure 1.1: (Left): a representation of a Wilsonian UV complete quantum field theory as an RG flow from a UV CFT to an IR CFT. (Right): many models may flow to the same IR Gaussian fixed point defined by a collection of non-interacting massless degrees-of-freedom. The low-energy EFT of these massless modes captures universal aspects of this class of models.

The second, equally important, reason that EFTs are an indispensible part of modern high-energy physics, is that we can construct EFTs following steps 1-4 above, even when the underlying model is not known. In this case the cutoff scale $\Lambda$ is usually not known *a priori*, but is assumed to be sufficiently greater than the masses of all of the fields included in the effective action that the perturbative low-energy expansion is justified. In this case the virtue of the EFT construction is that it gives a completely *systematic* approach to parametrizing physics at low-energies. This is important for both the obvious phenomenological application to the parametrization of new physics, as in the so-called *Standard Model Effective Field Theory* (SMEFT) [27], but as we will see repeatedly in this thesis, also for purely formal theoretical applications.

In some special cases, a model $M$ may in principle make sensible predictions over an *unbounded* domain $D$, whether or not these are expected to accurately describe a realistic physical system. These special models are called *UV complete*. Within the context of non-gravitational QFTs there is a well-understood formal definition of UV completeness within the context of the Wilsonian renormalization group [28], illustrated in Figure 1.1. A central problem in formal high-energy theory is the classification of all UV complete models of quantum field theory and quantum gravity. While in complete generality this is an extremely difficult and ambitious problem, progress can be made by focusing on subsets of models with certain special properties. A recent example of this has been the progress made in the classification of *conformally invariant* quantum field theories via the development of the numerical conformal bootstrap and related analytic approaches [29].

Such models correspond to *interacting* fixed points of the renormalization group [28].[6] In this same spirit, in this thesis we will be repeatedly motivated by the related problem of classifying models of QFT and quantum gravity which flow to an IR Gaussian (free) fixed point. This means that we are considering models with some spectrum of massless states and some spectral gap of size $\Lambda$ which makes is possible to define a low-energy EFT containing only the massless modes and calculate observables perturbatively in the ratio $E/\Lambda$.

Such an EFT captures aspects of low-energy physics which are common to a particular *universality class* of models defined by the assumed low-energy spectrum and symmetries. Moreover, the general parametrization of the EFT by Wilson coefficients gives us a systematic approach to studying how the low-energy EFT distinguishes between different UV completions. An example that will appear repeatedly in this thesis is the phenomenon of *symmetry emergence*. It is often the case that at the leading non-trivial order in the EFT expansion, the effective model (1.2.2) has strictly more symmetry than was assumed. Such symmetries are often called *accidental* or *emergent*, as they may not be present in certain UV completions of the given universality class. The true symmetry properties of the completion are then only realized at some sub-leading order in the EFT expansion. A relevant example of this phenomenon for the analysis of Chapter 5 is the emergence of an electromagnetic duality symmetry of models with a low-energy spectrum consisting of a massless photon and a graviton. At the lowest non-trivial (two-derivative) order, the constraint of diffeomorphism invariance uniquely fixes the effective action to be that of the Einstein-Maxwell model. Such a model has a $SO(2)$ duality symmetry of its equations of motion which acts by interchanging the electric and magnetic fields. This is certainly *not* a symmetry of UV completions of this model with generic spectra of massive charged states. To see this at low-energies requires continuing the parametrization of the low-energy EFT to the second non-trivial (four-derivative) order

$$S_{\text{eff}} = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{M_{\text{Pl}}^2}{4}R - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \right. \\ \left. + \frac{\alpha}{\Lambda^4}\left(F_{\mu\nu}F^{\mu\nu}\right)^2 + \frac{\beta}{\Lambda^4}\left(F_{\mu\nu}\tilde{F}^{\mu\nu}\right)^2 + \frac{\gamma}{\Lambda^4}F_{\mu\nu}F_{\rho\sigma}W^{\mu\nu\rho\sigma} + ...\right], \tag{1.2.3}$$

where $W^{\mu\nu\rho\sigma}$ is the Weyl tensor and $M_{\text{Pl}}$ is the Planck mass. Only for the special choice of Wilson coefficients $\alpha = \beta$ and $\gamma = 0$ is the duality symmetry preserved at four-derivative order. This simple analysis immediately tells us that the physical effects of the breaking of this symmetry in the UV completion are suppressed in low-energy photon scattering by a factor of $E^2 M_{\text{Pl}}^2/\Lambda^4$. Such emergent symmetries can have dramatic consequences, in Chapter 5 we give a simple proof using on-shell unitarity methods that the existence of this emergent symmetry at two-derivative

[6] Implicit in this discussion is the notion that at an RG fixed point the emergent dilatation symmetry is enhanced to the full conformal group. In $d = 2$ this has been rigorously proven by Zamolodchikov [30] and Polchinski [31]. In $d = 4$ there are no known counterexamples, see [32] for a contemporary review.

order implies a one-loop non-renormalization theorem of the four-derivative operators, which applies to *all* models in the universality class.

The understanding and study of the landscape of all possible low-energy EFTs is clearly an important step towards the goal of a general classification of UV complete models. At first glance it is not clear that there is an interesting classification problem here. A naive, but reasonable, guess would be that the landscape of low-energy EFTs consists of all possible massless spectra, all possible combinations of symmetries and all possible values for Wilson coefficients. As we will see throughout this thesis, this is not the case for two distinct reasons. First, not all combinations of spectra and symmetries are self-consistent, or consistent with general field theory properties such as locality, unitarity and Lorentz invariance. Additionally, consistency with any *non-linear* symmetries imposes an infinite number of constraints relating the Wilson coefficients. Since the pathologies associated with violating these constraints are visible in the low-energy EFT we refer to them as *bottom-up constraints*. In many cases, as we will see in Chapters 2 and 3, these inconsistent combinations are not always obviously problematic, often requiring a detailed analysis to expose the underlying sickness. Second, even a completely healthy looking low-energy EFT may fail to have a UV completion. Constraints of this kind require some general understanding of the UV physics and are quite poorly understood; we refer to these as *top-down* or *Swampland constraints* [33]. In the following subsections we give examples of both of these types of constraints relevant for later chapters.

### 1.2.1 Bottom-up Constraints: Low-Energy Theorems

In Chapters 3 and 4 an important role will be played by low-energy EFTs describing the dynamics of the Goldstone modes of some pattern of spontaneous symmetry breaking. The associated effective actions are strongly constrained by the requirement that the spontaneously broken symmetry is *nonlinearly realized* on the massless fields. That is, we must impose the additional constraint that the action is invariant under symmetries of the form

$$\phi \rightarrow \phi + a_0 + a_1\phi^2 + a_2\phi^3 + \dots \tag{1.2.4}$$

Such symmetries relate the *a priori* independent Wilson coefficients of the infinitely many operators with different multiplicities of the fundamental fields which may appear at a fixed order in the derivative expansion. For the simple case of the spontaneous breaking of internal symmetries, a completely systematic approach to constructing effective actions which automatically incorporate these constraints was developed by Callan, Coleman, Wess and Zumino (CCWZ) [34, 35]. Beyond the simplest case - incorporating arbitrarily complicated combinations of spacetime and super-symmetry breaking with additional unbroken (linearly realized) symmetries and couplings to generic matter fields - some partial extensions of the CCWZ formalism are know. However, they are neither exhaustive nor, due to the necessity of imposing certain *inverse Higgs* constraints, particularly simple to work with.[7] In Chapter 3 we develop a simpler approach, based on scattering amplitudes, to implementing these complicated bottom-up constraints known as the *soft bootstrap*. Here we will give a brief introduction to these ideas.

The starting point of our approach is the understanding that spontaneously broken symmetries manifest themselves in scattering amplitudes as low-energy or soft theorems. The relations among the Wilson coefficients are required to produce a cancellation between Feynman diagrams that in turn is responsible for the low-energy theorem. It is instructive to frame this idea in the context of an explicit example. We will consider the low-energy dynamics of a non-compact 3-brane embedded in 5d Minkowski space. While the complete physical system is invariant under the group $ISO(4,1)$ of 5d Poincare transformations, the ground state corresponds to a static and flat configuration of the brane, which is invariant only under the subgroup $ISO(3,1) \times \mathbb{Z}_2$. The spontaneous breaking of the spacetime symmetry is associated, via the Goldstone theorem, with a massless scalar boson which is identified with the modulus of the brane $\phi$. It is well-known that the leading low-energy dynamics is governed by the Dirac-Born-Infeld (DBI) action. In static gauge, it takes the form

$$S_{\text{DBI}} = \Lambda^4 \int d^4x \left( \sqrt{-\det\left(\eta_{\mu\nu} + \tfrac{1}{\Lambda^4}\partial_\mu\phi\partial_\nu\phi\right)} - 1 \right), \tag{1.2.5}$$

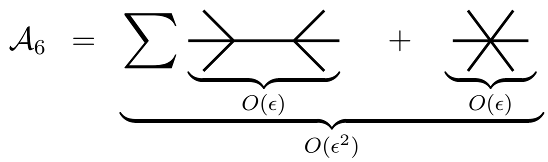
where the cutoff scale $\Lambda^4$ has the physical interpretation of measuring the tension of the 3-brane. This action trivially has a constant shift symmetry $\phi \to \phi + c$, the on-shell consequence of which is that the DBI amplitudes have *vanishing single-soft limits*. We can see this by means of the following simple argument: when one of its momentum lines is taken soft,

$$p^\mu_{\text{soft}} \to \epsilon\, p^\mu_{\text{soft}} \quad \text{with} \quad \epsilon \to 0\,, \tag{1.2.6}$$

the Feynman vertex it sits on goes to zero as $\mathcal{O}(\epsilon)$. There are no cubic interactions, so propagators remain finite. Hence, every tree-level Feynman diagram goes to zero as $\mathcal{O}(\epsilon)$. The general phenomenon of vanishing soft limits for Goldstone modes is well-known in the phenomenological literature of pion physics and is often referred to as an *Adler zero* [37]. What is less obvious is that a cancellation occurs between Feynman diagrams such that the soft behavior of any tree-level DBI $n$-point amplitude is enhanced to $\mathcal{O}(\epsilon^2)$. For example for the 6-point amplitude, the $\mathcal{O}(\epsilon)$-contributions of the pole diagrams cancel against those of the 6-point contact term, leaving an

[7] This does not mean that it is impossible, see [36] for an example of a non-linear realization based approach to the same classification problem we consider in Chapter 3 with several overlapping results.

overall $\mathcal{O}(\epsilon^2)$ soft behavior:




(1.2.7)

The *enhanced* soft limit of the DBI amplitudes is a consequence of the non-manifest 5d Lorentz symmetry of the action. In the static gauge action (1.2.5) this corresponds explicitly to an enhanced shift symmetry on the brane action of the form

$$\phi \to \phi + v^\mu x_\mu - \frac{1}{\Lambda^4} v^\mu \phi \partial_\mu \phi. \tag{1.2.8}$$

Importantly, we can invert the logic of this calculation, and by doing so *rediscover* the DBI action (1.2.5). A sketch of the inverted argument goes as follows: we begin with the most general effective action with the same EFT power-counting as DBI, but now with *a priori* independent Wilson coefficients $c_i$ (the ubiquitous field redefinition redundancy of the effective action is assumed to have also been completely fixed)

$$S_{\text{eff}} = \int \mathrm{d}^4 x \ \left[ (\partial\phi)^2 + \frac{c_1}{\Lambda^4} \partial^4 \phi^4 + \frac{c_2}{\Lambda^8} \partial^6 \phi^6 + ... \right]. \tag{1.2.9}$$

Imposing that the amplitudes of this model satisfy $\mathcal{O}(\epsilon^2)$ low-energy theorems generates an infinite set of relations among the $c_i$. We claim (though it is not at all obvious) that there is a unique solution to these constraints and, up to a non-physical field redefinition, this solution is precisely the DBI model. Thus we establish that DBI is the unique, leading-order, real single-scalar theory with $\mathcal{O}(\epsilon^2)$ low-energy theorems [38].

Something conceptually interesting has happened in this inverted line of reasoning. In the traditional approach, we first describe the symmetry breaking pattern and then construct the most general effective action which non-linearly realizes the symmetry. The phenomenon of the Adler zero, including the enhancement to quadratic order, is taken as a consequence of the non-linear symmetry. In the inverted approach, the degree of the Adler zero is taken as part of the data that defines the EFT together with the on-shell spectrum and the linearly realized symmetries. If a model which realizes this data can be constructed, then only *post hoc* is the vanishing soft limit interpreted in terms of a pattern of spontaneous symmetry breaking.

We will use this inverted approach in Chapter 3 to study the landscape of low-energy EFTs with non-linear symmetries, by-passing the complicated construction of the effective action entirely. Instead we will develop a systematic approach to directly generating tree-level scattering amplitudes which automatically satisfy the assumed linear symmetries and low-energy theorems. In some cases the object we construct will not be consistent with the inviolable constraints of locality and unitarity, and so we will be forced to conclude that our assumed combination of properties is inconsistent, and no such low-energy EFT can exist. In the context of the broader project of giving a complete classification of quantum field theories, the no-go theorems we derive using this on-shell approach have the physical interpretation that certain, naively sensible, patterns of spontaneous symmetry breaking can never actually be realized in a UV complete model.

### 1.2.2 Top-down Constraints: Weak Gravity Conjecture

The Weak Gravity Conjecture (WGC) is perhaps the oldest and best-studied example of a proposed *Swampland* criterion [33]. In general a Swampland conjecture proposes the following: all low-energy EFTs of quantum gravity which can *actually* be realized by matching to a UV completion have property X, therefore, if a low-energy EFT does not have property X it cannot have a UV completion and belongs to the swampland rather than the landscape of fundamentally meaningful low-energy EFTs. Since, in contrast with the well-understood Wilsonian picture of UV completion in non-gravitational QFT, we do *not* have a good understanding of the landscape of UV complete models of quantum gravity, such proposed criteria will likely remain as tantalizing conjectures for the forseeable future.

In its original form the WGC postulates that [39]: *in a UV complete model of quantum gravity, there should not exist an infinite tower of exactly stable states in a fixed direction in charge space.* Arguments against such an infinite tower include that it might lead to a species problem or remnant issues [40, 41]. No proof of this statement has been given (hence why it remains a conjecture), but it is consistent with all known explicit examples of string compactifications and is conceptually consistent with a number of other conjectures about quantum gravity, such as the finiteness principle and the absence of global symmetries [42].

The conjecture can be equivalently interpreted as a statement about the (in-)stability of asymptotically large extremal black holes. In quantum gravity, *elementary* states with super-Planckian masses can be expected to appear to distant observers as black hole solutions of some low-energy effective field theory (EFT) [43, 44]. The decay of such a state must have an equivalent semi-classical description as the discharge of the black hole, for example by Schwinger pair production of charged states near the horizon [45]. Since the relevant energy scale $\mu$ for the EFT calculation is here given by the scale of the black hole horizon $\mu \sim M_{\text{Pl}}^2/M$, asymptotically large black holes are well approximated by standard two-derivative Einstein gravity together with any additional massless degrees of freedom. All other details of the UV physics are integrated out and appear in the low-energy EFT as contributions to Wilson coefficients of higher-derivative effective operators that give subleading corrections to the black hole solutions. Models of quantum gravity can then be organized into universality classes according to their massless spectra and lowest dimension interactions; each class of model has an associated set of large black hole solutions that must then correspond to the asymptotic spectrum of super-Planckian elementary states.

In the analysis presented in Chapter 5 we consider the universality class of models in four-dimensions with zero cosmological constant and a massless spectrum of matter fields consisting of $N$ $U(1)$ gauge fields, while in Chapter 6 we consider models with a *negative* cosmological constant and a single $U(1)$ gauge field. For clarity, in these introductory remarks we review the statement of the WGC for a single $U(1)$ in flat space; in this class the spectrum of large black holes corresponds to the familiar Kerr-Newman solutions. Within a given charge sector, the lightest black hole corresponds to the extremal, non-rotating solution with $Q^2 = M^2/M_{\text{Pl}}^2$. If the WGC is true, then for all $Q^2$ greater than some critical value, the corresponding extremal black hole must be able to discharge. Whether this is kinematically possible depends on the spectrum of charged states with masses lighter than the black hole. For a general transition of the form

$$|Q, M\rangle \rightarrow |q_1, m_1\rangle \otimes |q_2, m_2\rangle \otimes ... \otimes |q_n, m_n\rangle, \tag{1.2.10}$$

where each of the final states is assumed to be localized and at rest asymptotically far away (with zero kinetic and gravitational potential energy), conservation of total energy and total charge requires

$$Q = q_1 + q_2 + ... + q_n, \qquad M = m_1 + m_2 + ... + m_n. \tag{1.2.11}$$

If the initial state is a large extremal black hole with $Q^2 = M^2/M_{\text{Pl}}^2$, then at least one of the daughter states $|q_i, m_i\rangle$ must be *self-repulsive*, meaning $q_i^2 \geq m_i^2/M_{\text{Pl}}^2$. Conversely, if there are no self-repulsive states then such a decay is impossible and an infinite tower of extremal black holes are exactly stable, violating the aforementioned Swampland criterion. This leads to the common formulation of the WGC:

**Weak Gravity Conjecture (Single Charge):** *In a UV complete model of quantum gravity there must exist some state with $Q^2 \geq M^2/M_{\text{Pl}}^2$.*

In the context of a specific model, to show that the WGC is violated requires complete knowledge of the spectrum of charged states. To show that it is satisfied however, requires only the existence of a single self-repulsive state. In our universe, for example, this condition is clearly satisfied by the existence of a particle such as the electron with

$$\frac{M_e^2}{M_{\text{Pl}}^2 Q_e^2} = \frac{4\pi\epsilon_0 G_N m_e^2}{q_e^2} \approx 2.3 \times 10^{-43}. \tag{1.2.12}$$

In general, it is useful to separate charged states into three regimes according to their masses:

1. **Particle regime** ($M \ll M_{\rm Pl}$): States in this regime are well-described by ordinary quantum field theory on a fixed spacetime background.

2. **Stringy regime** ($M \lesssim M_{\rm Pl}$): States in this regime are intrinsically related to the UV completion. They can usually only be calculated from a detailed understanding of the UV physics such as an explicit string compactification.

3. **Black hole regime** ($M \gg M_{\rm Pl}$): States in this regime are well-described by semi-classical black hole solutions in the relevant low-energy effective model of gravity.

In Chapters 5 and 6 we are analyzing the spectrum of charged states in the black hole regime. The corresponding analysis for a single $U(1)$ gauge field was made in [46], which we will now briefly review. Naively, it would seem impossible for a charged black hole to be self-repulsive since this would violate the extremality bound. The usual bound $Q^2 \leq M^2/M_{\rm Pl}^2$ is derived by requiring the existence of a horizon (by requiring Weak Cosmic Censorship). When the higher derivative corrections to the effective action are included, the black hole solutions and the associated extremality bounds are modified. For large black holes, with $Q^2 \gg 1$, these corrections can be calculated perturbatively in $1/Q^2$, with the leading corrections corresponding to four-derivative effective operators. The authors of [46] analyzed electrically charged solutions to the following effective action (1.2.3); at leading-order, the corrected extremality bound is

$$\frac{M_{\rm Pl}^2 Q^2}{M^2} \leq 1 + \frac{4M_{\rm Pl}^4}{5Q^2\Lambda^2}(2\alpha - \gamma) + \mathcal{O}\left(\frac{1}{Q^4}\right). \tag{1.2.13}$$

The $\mathcal{O}\left(1/Q^4\right)$ contributions correspond to next-to-leading-order in the four-derivative operators and leading-order in six-derivative operators. If the corrected extremality bound is positive

$$2\alpha - \gamma > 0, \tag{1.2.14}$$

then extremal black holes with finite charge are self-repulsive and the WGC is satisfied in the black hole regime. Conversely, if the corrected extremality bound is negative

$$2\alpha - \gamma < 0, \tag{1.2.15}$$

then the decay of asymptotically large extremal black holes into extremal black holes with large but finite charge is kinematically impossible. This does not mean that the WGC is violated, but rather that if it is valid then there must exist a self-repulsive state in either the stringy or particle regimes.

Various arguments have been given that (1.2.14) should always be true, including arguments from unitarity, causality [47], positivity of the S-matrix [48], shifts to entropy bounds [49], and renormalization group running [50]. Each of these arguments makes various assumptions and approximations, and the question of whether this formulation of the WGC (or one of the many stronger versions [33]) gives a *true* and *useful* constraint on low-energy EFTs of quantum gravity remains open.

## 1.3 Overview of this Thesis

In *Chapter 2* we use spinor-helicity methods to make a systematic study of soft limits for photons and gravitons in generic low-energy EFTs. The scope of this analysis includes all EFTs of massless states in $d = 4$ at tree-level. We prove that the subleading soft-graviton theorem is universal, while the subleading soft photon and sub-subleading soft graviton theorems are quasi-universal, receiving corrections from a set of effective operators for which we give a complete classification. In addition, from constraints arising by demanding that our master formula is consistent with spacetime locality, we reproduce known low-energy constraints on massless quantum field theories including charge conservation, the Einstein equivalence principle and various no-go results for massless higher-spin fields.

Based on:

- *Soft Photon and Graviton Theorems in Effective Field Theory* with Henriette Elvang and Stephen G. Naculich; Phys.Rev.Lett. 118 (2017) no.23, 231601; **arXiv:1611.07534** [51].

In *Chapter 3* we extend the soft bootstrap method for constraining the landscape of low-energy EFTs of Goldstone particles to incorporate spinning states. The implementation of the soft bootstrap uses the recently discovered method of soft subtracted recursion [76]. We derive a precise criterion for the validity of these recursion relations and show that they fail exactly when the assumed symmetries can be trivially realized by independent operators in the effective action. We use this to show that the possible pure (real and complex) scalar, fermion, and vector exceptional EFTs are highly constrained. Next, we prove how the soft behavior of states in a supermultiplet must be related and illustrate the results in extended supergravity. We demonstrate the power of the soft bootstrap in two applications. First, for the $\mathcal{N} = 1$ and $\mathcal{N} = 2$ $\mathbb{CP}^1$ nonlinear sigma models, we show that on-shell constructibility establishes the emergence of accidental IR symmetries. This includes a new on-shell perspective on the interplay between $\mathcal{N} = 2$ supersymmetry, low-energy theorems, and electromagnetic duality. We also show that $\mathcal{N} = 2$ supersymmetry requires 3-point interactions with the photon that make the soft behavior of the scalar $\mathcal{O}(1)$ instead of vanishing, despite the underlying symmetric coset. Second, we study Galileon theories, including aspects of supersymmetrization, the possibility of a vector-scalar Galileon EFT, and the existence of higher-derivative corrections preserving the enhanced special Galileon symmetry.

This is addressed by soft bootstrap and by application of double-copy/KLT relations applied to higher-derivative corrections of chiral perturbation theory.

Based on:

- *On the Supersymmetrization of Galileon Theories in Four Dimensions* with Henriette Elvang, Marios Hadjiantonis and Shruti Paranjape; Phys.Lett. B781 (2018) 656-663; **arXiv:1712.09937** [52].

- *Soft Bootstrap and Supersymmetry* with Henriette Elvang, Marios Hadjiantonis and Shruti Paranjape; JHEP 1901 (2019) 195; **arXiv:1806.06079** [53].

In *Chapter 4* we initiate a study of non-supersymmetric Born-Infeld electrodynamics in 4d at the quantum level. Explicit all-multiplicity expressions are calculated for the purely rational one-loop amplitudes in the self-dual $(+ + \ldots +)$ and next-to-self-dual $(- + \ldots +)$ helicity sectors. Using a supersymmetric decomposition, $d$-dimensional unitarity cuts of the integrand factorize into tree-amplitudes in a 4d model of Born-Infeld photons coupled to a massive complex scalar. The two-scalar tree-amplitudes needed to construct the Born-Infeld integrand are computed using two complimentary approaches: (1) as a double-copy of Yang-Mills coupled to a massive adjoint scalar with a dimensionally reduced form of Chiral Perturbation Theory, and (2) by imposing consistency with low-energy theorems under a reduction from 4d to 3d and T-duality. The Born-Infeld integrand is integrated in $d = 4-2\epsilon$ dimensions at order $\mathcal{O}(\epsilon^0)$ using the dimension-shifting formalism. We comment on the implications for electromagnetic duality in Born-Infeld theory and interpret our explicit results as evidence for the validity of duality at the quantum level.

Based on:

- *All-Multiplicity One-Loop Amplitudes in Born-Infeld Electrodynamics from Generalized Unitarity* with Henriette Elvang, Marios Hadjiantonis and Shruti Paranjape; JHEP 2003 (2020) 009; **arXiv:1906.05321** [54].

In *Chapter 5* we study the effect of higher-derivative corrections on asymptotically flat, four-dimensional, dyonic black holes in low-energy models of gravity coupled to $N\,U(1)$ gauge fields. For large extremal black holes, the leading $\mathcal{O}\left(1/Q^2\right)$ correction to the extremality bound is calculated from the most general low-energy effective action containing operators with up to four derivatives. Motivated by the multi-charge generalization of the Weak Gravity Conjecture, we analyze the necessary kinematic conditions for an asymptotically large extremal black hole to decay into a multi-particle state of extremal black holes. In the large black hole regime, we show that the convex hull condition degenerates to the requirement that a certain quartic form constructed from the Wilson coefficients of the four-derivative effective operators, is everywhere positive.

Using on-shell unitarity methods, we show that higher-derivative operators are renormalized at one-loop only if they generate local, on-shell matrix elements that are invariant tensors of the electromagnetic duality group $U(N)$. The one-loop logarithmic running of the four-derivative Wilson coefficients is calculated and shown to imply the positivity of the extremality form at some finite value of $Q^2$. This result generalizes an argument recently given by Charles [50], and shows that under the given assumptions the multi-charge Weak Gravity Conjecture is not a Swampland criterion.

Based on:

- *The Black Hole Weak Gravity Conjecture with Multiple Charges* with Brian McPeak; under review at JHEP; **arXiv:1908.10452** [55].

In *Chapter 6* we compute the four-derivative corrections to the geometry, extremality bound, and thermodynamic quantities of AdS-Reissner-Nordström black holes for general dimensions and horizon geometries. We confirm the universal relationship between the extremality shift at fixed charge and the shift of the microcanonical entropy, and discuss the consequences of this relation for the Weak Gravity Conjecture in AdS. The thermodynamic corrections are calculated using two different methods: first by explicitly solving the higher-derivative equations of motion and second, by evaluating the higher-derivative Euclidean on-shell action on the leading-order solution. In both cases we find agreement, up to the addition of a Casimir energy in odd dimensions. We derive the bounds on the four-derivative Wilson coefficients implied by the conjectured positivity of the leading corrections to the microcanonical entropy of thermodynamically stable black holes. These include the requirement that the coefficient of Riemann-squared is positive, meaning that the positivity of the entropy shift is related to the condition that $c - a$ is positive in the dual CFT. We discuss implications for the deviation of $\eta/s$ from its universal value and a potential lower bound.

Based on:

- *Higher-Derivative Corrections to Entropy and the Weak Gravity Conjecture in Anti-de Sitter Space* with Sera Cremonini, James T. Liu and Brian McPeak; **arXiv:1912.11161** [56].

# CHAPTER 2

# Singular Soft Limits in Gauge Theory and Gravity

There has been a recent resurgence of interest in the classic subject of studying the universal properties of scattering amplitudes of photons and gravitons, in the so-called *soft limit* in which the momentum of one of the external states vanishes [57–63]. Much of this recent activity has been driven by the discovery of a deep connection between soft limits and asymptotic symmetries in flat-space [64–69]. Despite enormous progress, much of the recent work has been confined to calculations at tree-level and at leading (two-derivative) order in the EFT expansion. Moving beyond these approximations is necessary to determine if the newly discovered relationships and structures have any fundamental physical significance or are merely an artifact of the approximations being made.

We can illustrate this point more clearly in the context of a concrete example. In [70] it had been explicitly demonstrated for tree-level graviton scattering, that the Ward identities of the asymptotic BMS group are equivalent to the leading-order Weinberg soft graviton theorem [60]. Equivalent here means that assuming one the other can be derived as a consequence. It had also recently been shown [64], using the BCFW construction of the tree-level S-matrix for Einstein gravity, that the universal factorized form of the Weinberg soft theorem continued to sub-subleading order in the soft expansion. A natural question is whether a similar correspondence with asymptotic symmetries would hold for the subleading corrections. Progress on this question was made in [65], it was demonstrated that if one assumed the newly discovered subleading soft theorem, and followed the same formal steps as were taken at leading order, then the on-shell matrix elements would satisfy the Ward identity of a hitherto unknown Virasoro symmetry. The interpretation was given that the true asymptotic symmetry group of quantum gravity in flat space was an enlargement of the BMS group which incorporated this new Virasoro symmetry. This bold conjecture rested on the fundamental status of the newly discovered subleading soft theorem, which had only been demonstrated at tree-level in two-derivative Einstein gravity. If it was truly a fundamental feature of quantum gravity then it should be robust against both quantum and higher-derivative corrections.

The purpose of this chapter is to make use of the modern technical tools developed in the context of the on-shell scattering amplitudes program to make a systematic analysis of the sub- and sub-subleading soft theorems in effective field theories of gauge theory and gravity. The scope of this analysis is to cover all massless quantum field theories in $d = 4$ at tree-level, importantly including all higher dimension effective operators. Such operators are natural in a low-energy EFT expansion of quantum gravity and can be thought of as parametrizing the effects of both $\alpha'$ corrections in string theory and the loop-level effects of integrating out massive states. In the context of the above discussion, one of our key results will be the demonstration that the subleading soft graviton theorem discovered in [64] is *unmodified by higher-derivative operators*. We will also demonstrate that the sub-subleading soft graviton and subleading soft photon theorems are *quasi*-universal, and are modified by a finite collection of operators for which we give a complete classification.

## 2.1 Systematics of the Soft Expansion

In this chapter we will be concerned with the structure of scattering amplitudes as a power series expansion around the soft limit at *subleading* orders. This requires a certain amount of care to ensure that the expansion is in harmony with the constraints of momentum conservation and on-shell-ness. The simplest approach to this is to construct an explicit function $p_k^\mu : \mathbb{C} \to \mathcal{K}_{n+1}$, where the kinematic space for scattering $n + 1$ massless particles is defined as

$$\mathcal{K}_{n+1} \equiv \{p_k^\mu \in \mathbb{C}^{4n+4} : p_k^2 = 0, \ \sum_{k=1}^{n+1} p_k^\mu = 0\}. \tag{2.1.1}$$

For such a function to probe a soft limit we further require $p_{n+1}^\mu(0) = 0$ (subsequently we will use the label $s$ in place of $n + 1$). Since we are assuming all momenta are massless, we can trivialize the on-shell condition by using spinor-helicity variables (our conventions are the same as [71]). The function we will use is the following

$$\begin{aligned} |\hat{s}\rangle &= \epsilon|s\rangle, \\ |\hat{i}] &= |i] - \epsilon\frac{\langle js\rangle}{\langle ji\rangle}|s], \\ |\hat{j}] &= |j] - \epsilon\frac{\langle is\rangle}{\langle ij\rangle}|s], \end{aligned} \tag{2.1.2}$$

where the unhatted spinors satisfy $n$-particle momentum conservation

$$\sum_{k=1}^{n} |k\rangle[k| = 0, \tag{2.1.3}$$

the spinors $|s\rangle$ and $|s]$ are independent, and all other spinors do not depend on $\epsilon$. Here the choice of two external momenta $p_i$ and $p_j$ with $i, j \in \{1, ..., n\}$ is necessary to satisfy momentum conservation; this choice only appears beyond leading order in the soft expansion and is a reflection of the fact that our stated conditions greatly underdetermine such a function. It is straightforward to verify that this definition provides a map into $\mathcal{K}_{n+1}$ with $p_s^\mu(0) = 0$.

The *soft expansion* is then defined by evaluating an $n+1$-point scattering amplitude $\mathcal{A}_{n+1} : \mathcal{K}_{n+1} \to \mathbb{C}$, on this function giving $\hat{\mathcal{A}}_{n+1} : \mathbb{C} \to \mathbb{C}$ and taking a Taylor (or Laurent) expansion around $\epsilon = 0$

$$\hat{\mathcal{A}}_{n+1}(\epsilon) = \mathcal{S}_{n+1}^{(0)} \epsilon^\sigma + \mathcal{S}_{n+1}^{(1)} \epsilon^{\sigma+1} + ... \tag{2.1.4}$$

where $\mathcal{S}_{n+1}^{(0)} \neq 0$ by assumption. This defines the *holomorphic soft weight* $\sigma$ as the leading non-zero power in the soft expansion. In (2.1.2) we have made the choice that the angle spinor $|s\rangle$ vanishes as $\epsilon \to 0$, while the square spinor $|s]$ does not, this defines the so-called *holomorphic soft limit*. We could have made the opposite choice and taken the square spinor to vanish, defining the *anti-holomorphic soft limit* or taken both to vanish together, only this final choice is compatible with real valued momenta. The value of $\sigma$ is dependent on this choice, but the functions $\mathcal{S}^{(0)}$ are not. Converting from holomorphic to anti-holomorphic soft limits is a simple little-group rescaling in which the amplitude acquires an overall factor

$$\mathcal{A}_{n+1}\left(\epsilon|s\rangle, |s], ...\right) \sim \epsilon^\sigma \iff \mathcal{A}_{n+1}\left(|s\rangle, \epsilon|s], ...\right) \sim \epsilon^{\sigma+2h_s}, \tag{2.1.5}$$

where $h_s$ is the helicity of particle $s$. For $h_s > 0$, the use of the holomorphic soft limit *minimizes* the value of $\sigma$, while the anti-holomorphic soft limit *maximizes* it. The significance of this choice is that for the holomorphic soft limit of a positive helicity state, the greatest number of orders in the soft expansion have negative powers of $\epsilon$; and as we will see, it is these singular orders that exhibit a universal or quasi-universal behaviour.

## 2.2 Master Equation for Singular Soft Limits

For $\sigma < 0$, the soft limit described in the previous subsection corresponds to a kinematic singularity. The structure of all such singularities is completely determined by locality and unitarity, our first job will be to make this statement precise. The function (2.1.2) is a particular example of a *linear momentum deformation*, which in general takes the form

$$\hat{p}_i^\mu(z) = p_i^\mu + z q_i^\mu. \tag{2.2.1}$$

For this to define a function onto $\mathcal{K}_n$, the vectors $q_i^\mu$ must satisfy the conditions

$$q_i^2 = 0, \quad p_i \cdot q_i = 0, \quad \sum_{i=1}^{n} q_i^\mu = 0. \tag{2.2.2}$$

We will describe such a deformation as *regular* if satisfies a further condition. Consider the set of all equations of the form

$$(\hat{p}_{i_1}(z) + \hat{p}_{i_2}(z) + ... + \hat{p}_{i_k}(z))^2 = 0, \quad \text{where} \quad \{i_1, i_2, ..., i_k\} \subset \{1, 2, ..., n\}. \tag{2.2.3}$$

It is straightforward to see that each such equation is quadratic in $z$ and therefore has two solutions $z^\pm_{i_1 i_2 \ldots i_k}$; the deformation is defined to be regular if no two solutions (on any of the equations) coincide. The importance of this definition is that it gives us a precise way to describe the constraints of *locality* and *unitarity* in complexified momentum space. In the following we define $\hat{\mathcal{A}}_n(z)$ as a tree-level scattering amplitude evaluated on a regular momentum deformation:

- **Locality:** An amplitude $\hat{\mathcal{A}}_n(z)$ is a meromorphic function with isolated singularities located only at solutions to equations (2.2.3). Moreover, such singularities are at most simple poles.

- **(Tree-level) Unitarity:** The residue of such a simple pole is completely determined by factorization into lower-multiplicity scattering amplitudes

$$P_I^2(z) \hat{\mathcal{A}}_n(z)\Big|_{z=z_I} = \sum_{\psi} \mathcal{A}_L(..., (-P_I)_\psi) \mathcal{A}_R((P_I)_\psi, ...), \tag{2.2.4}$$

  where $\psi$ denotes the set of all on-shell states exchanged in the channel $P_I^2 = 0$.

Naively, the combination of these properties should completely determine the structure of the singular terms in the soft expansion. Unfortunately, the holomorphic soft limit (2.1.2) is not a regular momentum deformation. It is straightforward to see this by considering two-particle factorization channels of the form

$$(\hat{p}_s(\epsilon) + p_k)^2 = \epsilon \langle sk \rangle [sk], \tag{2.2.5}$$

we see that for all $k \neq s$, this equation has a zero at $\epsilon = 0$. The structure of the singularity at this point, which is exactly the soft limit $p_s \to 0$, is not immediately given by the constraints of locality and unitarity given above. It is actually a good thing that this is the case, since it gives us a way out of what would otherwise be a paradox. In [64] it was shown that the holomorphic soft weight of a positive helicity graviton is $\sigma = -3$, and similarly in [72] that the corresponding weight of a positive helicity gluon is $\sigma = -2$. These values correspond to higher-order (non-simple) poles in apparent tension with naive locality.

Now we will show how the structure of these *degenerate* singularities are in fact determined by the locality and unitarity properties described above. The key technical innovation will be to introduce a second complex parameter $z$

$$
\begin{aligned}
|\hat{s}\rangle &= \epsilon|s\rangle - z|X\rangle, \\
|\hat{i}] &= |i] - \epsilon\frac{\langle js\rangle}{\langle ji\rangle}|s] + z\frac{\langle jX\rangle}{\langle ji\rangle}|s], \\
|\hat{j}] &= |j] - \epsilon\frac{\langle is\rangle}{\langle ij\rangle}|s] + z\frac{\langle iX\rangle}{\langle ij\rangle}|s].
\end{aligned} \tag{2.2.6}
$$

Here we have introduced an arbitrary spinor $|X\rangle$. Again, it is straightforward to verify that this defines a two-parameter function onto $\mathcal{K}_{n+1}$. Recalculating the previously degenerate factorization channel gives

$$
(\hat{p}_s(\epsilon, z) + p_k)^2 = \left(\epsilon - z\frac{\langle Xk\rangle}{\langle sk\rangle}\right)\langle sk\rangle[sk], \tag{2.2.7}
$$

the zeroes are these functions are located at

$$
\epsilon_k(z) = z\frac{\langle Xk\rangle}{\langle sk\rangle}. \tag{2.2.8}
$$

The important point is that for $z \neq 0$ these zeros are located at distinct locations for distinct values of $k$, and therefore (2.2.6) defines a regular momentum deformation. Our strategy will be to use our understanding of these factorization singularities to write down the parts of the amplitude which generate the singular terms in the soft expansion in the limit $z \to 0$. Explicitly this is

$$
\hat{\mathcal{A}}_{n+1}(\epsilon, z) = \sum_{k=1}^{n}\sum_{\psi_k}\frac{\mathcal{A}_3\left(s^{h_s}, k^{h_k}, (-\hat{P}_{sk})_{\psi_k}\right)\mathcal{A}_n\left((\hat{P}_{sk})_{\psi_k}, \ldots\right)}{\epsilon\left(1 - \frac{z}{\epsilon}\frac{\langle Xk\rangle}{\langle sk\rangle}\right)\langle sk\rangle[sk]} + \mathcal{R}(\epsilon, z), \tag{2.2.9}
$$

where $\psi_k$ indexes the on-shell states exchanged in the $P_{sk}^2 = 0$ channel. The *remainder* $\mathcal{R}(\epsilon, z)$ contains all other singular and non-singular terms. It is straightforward to see that no other poles in $\epsilon$ are generated as $z \to 0$ other than those that appear in the terms explicitly shown. In this expression $z$ is acting as a kind of regulator, blowing up a degenerate singularity into a sum over isolated simple poles with known residues.

We can be even more explicit by making use of the fact that massless three-particle amplitudes are essentially determined by little-group scaling up to an overall coefficient. To begin with we will solve for the internal on-shell spinors at the singularity $\hat{P}_{sk}^2 = 0$, using the Schouten identity it is straightforward to show that

$$\hat{p}_s(\epsilon_k(z), z) + p_k = -|k\rangle \left( [k| + z\frac{\langle sX\rangle}{\langle sk\rangle}[s| \right), \tag{2.2.10}$$

and therefore

$$|\hat{P}_{sk}\rangle = |k\rangle, \quad |\hat{P}_{sk}] = |k] + z\frac{\langle sX\rangle}{\langle sk\rangle}|s]. \tag{2.2.11}$$

Since the internal angle bracket is equal to $|k\rangle$, we must be working in the square-only branch of complex three-particle kinematics. The explicit form of the three-particle amplitude we need is[1]

$$\mathcal{A}_3\left(s^{h_s}, k^{h_k}, (-\hat{P}_{sk})_{\psi_k}\right) = c_{\psi_k}[sk]^{h_s+h_k-h_{\psi_k}}[k, \hat{P}_{sk}]^{h_k+h_{\psi_k}-h_s}[\hat{P}_{sk}, s]^{h_{\psi_k}+h_s-h_k}. \tag{2.2.12}$$

Inserting this expression into (2.2.9) gives

$$\hat{\mathcal{A}}_{n+1}(\epsilon, z) = \sum_{k=1}^{n}\sum_{\psi_k} c_{\psi_k}\frac{[sk]^{2h_s-a}\langle Xs\rangle^{1-a}}{\epsilon z^{a-1}\langle sk\rangle^{2-a}\left(1-\frac{z}{\epsilon}\frac{\langle Xk\rangle}{\langle sk\rangle}\right)}\hat{\mathcal{A}}_n^{(\psi_k)}(z) + \mathcal{R}(\epsilon, z), \tag{2.2.13}$$

where $a = h_s - h_k - h_{\psi_k} + 1$ and

$$\hat{\mathcal{A}}_n^{\psi_k}(z) =$$
$$\mathcal{A}_n\left(\left\{|k\rangle, |k] + z\frac{\langle sX\rangle}{\langle sk\rangle}\right\}_{\psi_k}, \{|i\rangle, |i] + z\frac{\langle Xs\rangle\langle kj\rangle}{\langle sk\rangle\langle ji\rangle}|s]\}, \{|j\rangle, |j] + z\frac{\langle Xs\rangle\langle ki\rangle}{\langle sk\rangle\langle ij\rangle}|s]\}, ...\right). \tag{2.2.14}$$

As described above, we can recover the soft expansion by taking the $z \to 0$ limit of the expression (2.2.13), this remarkably compact *master equation* encodes all singular soft theorems in four-dimensional massless field theories at tree-level. For a given model, the singular terms depend on a small amount of *soft data* consisting of: the spectrum of of on-shell states and the values of the three-particle coupling constants $c_{\psi_k}$. However, not all choices of this data correspond to a consistent quantum field theory.

By a cursory examination of the master equation (2.2.13), shows that in general it takes the form of a double Laurent series in $z$ and $\epsilon$

$$\hat{\mathcal{A}}_n(\epsilon, z) = \sum_{m,n} C_{mn} z^m \epsilon^n. \tag{2.2.15}$$

Terms can then be organized into three categories:

- **Important:** $m = 0$ and $n < 0$, these terms survive the $z \to 0$ limit and generate the soft

[1] Here we have assumed the phase convention $|-P] = |P], \; |-P\rangle = -|P\rangle$.

expansion (2.1.4).

- **Unimportant:** $m > 0$ or $n > 0$, these terms either vanish in the $z \to 0$ limit or generate terms which are non-singular (and hence non-universal) in the soft expansion.
- **Dangerous:** $m < 0$, these terms give a naive obstruction to taking the $z \to 0$ limit, and therefore must cancel among themselves in a consistent model.

## 2.3 Soft Limit Consistency Conditions

The assertion that the so-called *dangerous* terms in (2.2.13) with negative powers of $z$ must cancel requires some commentary. As described above, locality requires that the amplitude $\hat{\mathcal{A}}_{n+1}(\epsilon, z)$ can become singular only when some invariant mass (2.2.3) has a zero. For generic $\epsilon \neq 0$ and $z = 0$ it is straightforward to verify using the explicit form of the deformation (2.2.6) that no such zero occurs, and therefore the amplitude must be regular in the limit $z \to 0$. If the soft data of a model contains interactions with $a > 1$, then such dangerous terms will occur and must therefore cancel among themselves to restore locality. This constraint imposes consistency conditions on the soft data, which we will now explore.

### 2.3.1 Charge Conservation and the Equivalence Principle

We will first consider the constraints of the soft limit of a massless spin-1 particle $\gamma$ coupled to a set of massless particles with spin-0 labelled $\phi_i$. Without loss of generality the helicity of the soft particle is taken to be $h_s = +1$ and therefore the holomorphic soft limit described above is appropriate. The logic of this section is that we will assume no additional properties beyond what we have described (to expedite the analysis we will assume Bose/Fermi symmetry to exclude certain amplitudes at the beginning, though this is not necessary), we will pretend that we have never heard of gauge theories and their associated constraints, and instead we will rediscover such properties through the soft constraints. In addition to the assumed spectrum we will assume the model contains a set of interactions of the form

$$\mathcal{A}_3\left(1_\gamma^+, 2_{\phi_i}, 3_{\phi_j}\right) = c_{ij}\frac{[12][13]}{[23]}. \tag{2.3.1}$$

We can see immediately that Bose symmetry gives the constraint $c_{ij} = c_{ji}$. The value of the parameter $a$ for this interaction is $a = 2$ and therefore there is a $z^{-1}$ term in the master equation with an associated constraint. To write down this constraint explicitly we first note that on a relevant factorization channel

$$\mathcal{A}_n\left(1_\gamma^+, 2_{\phi_i}, ...\right) \xrightarrow{s_{12}=0} \sum_j \mathcal{A}_3\left(1_\gamma^+, 2_{\phi_i}, (-P_{12})_{\phi_j}\right) \mathcal{A}_{n-1}\left((P_{12})_{\overline{\phi}_j}, ...\right), \tag{2.3.2}$$

where $\overline{\phi}_j$ is the CPT conjugate of the state $\phi_j$, *a priori* these states may or may not be distinct. We will consider the constraints associated with a general amplitude of the form

$$\mathcal{A}_{n+1}\left(s_\gamma^+; 1_{\phi_{i_1}}, ..., m_{\phi_{i_m}}; (m+1)_\gamma, ..., n_\gamma\right). \tag{2.3.3}$$

Ignoring overall constants, the condition for the coefficient of the $z^{-1}$ term to vanish is

$$\sum_{l=1}^{m} \sum_j c_{i_l j} \mathcal{A}_n\left(1_{\phi_{i_1}}, ..., \hat{l}_{\overline{\phi}_j}, ..., m_{\phi_{i_m}}; (m+1)_\gamma, ..., n_\gamma\right) = 0, \tag{2.3.4}$$

where the notation means that we have made the replacement $\phi_{i_l} \to \overline{\phi}_j$. The only non-trivial way such a set of constraints could be satisfied in general is if the non-vanishing three-particle couplings are of the form

$$\mathcal{A}_3\left(1_\gamma^+, 2_{\phi_i}, 3_{\overline{\phi}_i}\right) = q_i \frac{[12][13]}{[23]}. \tag{2.3.5}$$

In this case the $n$-point amplitude factors out of the above and the constraint reduces to

$$\left(\sum_{l=1}^{m} q_{i_l}\right) \mathcal{A}_n\left(1_{\phi_{i_1}}, ..., m_{\phi_{i_m}}; (m+1)_\gamma, ..., n_\gamma\right) = 0. \tag{2.3.6}$$

We then arrive at a remarkable observation, first made by Weinberg long-ago [60, 73] using a different though related argument. If we can make the following assignment of charges $Q[\phi_i] = q_i$ and $Q[\overline{\phi}_i] = -q_i$, then either the sum of all out-going charges is zero or the amplitude vanishes. By insisting on locality, Lorentz invariance and unitarity of the tree-level S-matrix we discover that massless spin-1 fields are necessarily associated with a conserved global charge.

As a second example, we consider a model of a massless spin-2 particle $h$ coupled to a set of massless spin-0 particles $\phi_i$ through $a = 3$ interactions of the form

$$\mathcal{A}_3\left(1_h^+, 2_{\phi_i}, 3_{\phi_j}\right) = d_{ij} \frac{[12]^2[13]^2}{[23]^2}. \tag{2.3.7}$$

The master equation (2.2.13) for this model has both a $z^{-2}$ and $z^{-1}$ pole. The former must cancel independently so we will begin with the constraint in isolation. Using a similar notation to the above we have

$$\sum_{l=1}^{m}\sum_{j} d_{i_l j}[sl]\langle ls\rangle \mathcal{A}_n\left(1_{\phi_{i_1}},...,\hat{l}_{\overline{\phi}_j},...,m_{\phi_{i_m}};(m+1)_h,...,n_h\right)=0. \qquad (2.3.8)$$

Unlike the spin-1 case, there is no non-trivial choice of the couplings $d_{ij}$ that can satisfy this constraint. Even if we diagonalize the three-particle interactions as

$$\mathcal{A}_3\left(1_h^+,2_{\phi_i},3_{\overline{\phi}_i}\right)=d_i\frac{[12]^2[13]^2}{[23]^2}, \qquad (2.3.9)$$

the resulting expression has the form

$$\left(\sum_{l=1}^{m} d_l p_l^\mu\right)\mathcal{A}_n\left(1_{\phi_{i_1}},...,m_{\phi_{i_m}};(m+1)_h,...,n_h\right)=0. \qquad (2.3.10)$$

Again, as realized by Weinberg [60, 73], the only non-trivial way to satisfy this constraint is to choose the coupling strength to be universal $d_i=\kappa$, and to also include a self-interaction term

$$\mathcal{A}_3\left(1_h^+,2_h^+,3_h^-\right)=\kappa\frac{[12]^6}{[23]^2[31]^2}. \qquad (2.3.11)$$

In this case we then have

$$\left(\kappa\sum_{l=1}^{n} p_l^\mu\right)\mathcal{A}_n\left(1_{\phi_{i_1}},...,m_{\phi_{i_m}};(m+1)_h,...,n_h\right)=0, \qquad (2.3.12)$$

which is always true, by virtue of momentum conservation! So we discover the equally remarkable fact that if there is a massless spin-2 field in the spectrum, it must couple to every other state including itself with a universal strength $\kappa$, an on-shell statement of the Einstein equivalence principle.

We are not quite finished, we have only verified that the $z^{-2}$ poles cancel, there is also the subleading $z^{-1}$ pole, it is instructive to verify that this also cancels. There are two contributions, the first from Taylor expanding the denominator factor in (2.2.13)

$$\kappa\sum_{l=1}^{n}[sl]\langle lX\rangle\mathcal{A}_n\left(1_{\phi_{i_1}},...,m_{\phi_{i_m}};(m+1)_h,...,n_h\right), \qquad (2.3.13)$$

which clearly vanishes for the same reason as the leading order pole. There is also a second contribution from Taylor expanding $\hat{\mathcal{A}}_n^{\psi_k}(z)$ which simplifies to

$$J_{ab}\cdot\mathcal{A}_n\left(1_{\phi_{i_1}},...,m_{\phi_{i_m}};(m+1)_h,...,n_h\right), \qquad (2.3.14)$$

where

$$J_{ab} = \sum_{l=1}^{n} \left( [l|_a \nabla_b^{(l)} + [l|_b \nabla_a^{(l)} \right), \qquad \nabla_a^{(l)} = \partial_{|l]^a} - \frac{\langle jl \rangle}{\langle ji \rangle} \partial_{|i]^a} - \frac{\langle il \rangle}{\langle ij \rangle} \partial_{|j]^a}. \tag{2.3.15}$$

This also vanishes, a fact that is transparent once we recognize $J_{ab}$ as the generator of the Lorentz group. Somewhat miraculously, we find that the universal coupling also guarantees the absence of subleading $z$-poles, though the complete cancellation required appealing to the invariance of the amplitudes under the complete Poincare group.

### 2.3.2 No-Go for Massless Higher Spin

This modern version of the Weinberg argument naturally extends to massless particles with $s \geq 3$. The argument is identical except that we find, for diagonalized three-particle interactions

$$\mathcal{A}_3 \left( 1_X^s, 2_{\phi_i}, 3_{\overline{\phi}_i} \right) = e_i \left( \frac{[12][13]}{[23]} \right)^s, \tag{2.3.16}$$

a condition for the cancellation of a leading $z^{-(a-1)}$ pole which can only be satisfied in the trivial case $e_i = 0$, even if we allow for additional degrees of freedom and self-interactions. This result is usually interpreted as the statement that it is impossible for massless $s > 2$ particles to mediate long-range forces. Momentarily we will justify this statement as a corollary of a more general result constraining the allowed three-particle interactions in a local, unitary and Lorentz invariant model.

The starting point for this analysis is to consider the following four-particle amplitude

$$\mathcal{A}_4 \left( 1_{b_1}^{h_1}, 2_{b_2}^{h_2}, 3_{\bar{b}_1}^{-h_1}, 4_{\bar{b}_2}^{-h_2} \right). \tag{2.3.17}$$

Where here the bars denote CPT conjugate states. Taking $p_1 \to 0$ will in general generate contributions to the master equation (2.2.13) in all three channels. Without loss of generality we will assume that some contribution to the $s_{12}$-channel has the largest value of the parameter $a$ (defined below (2.2.13)) of any three-particle interaction in the model. As we have seen in the examples presented in the previous subsection, such a pole can cancel in a non-trivial way if the interactions have a certain amount of structure. The condition for the cancellation of the most singular $z$-pole is explicitly

$$\sum_c g_{b_1 b_2 c} \langle 12 \rangle^{a-2} [12]^{2h_1 - a} \mathcal{A}_3^{(12)} + \sum_d g_{b_1 \bar{b}_1 d} \langle 13 \rangle^{a-2} [13]^{2h_1 - a} \mathcal{A}_3^{(13)} + \sum_e g_{b_1 \bar{b}_2 e} \langle 14 \rangle^{a-2} [14]^{2h_1 - a} \mathcal{A}_3^{(14)} = 0. \tag{2.3.18}$$

Where here the sums over $c$, $d$ and $e$ are taken over any internal degrees of freedom that are relevant for generating the most singular pole. Until now we have been completely general, to proceed we will assume that $a > 3$ and demonstrate that there is no non-trivial choice of couplings which satisfies this constraint. The first step is to show that no cancellation between the factorization channels is possible. Recall that in our kinematics the spinors $|1\rangle$ and $|1]$ are independent, so we are free to set them to whatever value we like in this expression. We make the following choice

$$|1\rangle = x|3\rangle + y|4\rangle, \tag{2.3.19}$$

this gives

$$\begin{aligned}\sum_c g_{b_1 b_2 c} \left(x\langle 32\rangle + y\langle 42\rangle\right)^{a-3} [12]^{2h_1-a} \mathcal{A}_3^{(12)} + \sum_d g_{b_1 \bar{b}_1 d} y^{a-3} \langle 43\rangle^{a-3} [13]^{2h_1-a} \mathcal{A}_3^{(13)} \\ + \sum_e g_{b_1 \bar{b}_2 e} x^{a-3} \langle 34\rangle^{a-3} [14]^{2h_1-a} \mathcal{A}_3^{(14)} = 0.\end{aligned} \tag{2.3.20}$$

Here $x$ and $y$ are free-parameters and so this constraint must hold independently on distinct powers $x^m y^n$. Since mixed powers $m, n \neq 0$ only appear in the $s_{12}$-channel terms, these must sum to zero independently. Notice that this conclusion required $a > 3$ in order for $\left(x\langle 32\rangle + y\langle 42\rangle\right)^{a-3}$ to generate an $xy$ mixed term, hence this argument does not apply to the soft graviton analysis for which the interactions had $a = 3$. To complete the argument we insist that our model is CPT invariant, in which case the coupling constant associated with the amplitude $\mathcal{A}_3^{(12)}$ is simply the complex conjugate of $g_{b_1 b_2 c}$. The remaining spinor brackets are then common to each term in the sum over the internal quantum number $c$ and so factor out. The constraint that must be satisfied is then

$$\sum_c |g_{a_1 a_2 c}|^2 = 0. \tag{2.3.21}$$

Since this is manifestly a sum over non-negative numbers the only solution is given by $g_{b_1 b_2 c} = 0$. Hence, we have proven, using Lorentz invariance, locality, unitarity and CPT symmetry, that three-particle interactions with $a > 3$ are impossible.

This result has immediate and dramatic consequences for massless higher-spin fields. As described above, if a massless particle of spin-$s$ interacts *in any way* with a graviton, then it must couple to gravity directly with a universal strength $\kappa$

$$\mathcal{A}_3\left(1_h^+, 2_X^{+s}, 3_{\bar{X}}^{-s}\right) \propto \kappa. \tag{2.3.22}$$

If we take $X$ soft, then the relevant value is $a_X = 2s + 1$. Requiring $a_X \leq 3$ is equivalent to $s \leq 2$, and so we have proven that massless higher-spin fields cannot exist in a model of gravity. Different arguments leading to the same conclusion have previously been given [73–75].

A similar conclusion follows if we consider couplings to a photon

$$\mathcal{A}_3\left(1_\gamma^+, 2_{X_i}^{+s}, 3_{\bar{X}_i}^{-s}\right) \propto q_i. \tag{2.3.23}$$

As we demonstrated above $q_i$ is proportional to the electric charge of the state $X_i$. Again, taking a soft limit for $X$, the associated value is $a_X = 2s$. As a consequence of our general result we discover that massless states with $s \geq 3/2$ must have zero electric charge. This result demonstrates that the constraints on massless higher-spin particles extends beyond the requirements of a consistent coupling to Einstein gravity. One could imagine that there exists a landscape of gauge theories coupled to higher-spin massless fields which are classically consistent (in the sense that there exists a local, unitary and Lorentz invariant tree-level S-matrix) but which cannot be coupled to a dynamical gravitational field. Our general result excludes this possibility.

## 2.4 Soft Limits and Effective Field Theory

### 2.4.1 Higher-Derivative Corrections to Soft Photon Theorem

In the previous subsection we rediscovered the result of Weinberg [60, 73], that a massless spin-1 field $\gamma$ (henceforth we will call this a photon) interacting with a sector of matter fields $\phi_i$ via the $a = 2$ *minimal coupling* is consistent only if the strength of the coupling is proportional to a globally conserved additive charge $q_i$ carried by particle $i$. If this condition is satisfied, and all other couplings in the model have $a < 1$, then the master equation (2.2.13) predicts a soft expansion of the form

$$\hat{\mathcal{A}}_{n+1}\left(s_\gamma^+; 1, ..., n\right) = \sum_{i=1}^{n} q_i \left[\frac{1}{\epsilon^2}\frac{\langle Xi\rangle}{\langle Xs\rangle\langle si\rangle} + \frac{1}{\epsilon}\frac{1}{\langle si\rangle}\nabla_{si}\right]\mathcal{A}_n\left(1, ..., n\right) + \mathcal{O}\left(\epsilon^0\right). \tag{2.4.1}$$

This is the familiar soft photon theorem, including the subleading correction first discovered long-ago [57].

The leading $\epsilon^{-2}$ term is universal, any attempt to modify it requires adding $a = 2$ interactions for a soft photon. Such a deformation will reintroduce the dangerous $z^{-1}$ terms for which the only non-trivial solution is the minimal coupling of electrically charged matter fields. The subleading $\epsilon^{-1}$ term however is only *quasi*-universal, and may recieve additional contributions from higher-derivative effective operators. From (2.2.13) we can see that additional $\epsilon^{-1}$ terms may be generated if there are three-particle interactions of a positive helicity photon to two other particles with $a = 1$ of the form

$$\mathcal{A}_3\left(1_\gamma^{+1}, 2_X^{h_X}, 3_Y^{h_Y}\right) = g_{\gamma XY}[12]^{1+h_X-h_Y}[23]^{h_X+h_Y-1}[31]^{h_Y+1-h_2}, \tag{2.4.2}$$

with the constraint $h_X + h_Y = 1$. Such an interaction does not generate dangerous $z$-poles, and hence the associated couplings are unconstrained by consistency in the soft photon limit. Since the holomorphic soft limit () only has contributions from two-particle factorization channels in which the three-particle amplitude containing the soft photon is supported on the all-square branch of three-particle kinematic space we have the additional constraint that $h_X + h_Y \geq 0$. Without loss of generality we will assume $h_X \geq h_Y$ and hence $h_X > 0$; additionally, this requires that we impose the general constraint $a_X = h_X - h_Y \leq 3$, which together with the above gives $h_X \leq 2$. Putting all of these together we generate a finite list of allowed helicity values

$$(h_X, h_Y) \in \{(2,-1), (3/2,-1/2), (1,0), (1/2,1/2)\}. \tag{2.4.3}$$

When such an interaction is present the soft-photon theorem (2.4.1) contains new terms at subleading order of the form

$$\hat{\mathcal{A}}_{n+1}\left(s_\gamma^+; 1, ..., i_X, ..., n\right) \supset \frac{g_{\gamma XY}}{\epsilon} \frac{[si]}{\langle si\rangle} \mathcal{A}_n\left(1, ..., i_{\bar{Y}}, ..., n\right). \tag{2.4.4}$$

An important qualitative difference between (2.4.4) and (2.4.1), in the former the entire expression is factored into a product of a universal *soft factor* and the $n$-point scattering amplitude with the soft photon removed and the remaining hard states the same. In the latter, the amplitude that appears on the right-hand-side has the CPT conjugate of the state $Y$ in the place that on the left-hand-side was occupied by $X$. From the list of allowed helicities (2.4.3), we see that these states are necessarily different and therefore the addition of such interactions necessarily violates soft factorization at subleading order.

From the on-shell matrix elements we can reverse engineer the corresponding local operators that would appear in the Lagrangian. Since the mass-dimension of a three-particle amplitude in $d = 4$ is 1, and the mass-dimension of a spinor bracket is also 1, we learn that the couplings have dimension $[g_{\gamma XY}] = -1$, and the associated local operators must then have dimension 5. In the same order as (2.4.3), the explicit form of the local operators are

$$\mathcal{O} \in \{h_{\mu\nu} F^{+\mu\rho} F^{-\nu}_{\rho},\;\; \psi^{\dagger\mu} \bar{\sigma}^\nu \chi F^+_{\mu\nu},\;\; \phi (F^+)^2,\;\; \chi_i^\dagger \bar{\sigma}^{\mu\nu} \chi_j^\dagger F_{\mu\nu}\}, \tag{2.4.5}$$

where $F^{\pm}_{\mu\nu}$ denotes the (anti-)self-dual components of the Maxwell field strength tensor; additionally $h$, $\psi$, $\chi$ and $\phi$ are the standard elementary field operators for massless particles of spin $2$, $3/2$, $1/2$ and $0$ respectively.

### 2.4.2 Higher-Derivative Corrections to Soft Graviton Theorem

We can repeat the above discussion and determine the possibility of higher-derivative operators which generate quasi-universal corrections to the soft graviton theorem at sub- and sub-sub-leading order. First, from the master formula (2.2.13), the standard soft-graviton theorem to sub-sub-leading order is given by

$$\hat{\mathcal{A}}_{n+1}\left(s_h^{+2};1,...,n\right) = \kappa\sum_{i=1}^{n}\left[\frac{1}{\epsilon^3}\frac{[si]\langle Xi\rangle^2}{\langle si\rangle\langle Xs\rangle^2}+\frac{1}{\epsilon^2}\frac{\langle Xi\rangle[si]}{\langle Xs\rangle\langle si\rangle}\nabla_{si}+\frac{1}{2\epsilon}\frac{[si]}{\langle si\rangle}\nabla_{si}^2\right]\mathcal{A}_n\left(1,...,n\right)+\mathcal{O}\left(\epsilon^0\right). \quad (2.4.6)$$

The leading $\epsilon^{-3}$ term is the famous Weinberg soft graviton theorem [60], the sub- $\epsilon^{-2}$ and sub-subleading $\epsilon^{-1}$ terms were first written down in [64]. The analysis here is a little different from the above case of soft photons, in principle higher-derivative operators may modify the soft graviton theorem at either subleading or sub-subleading order. We will begin with the former case, following the same logic as above this requires an interaction with $a=2$ of the form

$$\mathcal{A}_3\left(1_h^{+},2_X^{h_X},3_Y^{h_Y}\right)=g_{\gamma XY}[12]^{2+h_X-h_Y}[23]^{h_X+h_Y-2}[31]^{h_Y+2-h_2}, \quad (2.4.7)$$

with

$$(h_X,h_Y)\in\{(2,-1),(3/2,-1/2),(1,0),(1/2,1/2)\}. \quad (2.4.8)$$

These interactions are constrained since they generate a dangerous $z^{-1}$ term. The explicit form of the constraint is

$$\sum_{i,c}g_{ic}[si]^2\mathcal{A}_n\left(1,...,\hat{i}_c,...,n\right)=0, \quad (2.4.9)$$

where the sum $i$ includes all external states $\{1,...,n\}$ with helicity $2,3/2,1,1/2,0,-1/2$ or $-1$, the sum over $c$ is over all possible internal quantum numbers and the notation $\hat{i}$ indicates that the external state is the CP conjugate of the remaining state $Y$ in the three-particle $s-i-Y$ interaction.

We claim that no non-trivial solution to such constraints is possible. The method for proving this is similar to the derivation of our general bound $a\leq 3$ above, we consider specially constructed four-particle amplitudes. Consider the four-particle amplitude

$$\mathcal{A}_4\left(1_h^{+2},2_h^{+2},3_h^{-2},4_h^{-2}\right). \quad (2.4.10)$$

The only possible factorization into one of the above interactions is in the $s_{12}$-channel $h_X=+2$. The argument is then the same as for the $a\leq 3$ bound, the constraint is proportional to $\sum_c|g_{2c}|^2$

which is zero only if $g_{2c} = 0$, hence such a coupling is impossible. Next we consider an amplitude

$$\mathcal{A}_4\left(1_h^{+2}, 2_{\psi_b}^{+3/2}, 3_h^{-2}, 4_{\psi_{\bar{b}}}^{-3/2}\right), \tag{2.4.11}$$

Where $\psi_{\bar{b}}$ indicates the CPT conjugate of $\psi_b$. Likewise this only has a contribution from the $s_{12}$-channel and so the couplings must be zero. Next we consider

$$\mathcal{A}_4\left(1_h^{+2}, 2_{\gamma_b}^{+1}, 3_h^{-2}, 4_{\gamma_{\bar{b}}}^{-1}\right). \tag{2.4.12}$$

This would naively have contributions in both the $s_{12}$-channel from the $h_X = 1$ interaction and in the $s_{14}$-channel from the $h_X = 2$ interaction. The latter however has already be proven to vanish, hence the former must also be zero. Finally we consider

$$\mathcal{A}_4\left(1_h^{+2}, 2_{\chi_b}^{+1/2}, 3_h^{-2}, 4_{\chi_{\bar{b}}}^{-1/2}\right). \tag{2.4.13}$$

Similar to the last case, this could have received a contribution from the $h_X = 1/2$ interaction in the $s_{12}$-channel and the $h_X = 3/2$ interaction in the $s_{14}$-channel, but the latter has already been shown to vanish and so must then the former.

All together we have proven that constraint (2.4.9) can never be non-trivially satisfied. The important consequence of this analysis is then that no higher-derivative operators can modify the soft graviton theorem at sub-leading order. With the possible exception of loops of massless states, this result indicates that the subleading soft graviton theorem is truly universal. In the context of the discussion at the beginning of this chapter, this result can be interpreted as the statement that the asymptotic Virasoro symmetry of quantum gravity conjectured in [65], is unmodified by $\alpha'$-type higher-derivative operators.

Now we move on to considering higher-derivative corrections to the sub-subleading soft graviton theorem. Here there are no dangerous terms and hence no constraints. We need interactions with $a = 1$, the possibilities are

$$(h_X, h_Y) \in \{(2,0), (3/2, 1/2), (1,1)\}. \tag{2.4.14}$$

In each case we generate a contribution to the soft expansion of the form

$$\hat{\mathcal{A}}_{n+1}\left(s_h^{+2}; 1, ..., i_X, ..., n\right) \supset \frac{g_{hXY}}{\epsilon}\frac{[si]^3}{\langle si\rangle}\mathcal{A}_n\left(1, ..., i_{\bar{Y}}, ..., n\right). \tag{2.4.15}$$

Simple dimensional analysis tells us that such interactions correspond to local operators with mass dimension 7, explicitly the operators are

$$\mathcal{O} \in \{\phi R^+_{\mu\nu\rho\sigma}R^{+\mu\nu\rho\sigma}, \quad R^{+\mu\nu\rho\sigma}\psi^\dagger_\mu\bar{\sigma}_{\nu\rho}\partial_\sigma\chi, \quad R^{+\mu\nu\rho\sigma}F^+_{\mu\nu}F^+_{\rho\sigma}\}, \tag{2.4.16}$$

where $F^+_{\mu\nu}$ denotes the self-dual components of the Maxwell field strength tensor and $R^+_{\mu\nu\rho\sigma}$ denotes the self-dual component of the Riemann tensor; additionally $\psi$, $\chi$ and $\phi$ are the standard elementary field operators for massless particles of spin $3/2$, $1/2$ and $0$ respectively.

# CHAPTER 3

# Vanishing Soft Limits and the Goldstone S-Matrix

## 3.1 Overview of Goldstone EFTs

The general motivation for approaching EFTs for Goldstone modes using scattering amplitudes was outlined in Section 1.2.1. The technical content of this chapter concerns the development of a numerical algorithm called the *soft bootstrap*, which implements the inverted approach of directly engineering a low-energy S-matrix satisfying assumed low-energy theorems, and then *post hoc* interpreting the result in terms of a pattern of spontaneous symmetry breaking. The numerical soft bootstrap allows us to efficiently answer two important questions:

- When does a model with an assumed spectrum, linear symmetries and low-energy theorems fail to exist?
- When is such a model unique?

Unfortunately we will not be able to prove that a given model exists, though in some cases once uniqueness is demonstrated the model can be identified with an explicit construction, as was the case for DBI refintro:let. This approach forgoes the use of off-shell effective actions entirely (and so avoids the problem of grappling with field-reparametrization redundancy), and instead provides an explicit recipe for the direct construction of the on-shell S-matrix through the recently discovered *soft subtracted recursion*.

The soft bootstrap program was initiated in [76], where it was used to explore the landscape of real scalar EFTs with vanishing low-energy theorems. The results are reviewed and extended in Section 3.3. This chapter should be understood as a continuation and generalization of this program, incorporating richer soft data including spinning particles and linearly realized supersymmetry. In this chapter, we extend the application of the soft bootstrap from real scalars to any massless helicity-$h$ particle and we derive a precise criterion for the validity of the soft subtracted recursion relations.

| Soft degree $\sigma$ | Spin $s$ | Type of symmetry breaking |
|---|---|---|
| 1 | 0 | Internal symmetry (symmetric coset) |
| 0 | 0 | Internal symmetry (non-symmetric coset) |
| 1 | 1/2 | Supersymmetry |
| 0 | 0 | Conformal symmetry |
| 0 | 1/2 | Superconformal symmetry |
| 2 | 0 | Higher-dimensional Poincaré symmetry |
| 0 | 0 | Higher-dimensional AdS symmetry |
| 3 | 0 | Special Galileon symmetry |

Table 3.1: The table lists soft weights $\sigma$ associated with the soft theorems $\mathcal{A}_n \to O(\epsilon^\sigma)$ as $\epsilon \to 0$ for several known cases. The soft limit is taken holomorphically in 4d spinor helicity, see Section 2.1 for a precise definition.

Table 3.1 summarizes the soft weights for various known cases of spontaneous symmetry breaking. Here follows a brief overview of the Goldstone EFTs that appear in this chapter. We include the connection between their soft behavior and Lagrangian shift symmetries:

- **DBI** can be extended to a complex scalar Dirac-Born-Infeld theory and coupled supersymmetrically to a fermion sector described by the **Akulov-Volkov** action of Goldstinos from spontaneous breaking of supersymmetry. In extended supersymmetric DBI, the vector sector is **Born-Infeld (BI)** theory. The soft weights are $\sigma_Z = 2$ for the complex scalars $Z$ of DBI, $\sigma_\psi = 1$ for the fermions of Akulov-Volkov, and $\sigma_\gamma = 0$ for the BI photon. The soft behaviors can be associated with shift symmetries $Z \to Z + c + v_\mu x^\mu$ and $\psi \to \psi + \xi$, where $\xi$ is a constant Grassmann-number.[1] $\mathcal{N} = 1$ **supersymmetric Born-Infeld** couples the BI vector to the Goldstino of Akulov-Volkov.

- **Nonlinear sigma models (NLSM)** describe the Goldstone modes of sponteneously broken internal symmetries and have scalars with constant shift symmetries that give $\sigma = 1$ soft weights in the low-energy theorems. A common example of an NLSM is **chiral perturbation theory** in which the scalars live in a coset space $U(N) \times U(N)/U(N)$.

  The complex scalar $\mathbb{CP}^1$ NLSM can be supersymmetrized with a fermion sector that is **Nambu-Jona-Lasinio (NJL)** model. The complex scalars have shift symmetry $Z \to Z + c$ and $\sigma_Z = 1$ while the fermions have no shift symmetry and $\sigma_\psi = 0$. We study both the $\mathcal{N} = 1$ and 2 supersymmetric $\mathbb{CP}^1$ NLSM.[2]

- A NLSM can have a non-trivial subleading operator that respects the shift symmetry and

[1] We leave out field-dependent terms for simplicity when stating the shift symmetries.

[2] In Section 3.5.2 we show that the $\mathcal{N} = 2$ $\mathbb{CP}^1$ NLSM requires the presence of 3-point interactions and the soft weight of the scalar is reduced to $\sigma_Z = 0$.

hence also the low-energy theorems with $\sigma = 1$. This operator is known as the **Wess-Zumino-Witten (WZW)** term and has a leading 5-point interaction.

- **Galileon** scalar EFTs arise in various contexts and have the extended shift symmetry $\phi \to \phi + c + v_\mu x^\mu$ that gives low-energy theorems with $\sigma = 2$. As such they can be thought of as subleading operators of the DBI action, and are called **DBI-Galileons**. They can also be decoupled from DBI (at the cost of having no UV completion).

  In 4d there are two independent Galileon operators: the quartic and quintic Galileon. (By a field redefinition, the cubic Galileon is not independent from the quartic and quintic.) When decoupled from DBI, the quartic Galileon has an even further enhanced shift symmetry $\phi \to \phi + c + v_\mu x^\mu + s_{\mu\nu} x^\mu x^\nu$ that gives low-energy theorems with soft weight $\sigma = 3$ and is then called the **Special Galileon** [76,77].

- The quartic Galileon has a complex scalar version with $\sigma_Z = 2$ (but it cannot have $\sigma_Z = 3$). It has an $\mathcal{N} = 1$ supersymmetrization [52,78] in which the fermion sector trivially realizes a constant shift symmetry that gives $\sigma_\psi = 1$.

- There is evidence [52] that the quintic Galileon may have an $\mathcal{N} = 1$ supersymmetrization. This involves a complex scalar whose real part is a Galileon with $\sigma = 2$ and imaginary part is an R-axion with $\sigma = 1$.

### 3.1.1 Structure of the Effective Action

The low-energy dynamics of a physical system can be described by a Wilsonian effective action containing a set of local quantum fields for each of the on-shell asymptotic states with *all possible* local interactions allowed by the assumed symmetries:

$$S_{\text{effective}} = S_0 + \sum_{\mathcal{O}} \frac{c_{\mathcal{O}}}{\Lambda^{\Delta[\mathcal{O}]-4}} \int \mathrm{d}^4 x \, \mathcal{O}(x) \,. \tag{3.1.1}$$

Here $S_0$ denotes the free theory, i.e. the kinetic terms, $\Lambda$ is a characteristic scale of the problem, and $c_{\mathcal{O}}$ are dimensionless constants. The sum is over all local Lorentz invariant operators $\mathcal{O}(x)$ of the schematic form

$$\mathcal{O}(x) \sim \partial^A \phi(x)^B \psi(x)^C F(x)^D \,, \tag{3.1.2}$$

where $A, \ldots, D$ are integer exponents. In this chapter we focus on EFTs in which the operators $\mathcal{O}$ are manifestly gauge invariant.[3]

[3]This need not be the case in more general scenarios (though of course we insist on overall gauge invariance). For example in Yang-Mills theory, the gauge invariant operator $\mathrm{tr} F^2$ has a quadratic term which we group into the free part $S_0$ of the action while the interaction terms would be accounted for in the sum of all operators $\mathcal{O}$ in (3.1.1). Similarly, for massless spin-2 fields when $\sqrt{-g}R$ is expanded around flat space.

We assign the following quantities to a local operator

- *Dimension:* $\Delta[\mathcal{O}]$ defined as the engineering dimension with bosonic fields of dimension 1 and fermionic fields of dimension 3/2.
- *Valence:* $N[\mathcal{O}]$ defined as the sum of the total number of field operators appearing. Equivalently, this is the valence of the Feynman vertex derived from such an interaction.

The schematic operator in (3.1.2) has $\Delta[\mathcal{O}] = A + B + \frac{3}{2}C + 2D$ and $N[\mathcal{O}] = B + C + D$.

In standard EFT lore, operators of lowest dimension dominate in the IR. In many cases this means the marginal and relevant interactions dominate and the irrelevant interactions are sub-dominant and suppressed by powers of the UV scale $\Lambda$. In other cases, such as effective field theories describing the dynamics of Goldstone modes, there are only irrelevant interactions and it may be less clear which operators dominate. It is therefore useful to introduce the *reduced* dimension

$$\tilde{\Delta}[\mathcal{O}] = \frac{\Delta[\mathcal{O}] - 4}{N[\mathcal{O}] - 2} \tag{3.1.3}$$

for the operator basis (3.1.1). Operators that minimize $\tilde{\Delta}$ dominate in the IR.

The authors of [38, 76, 79] consider only scalar EFTs and therefore operators of the form $\mathcal{O} \sim \partial^m \phi^n$. They define a quantity

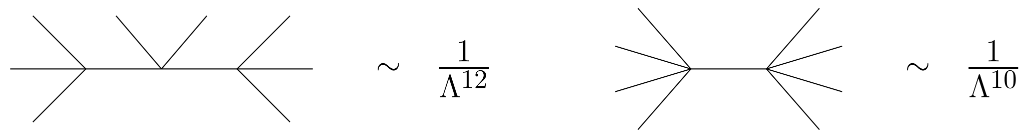
$$\rho \equiv \frac{m-2}{n-2} = \tilde{\Delta}[\mathcal{O}] - 1\,, \tag{3.1.4}$$

to determine when two operators of this form produce tree-level diagrams with couplings of the same mass dimension. Morally $\rho$ is the same as the reduced dimension $\tilde{\Delta}[\mathcal{O}]$. The latter is the natural generalization of $\rho$ to operators containing particles of all spins.

The quantity $\tilde{\Delta}$ is useful for clarifying the notion of what it means for an interaction to be leading order in an EFT with only irrelevant interactions. In the deep IR, the relative size of the dimensionless Wilson coefficients in the effective action is unimportant since lower dimension operators will *always* dominate over higher dimension operators. It is therefore only necessary to isolate the contributions that are leading in a power series expansion of the amplitudes in the inverse UV cutoff scale $\Lambda^{-1}$. The dominant interactions in the deep IR are generated by operators that *minimize* this quantity. As an illustrative example, consider an effective action for scalars with interaction terms of the form

$$S_{\text{effective}} \supset \int \mathrm{d}^4x \left[\frac{c_4}{\Lambda^4}\partial^4\phi^4 + \frac{c_5}{\Lambda^5}\partial^4\phi^5\right]. \tag{3.1.5}$$

The reduced dimensions $\tilde{\Delta}$ are $2$ and $5/3$ for the quartic and quintic interactions respectively. The quintic interaction should therefore dominate over the quartic in the deep IR. To see this explicitly we have to compare amplitudes with the *same* number of external states, so we compare the contributions from tree-level Feynman diagrams to the 8-point amplitude:



This confirms that the diagrams arising from the quintic interaction dominate the 8-point amplitude.

It is useful to introduce the notion of *fundamental interactions* (or *fundamental operators*) in an EFT. These are the lowest dimension operator(s) whose on-shell matrix elements can be recursed to define all matrix elements of the theory at leading order in the low-energy expansion.

Consider the DBI action. The leading interaction comes from an operator of the form $\frac{1}{\Lambda^4}\partial^4\phi^4$ and as discussed in the introduction, with the associated 4-point amplitude as input, all other $n$-point amplitudes in DBI can be constructed with soft subtracted recursion relations. If the action had contained an interaction term of the form $\frac{c_5}{\Lambda^5}\partial^5\phi^4$, then $\frac{1}{\Lambda^4}\partial^4\phi^4$ would not be sufficient to determine dominating contributions at $n$-point order, i.e. both interactions would need to be considered fundamental for soft recursion.

The operators immediately subleading to DBI in the brane-effective action are encoded in the DBI-Galileon. In 4d, there are two such independent couplings,[4] namely for a quartic interaction of the schematic form $\frac{b_4}{\Lambda^6}\partial^6\phi^4$ and a quintic interaction of the form $\frac{b_5}{\Lambda^9}\partial^8\phi^5$; these both have $\tilde{\Delta} = 3$ whereas DBI has $\tilde{\Delta} = 2$. Thus the DBI-Galileon has a total of three fundamental operators: the 4-point DBI interaction and the 4- and 5-point Galileon interactions.

## 3.2 Subtracted Recursion Relations

We review on-shell subtracted recursion relations for scattering amplitudes of Goldstone modes [38, 76, 79–81] and derive a new precise criterion for their validity.

### 3.2.1 Review of Soft Subtracted Recursion Relations

We consider complex momentum deformations of the form

$$p_i \to \hat{p}_i = (1 - a_i z)p_i \quad \text{with} \quad \sum_{i=1}^{n} a_i p_i = 0\,. \tag{3.2.1}$$

[4]The cubic Galileon interaction is equivalent to a particular linear combination of the quartic and quintic Galileon after a field redefinition.

The label $i = 1, 2, \ldots, n$ runs over the $n$ massless particles in the scattering amplitude. The shifted momenta $\hat{p}_i$ are on-shell by virtue of $p_i^2 = 0$ and satisfy momentum conservation when the shift coefficients $a_i$ satisfy the condition in (3.2.1). (We discuss the solutions to this condition in Section 3.2.4.) When evaluated on the shifted momenta $\hat{p}_i$, an $n$-point amplitude becomes a function of $z$ and we write it as $\hat{\mathcal{A}}_n(z)$.

The *subtracted recursion relations* for an $n$-point tree-level amplitude $\mathcal{A}_n$ are derived from the Cauchy integral

$$\oint \frac{dz}{z} \frac{\hat{\mathcal{A}}_n(z)}{F(z)} \;=\; 0\,, \tag{3.2.2}$$

where the contour surrounds all the poles at finite $z$ and the function $F$ is defined as

$$F(z) = \prod_{i=1}^{n} (1 - a_i z)^{\sigma_i}\,. \tag{3.2.3}$$

The vanishing of the integral in (3.2.2) requires absence of a simple pole at $z = \infty$. We derive a sufficient criterion for this behavior in Section 3.2.2.

The shift (3.2.1) is implemented on the spinor helicity variables according to the sign of the helicity $h_i$ of particle $i$ as

$$\begin{aligned} h_i \geq 0:& \qquad |i\rangle \to (1 - a_i z)|i\rangle\,, \qquad & |i] \to |i]\,, \\ h_i < 0:& \qquad |i\rangle \to |i\rangle\,, & |i] \to (1 - a_i z)|i]\,. \end{aligned} \tag{3.2.4}$$

The limit $z \to 1/a_i$ is then precisely the soft limit $\hat{p}_i \to 0$ of the $i$th particle in the deformed amplitude. Hence, *if the amplitude satisfies low-energy theorems of the form (3.4.5) with weights $\sigma_i$ for each particle $i$*, the integral (3.2.2) will not pick up any non-zero residues from poles arising from the function $F$ when it is chosen as in (3.2.3). Therefore the only simple poles in (3.2.2) arise from $z = 0$ and factorization channels in the deformed tree amplitude. They occur where internal momenta go on-shell, $\hat{P}_I^2 = 0$. The residue theorem then states that the residue at $z = 0$ equals minus the sum of all such residues, and factorization on these poles gives

$$\mathcal{A}_n = \hat{\mathcal{A}}_n(z=0) = \sum_I \sum_{|\psi^{(I)}\rangle} \sum_{\pm} \frac{\hat{\mathcal{A}}_L^{(I)}(z_I^{\pm})\hat{\mathcal{A}}_R^{(I)}(z_I^{\pm})}{F(z_I^{\pm}) P_I^2 (1 - z_I^{\pm}/z_I^{\mp})}\,. \tag{3.2.5}$$

The sums are over all factorization channels $I$, the two solutions $z_I^{\pm}$ to $\hat{P}_I^2 = 0$, and all possible particle types $|\psi^{(I)}\rangle$ that can be exchanged in channel $I$. These recursion relations are called *soft subtracted recursion relations*. When $F = 1$, the recursion is called *unsubtracted*.

The expression for the solutions $z_I^{\pm}$ to the quadratic equation $\hat{P}_I^2 = 0$ involves square roots, but those must cancel since the tree amplitude is a rational function of the kinematic variables.

On channels where the amplitude factorizes into two *local* lower-point amplitudes (meaning that they have no poles), the cancellations of the square roots can be made manifest. This is done by a second application of Cauchy's theorem, which for each channel $I$ converts the sum of residues at $z = z_I^\pm$ to the sum of the residues at $z = 0$ and $z = 1/a_i$ for all $i$. Details are provided in Appendix A, here we simply state the result: *if $\mathcal{A}_L^{(I)}$ and $\mathcal{A}_R^{(I)}$ are local for all factorization channels*, the soft recursion relations take the form

$$\mathcal{A}_n = \sum_I \sum_{|\psi^{(I)}\rangle} \left( \frac{\hat{\mathcal{A}}_L^{(I)}(0)\hat{\mathcal{A}}_R^{(I)}(0)}{P_I^2} + \sum_{i=1}^n \text{Res}_{z=\frac{1}{a_i}} \frac{\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)}{z\, F(z)\, \hat{P}_I^2} \right). \tag{3.2.6}$$

Note that this form of the recursion relation is typically only valid at low points since it requires that the amplitude factorizes into a form where all subamplitudes are local. The recursion relation in the form (3.2.6) is manifestly rational in the kinematic variables, and we will be using (3.2.6) for the applications in this chapter. Note that only the first term in (3.2.6) has poles. Therefore the sum of the $1/a_i$ residues over all channels must be a local polynomial in the momenta.

### 3.2.2 Validity Criterion

The purpose of including $F(z)$ in (3.2.2) is to improve the large-$z$ behavior of the integrand so that one can avoid a pole at $z = \infty$. This is necessary in EFTs, where the large-$z$ behavior of the amplitude typically does not allow for unsubtracted recursion relations with $F(z) = 1$ to be valid without a boundary term from $z = \infty$. A sufficient condition for absence of a simple pole at infinity is that the deformed amplitude vanishes as $z \to \infty$. Below we show that for a theory with a single fundamental interaction of valence $v$ and coupling of mass-dimension $[g_v]$ the criterion for validity of the subtracted recursion relations is

$$\boxed{4 - n - \frac{n-2}{v-2}\,[g_v] - \sum_{i=1}^n s_i - \sum_{i=1}^n \sigma_i < 0\,.} \tag{3.2.7}$$

Here $s_i$ is the spin (not helicity) of particle $i$ and $\sigma_i$ is its soft behavior (3.4.5). Alternatively, one can write the constructibility criterion in terms of the reduced dimension $\tilde{\Delta}$ as

$$\boxed{4 - n + (n-2)\tilde{\Delta} - \sum_{i=1}^n s_i - \sum_{i=1}^n \sigma_i < 0\,.} \tag{3.2.8}$$

The criterion generalizes to theories with more than one fundamental coupling by replacing $\frac{n-2}{v-2}\,[g_v]$ in (3.2.7) by the sum over all couplings contributing to the diagrammatic expansion of the amplitude in question; the precise criterion is given in (3.2.16).

**Proof of the criterion (3.2.7)**

To avoid a pole at infinity in the Cauchy integral (3.2.2), it is sufficient to require $\hat{\mathcal{A}}_n(z)/F(z) \to 0$ as $z \to \infty$. To start with, we determine the large-$z$ behavior of the deformed amplitude $\hat{\mathcal{A}}_n(z)$.

Generically, in a theory of massless particles with couplings $g_k$, a tree-level amplitude takes the form

$$\mathcal{A}_n = \sum_j \left( \prod_k g_k^{n_{jk}} \right) M_j \,, \tag{3.2.9}$$

where $\prod_k g_k^{n_{jk}}$ is a product of coupling constants and $M_j$ is a function of spinor brackets only. Since there can be no other dimensionful quantities entering $M_j$, the mass dimension $[M_j]$ can be determined via a homogenous scaling of all spinors:

$$|i\rangle \to \lambda^{1/2}|i\rangle \quad \text{and} \quad |i] \to \lambda^{1/2}|i] \qquad \Longrightarrow \qquad M_j \to \lambda^{[M_j]} M_j \,. \tag{3.2.10}$$

The mass dimension is also fixed by simple dimensional analysis to be

$$[M_j] = 4 - n - \sum_k n_{jk}[g_k] \,, \tag{3.2.11}$$

since an $n$-point scattering amplitude in 4d has to have mass-dimension $4 - n$.

It is useful to consider a modified scale transformation defined as

$$\begin{aligned} h_i \geq 0 &: \quad |i\rangle \to \lambda|i\rangle \,, \qquad |i] \to |i] \,, \\ h_i < 0 &: \quad |i\rangle \to |i\rangle \,, \qquad |i] \to \lambda|i] \,. \end{aligned} \tag{3.2.12}$$

The effect of this scaling can be obtained from the uniform scaling (3.2.10) via a little group transformation on all momenta with $t = \lambda^{1/2}$. Therefore under (3.2.12), $M_j$ scales as $M_j \to \lambda^{[M_j] - \sum_i s_i} M_j$, where $s_i$ is the spin (*not* helicity) of particle $i$.

For the case of a theory with a single fundamental interaction of valence $v$ with coupling $g_v$, the number of couplings appearing in an $n$-point amplitude is $\frac{n-2}{v-2}$, and therefore we have

$$\mathcal{A}_n \to \lambda^D \mathcal{A}_n \,, \quad D = 4 - n - \frac{n-2}{v-2}[g_v] - \sum_i s_i \tag{3.2.13}$$

under the modified scale transformation (3.2.12).

Under the momentum shift (3.2.4), the deformed tree amplitude $\hat{\mathcal{A}}_n(z)$ can be written

$$\begin{aligned} \hat{\mathcal{A}}_n(z) &= \hat{\mathcal{A}}_n\big( \dots \{(1 - a_i z)|i\rangle, |i]\}_+ \dots \{|j\rangle, (1 - a_j z)|j]\}_- \big) \\ &= \hat{\mathcal{A}}_n\big( \dots \{z(1/z - a_i)|i\rangle, |i]\}_+ \dots \{|j\rangle, z(1/z - a_j)|j]\}_- \big) \\ &= z^D \, \hat{\mathcal{A}}_n\big( \dots \{(1/z - a_i)|i\rangle, |i]\}_+ \dots \{|j\rangle, (1/z - a_j)|j]\}_- \big) \,, \end{aligned} \tag{3.2.14}$$

where the subscripts $\pm$ refer to the sign of the helicity of each particle. In the last line we used the behavior (3.2.13) under the modified scaling (3.2.12).

At large $z$, the amplitude in the last line of (3.2.14) is the original unshifted amplitude evaluated at a momentum configuration with $q_i = -a_i p_i$. These momenta are all on-shell and satisfy, via (3.2.1), momentum conservation. The only way the tree amplitude could have a singularity at this momentum configuration would be if an internal line went on-shell. This can always be avoided for generic momenta.[5] Thus we conclude from (3.2.14) that for large $z$, the deformed amplitude behaves as

$$\hat{\mathcal{A}}_n(z) \to z^N \quad \text{with} \quad N \leq D\,, \tag{3.2.15}$$

where $D$ is given in (3.2.13). The inequality allows for the possibility that $\mathcal{A}_n$ could have a zero at $q_i = -a_i p_i$.

Our mission was to find a criterion for $\hat{\mathcal{A}}_n(z)/F(z) \to 0$ as $z \to \infty$. By the definition (3.2.3), we have $F(z) \to z^{\sum_i \sigma_i}$ for large $z$. From our analysis of the large-$z$ behavior of $\hat{\mathcal{A}}_n(z)$, we can therefore conclude that, at worst, $\hat{\mathcal{A}}_n(z)/F(z) \to z^{D-\sum_i \sigma_i}$. The sufficient criterion for absence of a pole at infinity, and hence for validity of the subtracted recursion relation, is then $D - \sum_i \sigma_i < 0$. This is precisely the condition (3.2.7). This concludes the proof.

It is straightforward to generalize the constructibility criterion to EFTs with more than one fundamental interaction,

$$4 - n - \min_j\Big(\sum_k n_{jk}[g_k]\Big) - \sum_{i=1}^n s_i - \sum_{i=1}^n \sigma_i < 0\,. \tag{3.2.16}$$

Recall that in effective field theories, the couplings have negative mass-dimension. This means that the constructibility criterion tends to be dominated by the fundamental interactions associated with operators of the highest mass-dimension that can contribute to the $n$-point amplitude.

**Example 1**

Let us once again return to the example of DBI. The action has a fundamental quartic vertex $g_4(\partial\phi)^4$ with a coupling of mass-dimension $[g_4] = -4$. The constructibility criterion (3.2.7) for the $n$-scalar amplitude is $n(1-\sigma_S) < 0$, where $\sigma_S$ is the soft behavior of the scalar $\phi$. Since $\sigma_S = 2$ in DBI, all DBI tree amplitudes are constructible via the subtracted soft recursion relations, as claimed in the introduction.

The failure of the constructibility criterion for $\sigma_S = 1$ is simply the statement that an EFT whose interactions are built from powers of $(\partial\phi)^2$ trivially has a constant shift symmetry and hence

[5] The condition (3.2.1) has a trivial solution with all $a_i$ equal. Therefore any solution to (3.2.1) can be shifted uniformly $a_i \to a_i + a$ for any real number $a$. Hence, we can always avoid the discrete set of momentum configurations for which an internal line in $\mathcal{A}_n$ goes on-shell.

$\sigma_S = 1$, so there are no constraints from shift symmetry on the coefficients of $(\partial\phi)^{2k}$ in terms of that of $(\partial\phi)^4$ and then one has no chance of recursing $\mathcal{A}_4$ to get all-point amplitudes.

**Example 2**

Consider a theory of massless fermions with quartic coupling of mass-dimension $[g_4] = -2$. The criterion (3.2.7) says that the $n$-fermion amplitudes are constructible when $4 < n(1 + 2\sigma_\psi)$. Thus all $n > 4$ point tree-amplitudes are constructible by (3.2.5) for any soft weight $\sigma_\psi \geq 0$. No such theory exists for $\sigma_\psi > 0$ (as we prove in Section 3.3.2), but for $\sigma_\psi = 0$ this is exactly the Nambu–Jona-Lasinio (NJL) model, which consists of the simple 4-fermion interaction $\psi^2\bar{\psi}^2$ [82].

### 3.2.3 Non-Constructibility = Triviality

We have derived a constructibility criterion, but what does it mean? The answer is quite simple: if an $n$-point amplitude can be constructed recursively from lower-point on-shell amplitudes, there cannot exist a local gauge-invariant $n$-field operator that contributes to the amplitude without modifying its soft behavior. We define a *trivial operator* to be one with at least 4 fields whose matrix elements manifestly have a given soft weight $\sigma$. Let us now assess what it takes to make an operator of scalar, fermion, and vector fields trivial.

**Triviality.**

*Scalars.* Operators with at least $m$ derivatives on each scalar field will trivially have single-soft scalar limits with $\sigma_S = m$.

*Fermions.* We have chosen the soft limit (3.4.5) according to the helicity such that the fermion wavefunctions do not generate any soft factors of $\epsilon$. Thus a trivial soft behavior must come from derivatives on each fermion field in the Lagrangian. We conclude that the trivial soft behavior $\sigma_F =$ smallest number of derivatives on each fermion field.

*Photons.* Gauge invariance tells us that we should construct the interaction terms using the field strength $F_{\mu\nu}$.[6] When associated with an external photon, the Feynman rule for $F_{\mu\nu}$ gives $p_\mu\epsilon_\nu - p_\nu\epsilon_\mu$. Naively, it may seem to be linear in the soft momentum, but under the holomorphic soft shift (3.2.4) it is actually $\mathcal{O}(\epsilon^0)$. Recall that in spinor helicity formalism, a positive helicity vector polarization takes the form $\epsilon_+^\mu \bar{\sigma}_\mu^{\dot{a}b} = \epsilon_+^{\dot{a}b} = |q\rangle^{\dot{a}}[p|^b/\langle pq\rangle$, where $q$ is a reference spinor. Hence, for a positive helicity photon we have

$$(F_+)_a{}^b \equiv (\sigma^{\mu\nu})_a{}^b F_{\mu\nu} \longrightarrow (\sigma^{\mu\nu})_a{}^b(p_\mu\epsilon_{+\nu} - p_\nu\epsilon_{+\mu}) \sim |p]_a\langle p|_{\dot{c}}\frac{|q\rangle^{\dot{c}}[p|^b}{\langle pq\rangle} = |p]_a[p|^b\,. \qquad (3.2.17)$$

This is explicitly independent of the reference spinor $q$ because $F_{\mu\nu}$ is gauge invariant. For a

[6] Or covariant derivatives $D_\mu = \partial_\mu + igA_\mu$. In this chapter, we focus on scalars and fermions that do not transform under any gauge-$U(1)$, therefore photons must couple via $F_{\mu\nu}$.

positive helicity particle, we take the soft limit holomorphically as $|p\rangle \to \epsilon|p\rangle$ (while $|p] \to |p]$), so we explicitly see that $F_{\mu\nu} \longrightarrow |p][p|$ is $\mathcal{O}(\epsilon^0)$ when $p$ is taken soft. Likewise, for a negative helicity photon, $(F_-)^{\dot{a}}{}_{\dot{b}} \longrightarrow |p\rangle\langle p|$. We conclude that an operator with photons has trivial soft behavior that is determined by the smallest number of derivatives on each field strength $F_{\mu\nu}$.

In an EFT where photon interactions are built only from the field strengths, the matrix elements are $\mathcal{O}(1)$ when a photon is taken soft. This, for example, is exactly the case for Born-Infeld theory in which the photons have $\sigma = 0$.

**Constructibility.** Suppose we study an $n$-particle amplitude with $n_s$ scalars, $n_f$ fermions, and $n_\gamma$ photons in an EFT whose fundamental $v$-particle interactions all have couplings of the same mass-dimension $[g_v]$. The criterion (3.2.7) for constructibility via subtracted soft recursion relations can be written as

$$4 - n - n_v[g_v] - \frac{1}{2}n_f - n_\gamma - n_s\sigma_s - n_f\sigma_f - n_\gamma\sigma_\gamma < 0\,, \tag{3.2.18}$$

where $n_v = (n-2)/(v-2)$ is the number of vertices needed at $n$-point.

**Non-constructibility = Triviality.** Let us assess if there can be a local contact term for an $n$-particle amplitude with $n_s$ scalars, $n_f$ fermions, and $n_\gamma$ photons and soft behaviors $\sigma_s$, $\sigma_f$, and $\sigma_\gamma$, respectively. As discussed above, a contact term that has such trivial soft behavior takes the form

$$g_n \underbrace{(\partial^{\sigma_s}\phi)\cdots(\partial^{\sigma_s}\phi)}_{n_s}\underbrace{(\partial^{\sigma_f}\psi)\cdots(\partial^{\sigma_f}\psi)}_{n_f}\underbrace{(\partial^{\sigma_\gamma}F)\cdots(\partial^{\sigma_\gamma}F)}_{n_\gamma} \tag{3.2.19}$$

(for brevity we have not distinguished $\psi$ and $\bar{\psi}$). In 4d, the mass-dimension of the coupling $g_n$ is easily computed as

$$[g_n] = 4 - \big(n_s + n_s\sigma_s\big) - \big(\tfrac{3}{2}n_f + n_f\sigma_f\big) - \big(2n_\gamma + n_\gamma\sigma_\gamma\big)\,. \tag{3.2.20}$$

Using $n = n_s + n_f + n_\gamma$, we can rewrite this as

$$4 - n - [g_n] - \frac{1}{2}n_f - n_\gamma - n_s\sigma_s - n_f\sigma_f - n_\gamma\sigma_\gamma = 0\,. \tag{3.2.21}$$

Compare this with (3.2.18); we note that the constructibility criterion is simply that $n_v[g_v] > [g_n]$, or maybe more intuitively, that $g_n$ has more negative mass-dimension than $n_v$ $g_v$-vertices. So, when constructibility holds, the $n$-particle amplitude constructed from the $n_v$ $v$-valent vertices *cannot* be influenced by a contact term that trivially has the soft behavior: such a contact term would be too high order in the EFT due to all the derivatives needed to trivialize the soft behavior. That of course makes sense; were there such an independent local contact term, it could be added to the result of recursion with any coefficient without changing any of the properties of the amplitude. Hence recursion cannot possibly work in that case. (This is analogous to the example in [71, 83] for constructibility in scalar-QED via BCFW; the difference here is that the subtracted soft recursion relations "know" about the soft behavior in addition to gauge-invariance.)

The argument is easily extended to the case where the theory has fundamental vertices of different valences and mass-dimensions. We conclude that the constructibility criterion (3.2.7) is equivalent to the non-existence of local $n$-particle operators with couplings of the same mass-dimension and trivial soft behavior: Non-constructibility = Triviality.

### 3.2.4 Implementation of the Subtracted Recursion Relations

Here we present details relevant for the practical implementation of the soft subtracted recursion relations.

**Solving the shift constraints.** Conservation of the momentum for the shifted momenta $\hat{p}_i$ (3.2.1) requires the shift variables $a_i$ to satisfy

$$\sum_i a_i p_i^\mu = 0. \tag{3.2.22}$$

In 4d, the LHS can be viewed as a $4\times n$ matrix $p_i^\mu$ of rank 4 (if $n \geq 5$) multiplying an $n$-component vector $a_i$. Hence the valid choices of parameters $a_i$ form a vector space given by the kernel of the matrix $p_i^\mu$. For $n \geq 5$ any subset of four momenta are generically linearly independent, so the $p_i^\mu$-matrix has full rank. By the rank-nullity theorem, the dimension of the kernel is therefore $n-4$. However, there is always a trivial solution which consists of all $a_i$'s equal, hence non-trivial solutions to (3.2.22) exist only when $n \geq 6$.

Practically, the linear system of equations is solved by dotting in $p_j$, i.e. we have

$$\sum_i s_{ji}\, a_i = 0 \quad \text{for} \quad j = 1, 2, \ldots, n\,. \tag{3.2.23}$$

The symmetric $n \times n$-matrix with entries $s_{ji}$ has rank 4, so the linear system (3.2.23) can be solved for say $a_1$, $a_2$, $a_3$, and $a_4$ in terms of the $n-4$ other $a_i$'s.

**Soft bootstrap.** Subtracted recursion relations can be used to calculate tree amplitudes in EFTs of Goldstone modes in theories we already know well, such as DBI, Akulov-Volkov etc. However, the soft subtracted recursion relations can also be used as a tool to *classify* and assess the existence of EFTs with a given spectrum of massless particles and low-energy theorems with given weights $\sigma$.

The approach to the classification of EFTs is as follows:

(1) Model input: the spectrum of massless particles and the coupling dimensions of the fundamental interactions in the model.

(2) Symmetry assumptions: the $n$-particle amplitudes have soft behavior with weight $\sigma_i$ for the $i$th particle.

If the constructibility criterion (3.2.7) is *not* satisfied, the assumptions (1) and (2) are trivially satisfied and we cannot constrain the couplings in the EFTs.

If the constructibility criterion (3.2.7) *is* satisfied for input (1) and (2), one can use the soft subtracted recursion relations to test whether a theory can exist with the above assumptions. One proceeds as follows.

The fundamental vertices give rise to local amplitudes which must be polynomials[7] in the spinor helicity brackets, and it is simple to construct the most general such ansatz for the local input amplitudes. One can further restrict this ansatz by imposing on it the soft behaviors associated with the assumed symmetries. The result of recursing this input from the fundamental vertices is supposed to be a physical amplitude and therefore it must necessarily be independent of the $n-4$ parameters $a_i$ that are unfixed by (3.2.22). If that is *not* the case for any ansatz of the fundamental input amplitudes (vertices), we learn that there cannot exist a theory with the properties (1) and (2) above. On the other hand, an $a_i$-independent result is evidence (but not proof) of the existence of such a theory. It may well be that $a_i$-independence requires some of the free parameters in the input amplitudes to be fixed in certain ways and this can teach us important lessons about the underlying theory. The test of $a_i$-independence can be done efficiently numerically, and this way one can scan through theory-space to test which symmetries are compatible with a given model input.

Additionally, one can impose further constraints from unbroken global symmetries, for example, one can restrict the input from the fundamental amplitudes by imposing the supersymmetry Ward identities. We shall see examples of this in later sections.

**4d and 3d consistency checks.** There is a subtlety that must be addressed for $n = 6$. In that case, the solution space is 2-dimensional, but one solution is the trivial one with all $a_i$ equal. Furthermore, one can rescale all $a_i$. This means that if the recursed result for the amplitude depends on the $a_i$ only through ratios of the form

$$\frac{(a_i - a_j)}{(a_k - a_l)}\,, \tag{3.2.24}$$

it will appear to be $a_i$-independent numerically, but the result will nonetheless have spurious poles. To detect this problem numerically, we dimensionally reduce the recursed result to 3d.[8]. Then the space of solutions to (3.2.22) is $(n-3)$-dimensional, so there are non-trivial solutions

[7] This is true at 4-point and higher; for 3-point, massless particle amplitudes are uniquely fixed by the little group scaling.

[8] The dimensional reduction from 4d to 3d is carried out by simply replacing all square spinors by angle spinors.

and a numerical 3d test will reveal dependence on ratios such as (3.2.24) for $n = 6$.

We refer to the consistency checks of $a_i$-independence as *4d and 3d consistency checks*, respectively, or simply as *$n$-point tests* when applied to construction of $n$-point amplitudes. In this chapter, we use 6-, 7- and 8-point tests. In Section 3.3, we present an overview of the resulting space of exceptional pure real and complex scalar, fermion, and vector EFTs.

**Special requirements for non-trivial 5-point interactions.** Consider 5-particle interactions which are non-trivial with respect to a given soft behavior. This could for example be the Wess-Zumino-Witten (WZW) term, which with 4 derivatives on 5 scalars has a non-trivial $\sigma = 1$ soft behavior. Or the 5-point Galileon, which with 8 derivatives on 5 scalars has a non-trivial $\sigma = 2$. Constructibility tells us that one must be able to calculate such 5-point amplitudes from soft recursion relations via factorization, i.e.

$$\mathcal{A}_5 = \sum_I \frac{\hat{\mathcal{A}}_3 \hat{\mathcal{A}}_4}{P_I^2} \,. \tag{3.2.25}$$

However, there are no 3-point amplitudes available that could possibly make this work. The reason is that the only 3-scalar interaction with a non-zero on-shell amplitude is $\phi^3$, which gives rise to amplitudes with $\sigma = -1$ [51]. So we appear to have a contradiction: the constructibility criterion tells us that these 5-particle amplitudes are recursively constructible, but it is obviously impossible to construct them from lower-point input.

What goes wrong is that at 5-points, there are no non-trivial choices of the $a_i$ parameters that give valid recursion relations in 4d. So we have to go to 3d kinematics to resolve this issue. The above contradiction persists in 3d, so the only resolution is that these non-trivial constructible 5-point amplitudes must vanish in 3d kinematics.

Indeed they do: for WZW term and the quintic Galileon, the 5-point matrix elements are

$$A_5^{\rm WZW} = g_5 \, \epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu p_3^\rho p_4^\sigma \,, \qquad A_5^{\rm Gal} = g_5' \left(\epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu p_3^\rho p_4^\sigma\right)^2 . \tag{3.2.26}$$

The Levi-Civita contraction makes it manifest that these amplitudes vanish in 3d.

We conclude that *any* non-trivial (in the sense of soft behavior) 5-particle interaction must vanish in 3d. Thus, it is no coincidence that the WZW and quintic Galileon 5-point amplitudes are proportional to Levi-Civita contractions.

## 3.3 Soft Bootstrap

We now turn to examples of how the soft recursion relations can be used to examine the existence of EFTs with assumed low-energy theorems. The landscape of real scalar theories was previously studied in [38, 76, 79, 84]. We outline it briefly below for completeness, but otherwise focus on new results, in particular for complex scalars, fermions, and vectors. This section considers only theories with one kind of massless particle. One can of course also couple scalars, fermions, and vectors in EFTs, and this is discussed in Sections 3.5 and 3.6.

### 3.3.1 Pure Scalar EFTs

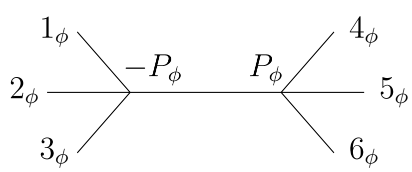
Consider an EFT with a single real scalar field $\phi$. There can only be non-vanishing 3-point amplitudes in $\phi^3$-theory and this gives amplitudes with soft weight $\sigma = -1$. Focusing on EFTs with soft weights $\sigma \geq 0$, the lowest-point amplitude is 4-point.

The on-shell factorization diagrams that contribute in the recursion relations (3.2.6) for

$$\mathcal{A}_6(1_\phi\, 2_\phi\, 3_\phi\, 4_\phi\, 5_\phi\, 6_\phi) \tag{3.3.1}$$

are composed of a product of two 4-point amplitudes, for example the 123-channel diagram is

$\mathcal{A}_6^{(123)} =$




$$= \frac{\hat{\mathcal{A}}_L(0)\hat{\mathcal{A}}_R(0)}{P_{123}^2} + \sum_{i=1}^{6} \text{Res}_{z=\frac{1}{a_i}} \frac{\hat{\mathcal{A}}_L(z)\hat{\mathcal{A}}_R(z)}{z\, F(z)\, \hat{P}_{123}^2}\,,$$

where $\hat{\mathcal{A}}_L = \hat{\mathcal{A}}_4(1_\phi\, 2_\phi\, 3_\phi\, -P_\phi)$ and $\hat{\mathcal{A}}_R = \hat{\mathcal{A}}_4(P_\phi\, 4_\phi\, 5_\phi\, 6_\phi)$.[9] One sums over the 10 independent permutations corresponding to the 10 distinct factorization channels.[10]

For complex scalars, we assume that the input 4-point amplitudes are of the form $\mathcal{A}_4(1_Z\, 2_{\bar{Z}}\, 3_Z\, 4_{\bar{Z}})$;[11] one can also consider more general input but it would not be compatible with supersymmetry, so in the present chapter we do not discuss such options. At 6-point, there is only one type of amplitude that can arise from such 4-point input via recursion, and that is $\mathcal{A}_6(1_Z\, 2_{\bar{Z}}\, 3_Z\, 4_{\bar{Z}}\, 5_Z\, 6_{\bar{Z}})$. The

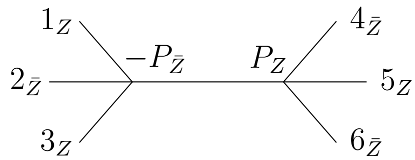
[9]The momenta in the hatted amplitudes are shifted; for simplicity, we do not write the hats on the momentum variables explicitly. Note that in particular $P_\phi$ should really be understood as $\hat{P}_\phi$ with $\hat{P}_\phi^2 = 0$.

[10]We do not consider color-ordering in this section. With color-ordering, one only includes the factorization diagrams from cyclic permutations of the external lines.

[11]There is no color-ordering implied in any of the amplitudes here. We simply alternate $Z$ and $\bar{Z}$ states as odd/even numbered momentum lines. In later sections, other helicity states are grouped similarly, in particular for supersymmetric cases, states that belong to the positive helicity sector sit on odd-numbered lines and negative helicity sector states on even-numbered lines. This is convenient for the practical implementation but should not be misunderstood as an indication of color-ordering.

123-channel diagram is

$$\mathcal{A}_6^{(123)} = \qquad (3.3.2)$$




To get the full amplitude, one must sum over all factorization channels:

$$\mathcal{A}_6(1_Z\, 2_{\bar{Z}}\, 3_Z\, 4_{\bar{Z}}\, 5_Z\, 6_{\bar{Z}}) = \Big(\mathcal{A}_6^{(123)} + (2 \leftrightarrow 4) + (2 \leftrightarrow 6)\Big) + (1 \leftrightarrow 5) + (3 \leftrightarrow 5)\,. \qquad (3.3.3)$$

In the following we consider real and complex scalar theories with 4- and 5-point fundamental vertices.

#### 3.3.1.1 Fundamental 4-point Interactions

Consider a theory of a single real scalar with fundamental 4-point interactions. We parameterize $\mathcal{A}_4^{\text{ansatz}}$ as the most general polynomial in the Mandelstam variables $s, t, u$ (with $s + t + u = 0$) and full Bose symmetry. We subject the recursed result for $\mathcal{A}_6$ to the test of $a_i$-independence, as described in Section 3.2.4. The result is

$$\partial^{2m}\phi^4 \qquad (3.3.4)$$

| - [g] | $m$ | $\mathcal{A}_4^{\text{ansatz}}(1_\phi\, 2_\phi\, 3_\phi\, 4_\phi)$ | $\sigma = 0$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| 0 | 0 | $g$ | $\phi^4$-theory | F | F | F | F |
| 2 | 1 | 0 | – | F | F | F | F |
| 4 | 2 | $g(s^2 + t^2 + u^2)$ | – | – | DBI | F | F |
| 6 | 3 | $g\, stu$ | – | – | Gal$_4$ | Spec Gal$_4$ | F |
| 8 | 4 | $g(s^4 + t^4 + u^4)$ | – | – | – | F | F |

In the table, we list the coupling dimension $[g]$ of the fundamental quartic couplings along with the most general ansatz for the corresponding 4-point amplitude. The dash, –, indicates that the constructibility criterion (3.2.7) fails; this means "triviality" in the sense described in Section 3.2.3). "F" indicates that the soft recursion fails to give an $a_i$-independent result, and hence no such theory can exist with the given assumptions. When a case passes the 6-point test, we are able to uniquely identify which theory it is. In the above table, the non-trivial theories that pass the 6-point test are: $\phi^4$-theory, DBI, and the quartic Galileon. The latter automatically has $\sigma = 3$ (which is called the Special Galileon) and passes 6-point test for both $\sigma = 2$ and $\sigma = 3$.

The analysis for complex scalars proceeds similarly and the results are

$$\partial^{2m} Z^2 \bar{Z}^2 \tag{3.3.5}$$

| - [g] | $m$ | $\mathcal{A}_4^{\text{ansatz}}(1_Z, 2_{\bar{Z}}, 3_Z, 4_{\bar{Z}})$ | $\sigma = 0$ | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| 0 | 0 | $g$ | $\lvert Z\rvert^4$-theory | F | F | F |
| 2 | 1 | $gt$ | – | $\mathbb{CP}^1$ NLSM | F | F |
| 4 | 2 | $gt^2 + g'su$ | – | – | $g' = 0$ cmplx DBI | F |
| 6 | 3 | $gt^3 + g'stu$ | – | – | $g = 0$ cmplx Gal$_4$ | F |
| 8 | 4 | $gt^4 + g't^2su + g''s^2u^2$ | – | – | – | F |

The non-trivial theories are $|Z|^4$-theory, the $\mathbb{CP}^1$ NLSM (which is studied in further detail in Section 3.5), and the complex scalar versions of DBI and the quartic Galileon. Note that there does not exist a complex scalar version of the Special Galileon with $\sigma = 3$. The results for the 6-point amplitudes of each of the theories with $\sigma > 0$ can be found in Appendix B.

#### 3.3.1.2 Fundamental 5-point Interactions

At 5-point, the input amplitudes are constructed as polynomials of Mandelstam variables $s_{ij}$ and Levi-Civita contractions of momenta. They must obey (1) momentum conservation, (2) Bose symmetry, and (3) assumed soft behavior $\sigma$. In many cases, these constraints on the 5-point input amplitudes are sufficient to rule out such theories (assuming no other interactions) without even applying soft recursion.

As discussed at the end of Section 3.2.4, non-trivial 5-point amplitudes must vanish in 3d kinematics, so they are naturally written using the Levi-Civita tensor, as in the two cases of WZW and the quintic Galileon (3.2.26).

We can summarize the results in the following:

- **1 real scalar.** There are only two non-trivial theories based on a fundamental 5-point interaction, namely $\phi^5$-theory, which has $[g_5] = -1$ and $\sigma = 0$, and the quintic Galileon, which has $[g_5] = -9$ and $\sigma = 2$.

- **1 complex scalar.** We assume input amplitudes of the form $\mathcal{A}_5(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}} 5_Z)$. Two cases pass the 8-point test:

  The quintic $g_5(Z^3\bar{Z}^2 + Z^2\bar{Z}^3)$-theory with $[g_5] = -1$ has $\sigma_Z = 0$.

  The complex-scalar version of the quintic Galileon with $[g_5] = -9$ and $\sigma_Z = 2$. The 5-point amplitude is

$$\mathcal{A}_5(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}} 5_Z) = g_5(\epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu p_3^\rho p_4^\sigma)^2 \,, \tag{3.3.6}$$

same as for the real-scalar quintic Galileon. The fact that it passes the 8-point test is somewhat trivial: because of the two explicit factors of momentum for 4 out of 5 particles, the residues at $1/a_i$ vanish identically for each factorization channel. The same is true for the real Galileon, so the 8-point test is not really effective as an indicator of whether such a theory may exist.

Suppose the putative complex-scalar quintic Galileon is coupled to the complex scalar DBI. Then we can conduct a 7-point test based on factorization into a quantic Galileon and a quartic DBI subamplitude. The test of $a_i$-independence requires the coupling constant $g_5$ to vanish. This means that the DBI-Galileon with a complex scalar cannot have a 5-point interaction.

At $[g_5] = -9$, there is a 6-parameter family of 5-point amplitudes with $\sigma_Z = 1$. The EFT with such amplitudes is generally non-constructible. However, a 1-parameter sub-family is compatible with the constraints of supersymmetry. As discussed in [52] and further in Section 3.6.1 this may be a candidate for a supersymmetric quintic Galileon with a limited sector of constructible amplitudes.

### 3.3.2 Pure Fermion EFTs

Let us now consider EFTs with only fermions and fundamental interactions of the form $\partial^{2m}\psi^2\bar{\psi}^2$. This is not the only choice, but it is the option compatible with supersymmetry. Moreover, we have found that couplings of "helicity violating" 4-point interactions in the fermion sector must vanish by the 6-point test in all pure-fermion cases we tested. The calculations proceed much the same way as for scalars, except that one must be more careful with signs when inserting fermionic states on the internal line. The diagrams needed for the recursive calculation of the 6-fermion amplitude $A_6(1^+_\psi\, 2^-_\psi\, 3^+_\psi\, 4^-_\psi\, 5^+_\psi\, 6^-_\psi)$ are just like those in the scalar case (3.3.2), but now the permutations have to be taken with a sign:

$$\mathcal{A}_6(1^+_\psi\, 2^-_\psi\, 3^+_\psi\, 4^-_\psi\, 5^+_\psi\, 6^-_\psi) = \Big(\mathcal{A}_6^{(123)} - (1 \leftrightarrow 5) - (3 \leftrightarrow 5)\Big) - (2 \leftrightarrow 4) - (2 \leftrightarrow 6). \tag{3.3.7}$$

The input 4-point amplitudes $\mathcal{A}_4(1^+_\psi\, 2^-_\psi\, 3^+_\psi\, 4^-_\psi)$ are fixed by little group scaling to be $\langle 24\rangle[13]$ times a Mandelstam polynomial of degree $m-1$ that must be symmetric under $s \leftrightarrow u$ to ensure Fermi antisymmetry for identical fermions. The most general input amplitudes for low values of

$m$ are summarized in the table below that also shows the result of the recursive 6-point test:

$$\partial^{2m}\psi^2\bar{\psi}^2 \tag{3.3.8}$$

| - [g] | $m$ | $\mathcal{A}_4(1^+_\psi\, 2^-_\psi\, 3^+_\psi\, 4^-_\psi) = \langle 24\rangle[13]\times$ | $\sigma = 0$ | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| 2 | 0 | $g$ | NJL | F | F | F |
| 4 | 1 | $gt$ | – | A-V | F | F |
| 6 | 2 | $gt^2 + g'su$ | – | – | F | F |
| 8 | 3 | $gt^3 + g'stu$ | – | – | $g = 0$ new | F |

We comment briefly on these results:

- The **NJL model** has the fundamental 4-fermion interaction $\bar{\psi}^2\psi^2$ and the result of recursing it to 6-point is given in Appendix B.1. The relevance of this model will for our purposes be as part of the supersymmetrization of the NLSM (see Section 3.5).

- **Akulov-Volkov theory** of Goldstinos is the only non-trivial EFT with coupling of mass-dimension $-4$. The Goldstinos in this theory have low-energy theorems with $\sigma = 1$.

- There are no constructible purely fermionic EFTs with fundamental quartic coupling $[g_4] = -6$. Nonetheless, as was shown in [52], the quartic Galileon has a supersymmetrization with a 4-fermion fundamental interaction, however, the fermion has $\sigma = 1$, so the all-fermion amplitudes in that theory are not constructible by soft recursion: one needs additional input from supersymmetry. We refer the reader to [52] and present some further details in Section 3.6.1.

- For $[g] = -8$ and $\sigma = 2$, the 6-point numerical test is passed in 4d kinematics without constraints on $g$ and $g'$; that is because the recursed result depends only on ratios (3.2.24). When the 3d consistency check is employed, we learn that we must set $g = 0$ to ensure $a_i$-independence. (This is not a strong test since the particular form of the interaction, $stu$, ensures that all $1/a_i$-poles cancel in each factorization individual diagram.) Hence, the theory that passes the 6-point test with $\sigma = 2$ has $\mathcal{A}_4(1_\psi, 2_{\bar{\psi}}, 3_\psi, 4_{\bar{\psi}}) = g'\langle 24\rangle[13]stu$. The subtracted recursion relations fail at $n > 6$, which means that at 8-point and higher, this model is not uniquely determined by its symmetries. The Lagrangian construction of this theory has been studied as a fermionic generalization of the scalar Galileon [85].

### 3.3.3 Pure Vector EFTs

Pure Abelian vector EFTs consist of interaction terms built from $F_{\mu\nu}$-contractions, possibly dressed with extra derivatives. In 4d, the Cayley-Hamilton relations imply that theories built from just field strengths $F_{\mu\nu}$ can be constructed from two types of index-contractions, namely (see for example [1])

$$f = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad \text{and} \quad g = -\frac{1}{4}F_{\mu\nu}\tilde{F}^{\mu\nu}\,, \tag{3.3.9}$$

where $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. If one assumes parity, the Lagrangian can only contain even powers of $g$. One can then write an ansatz for the Lagrangian as

$$\mathcal{L} = f + \frac{b_1}{\Lambda^4}f^2 + \frac{b_2}{\Lambda^4}g^2 + \frac{b_3}{\Lambda^8}f^3 + \frac{b_4}{\Lambda^8}fg^2 + \ldots \tag{3.3.10}$$

As established in Section 3.2.3, a model with photon interactions built of $F_{\mu\nu}$-contractions only have soft behavior $\sigma = 0$. The simplest 4-photon interactions may naively look like the vector equivalent of the constructible $\phi^4$ scalar EFT. However, that is not the case. For the scalar, the 6-particle operator $\frac{1}{\Lambda^2}\phi^6$ is subleading to the pole contributions with two $\phi^4$-vertices. However, for photons the pole terms with two $\frac{1}{\Lambda^4}F^4$-vertices are exactly the same order as $\frac{1}{\Lambda^8}F^6$. Therefore amplitudes in a theory with $F^n$ interactions and $\sigma = 0$ are non-constructible, in other words it is trivial to have $\sigma = 0$ for any choice of coefficients $b_i$. One may ask if it is possible to choose the parameters $b_i$ in (3.3.10) such that the amplitudes have enhanced soft behavior $\sigma > 0$. The 6-point soft recursive test shows that this is impossible, i.e. no models exist with Lagrangians of the form (3.3.10) and $\sigma > 0$.

Nonetheless, the class of theories with pure $F^n$-interactions do include one particularly interesting case, namely Born-Infeld (BI) theory. The BI Lagrangian can be written in 4d as

$$\mathcal{L}_{\text{BI}} = \Lambda^4\left(1 - \sqrt{-\det\left(\eta_{\mu\nu} + F_{\mu\nu}/\Lambda^2\right)}\right). \tag{3.3.11}$$

Upon expansion, the Lagrangian will take the form (3.3.10) with some particular coefficients $b_i$. As noted, those particular coefficients do not change the single-soft behavior of amplitudes, the BI photon also has $\sigma = 0$. Nonetheless, BI theory does have the distinguishing feature of being the vector part of a supersymmetric EFT. In particular, $\mathcal{N} = 1$ supersymmetric Born-Infeld theory couples the BI vector to a Goldstino mode whose self-interactions are described by the Akulov-Volkov action. One can also view Born-Infeld as the vector part of the $\mathcal{N} = 2$ or $\mathcal{N} = 4$ supersymmetrization of DBI. It was argued recently [1] that supersymmetry ensures BI amplitudes to vanish in certain multi-soft limits. Based on that, the BI amplitudes can be calculated unambiguously using on-shell techniques [1].

Next, one can consider EFTs in which the field strengths are dressed with derivatives, for example

$$\mathcal{L} = -\frac{1}{4}F^2 + \frac{c_1}{\Lambda^6}\partial^2F^4 + \frac{c_1}{\Lambda^{12}}\partial^4F^6 + \ldots \tag{3.3.12}$$

Theories with fundamental 4-point interactions are non-constructible for $\sigma = 0$ and fail the soft recursion $a_i$-independence 6-point test for $\sigma > 0$. One implication of this is that there can be no vector Goldstone bosons with vanishing low-energy theorems. This conclusion was also reached in [36], but from a very different algebraically-based analysis. A second implication is that the pure vector sector of an $\mathcal{N} \geq 2$ Galileon model is non-constructible with the basic soft recursion, and other properties (such as supersymmetry) have to be specified in order to determine those amplitudes recursively.

There are other interesting vector EFTs: we study in detail the $\mathcal{N} = 2$ supersymmetric NLSM in Section 3.5. Furthermore, massive gravity [86–88] motivates the existence of a vector-scalar theory coupling Galileons to a vector field; we explore this in Section 3.6.4.

## 3.4 Soft Limits and Supersymmetry

For models with unbroken supersymmetry, the on-shell amplitudes satisfy a set of linear relations known as the *supersymmetry Ward identities* [89, 90]. (For recent reviews and results, see [71, 83, 91].) In this section, we use $\mathcal{N} = 1$ supersymmetry to derive general consequences for the soft behavior for massless particles in the same supermultiplet. It is not assumed that these particles are Goldstone or quasi-Goldstone modes; the results apply to all $\mathcal{N} = 1$ supermultiplets of massless particles. The consequences for extended supersymmetry are directly inferred from the $\mathcal{N} = 1$ constraints.

### 3.4.1 $\mathcal{N} = 1$ Supersymmetry Ward Identities

We consider $\mathcal{N} = 1$ *chiral* and *vector* supermultiplets. We use the following shorthand for the action of the supercharges on individual particles with momentum label $i$: for chiral multiplets

| state $i$ | $\mathcal{Q} \cdot i$ | $\mathcal{A}_n$ prefactor | $\mathcal{Q}^\dagger \cdot i$ | $\mathcal{A}_n$ prefactor |
|---|---|---|---|---|
| $\psi^+$ | $Z$ | $\vert i]$ | $0$ | $0$ |
| $Z$ | $0$ | $0$ | $\psi^+$ | $-\vert i\rangle$ |
| $\overline{Z}$ | $\psi^-$ | $\vert i]$ | $0$ | $0$ |
| $\psi^-$ | $0$ | $0$ | $\overline{Z}$ | $-\vert i\rangle$ |

(3.4.1)

where $Z$ is a complex scalar and $\psi$ is a Weyl fermion. The superscripts $\pm$ refer to the helicity of the particle. $\mathcal{Q}^\dagger$ raises helicity by $1/2$ while $\mathcal{Q}$ lowers it by $1/2$. The prefactor is what goes outside the amplitude when the supercharge acts on it, e.g.

$$\begin{aligned}\mathcal{Q} \cdot \mathcal{A}_n\big(1_Z\, 2_\psi^+\, 3_\psi^+\, 4_{\overline{Z}} \ldots\big) = \ & 0 \ + \ |2]\mathcal{A}_n\big(1_Z\, 2_Z\, 3_\psi^+\, 4_{\overline{Z}} \ldots\big) \ - \ |3]\mathcal{A}_n\big(1_Z\, 2_\psi^+\, 3_Z\, 4_{\overline{Z}} \ldots\big) \\ & + \ |4]\mathcal{A}_n\big(1_Z\, 2_\psi^+\, 3_\psi^+\, 4_\psi^- \ldots\big) \ + \ldots\end{aligned} \tag{3.4.2}$$

Due to the Grassmann nature of the supercharges, there is a minus sign for each fermion that the supercharge has to move past to get to the $i$th state.

Similarly for a vector multiplet:

| state $i$ | $\mathcal{Q}\cdot i$ | $\mathcal{A}_n$ prefactor | $\mathcal{Q}^\dagger\cdot i$ | $\mathcal{A}_n$ prefactor |
|---|---|---|---|---|
| $\gamma^+$ | $\psi^+$ | $\vert i]$ | $0$ | $0$ |
| $\psi^+$ | $0$ | $0$ | $\gamma^+$ | $-\vert i\rangle$ |
| $\psi^-$ | $\gamma^-$ | $-\vert i]$ | $0$ | $0$ |
| $\gamma^-$ | $0$ | $0$ | $\psi^-$ | $\vert i\rangle$ |

(3.4.3)

where $\psi$ is a Weyl fermion and $\gamma$ is a vector boson.

In this notation, the supersymmetry Ward identities are equivalent to the statement that the following action of the supercharges annihilates the amplitude [71, 83, 91]

$$\begin{aligned} 0 = \mathcal{Q}\cdot\mathcal{A}_n\left(1,\dots,n\right) &= \sum_{i=1}^{n}(-1)^{L_i+P_i}|i]\mathcal{A}_n\left(1,\dots,\mathcal{Q}\cdot i,\dots,n\right), \\ 0 = \mathcal{Q}^\dagger\cdot\mathcal{A}_n\left(1,\dots,n\right) &= \sum_{i=1}^{n}(-1)^{L_i+P_i}|i\rangle\mathcal{A}_n\left(1,\dots,\mathcal{Q}^\dagger\cdot i,\dots,n\right), \end{aligned} \tag{3.4.4}$$

where $L_i$ is equal to the number of fermions to the *left* of $\mathcal{Q}^{(\dagger)}\cdot i$ and the factors $P_i = 0$ or $1$ correspond to the additional minus signs associated with the spinor prefactors as described in Tables 3.4.1 and 3.4.3. Note that the action of the supercharges always changes the number of fermions by $\pm1$, but that amplitudes are non-vanishing only if the number of fermions is even. So to get an interesting relation among amplitudes on the right-hand-side, the amplitude on the left-hand-side must vanish identically.

### 3.4.2 Soft Limits and Supermultiplets

We consider the chiral multiplet and vector multiplet separately and then extend the results to enhanced supersymmetry.

**Chiral multiplet.** Define the *soft factors* $\mathcal{S}_n^{(i)}$ as the momentum dependent coefficients in the holomorphic soft expansion taken here for simplicity on the first particle

$$\begin{aligned} \mathcal{A}_n\left(\{\epsilon|1\rangle,|1]\}_Z,\dots\right) &\rightarrow \mathcal{S}_n^{(0)}(1_Z,\dots)\,\epsilon^{\sigma_Z} + \mathcal{S}_n^{(1)}(1_Z,\dots)\,\epsilon^{\sigma_Z+1} + \mathcal{O}\left(\epsilon^{\sigma_Z+2}\right), \\ \mathcal{A}_n\left(\{\epsilon|1\rangle,|1]\}_\psi^+,\dots\right) &\rightarrow \mathcal{S}_n^{(0)}(1_\psi^+,\dots)\,\epsilon^{\sigma_\psi} + \mathcal{S}_n^{(1)}(1_\psi^+,\dots)\,\epsilon^{\sigma_\psi+1} + \mathcal{O}\left(\epsilon^{\sigma_\psi+2}\right). \end{aligned} \tag{3.4.5}$$

The soft weights are $\sigma_Z$ and $\sigma_\psi$ for the scalar and fermion, respectively. To see how supersymmetry forces relations among the soft weights and soft factors we use (3.4.4) to write

$$\begin{aligned}\mathcal{A}_n\left(1_Z,\ldots,n\right) &= \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{[Xi]}{[X1]}\,\mathcal{A}_n\left(1_\psi^+,\ldots,\mathcal{Q}\cdot i,\ldots,n\right),\\ \mathcal{A}_n\left(1_\psi^+,\ldots,n\right) &= \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{\langle Xi\rangle}{\langle X1\rangle}\,\mathcal{A}_n\left(1_Z,\ldots,\mathcal{Q}^\dagger\cdot i,\ldots,n\right),\end{aligned} \tag{3.4.6}$$

where the arbitrary $X$-spinor cannot be proportional to $|1\rangle$ or $|1]$.

Taking the holomorphic soft expansion on the right-hand-side of these expressions, in the second line only, an extra power of $\epsilon$ appears in the denominator and we find

$$\mathcal{S}_n^{(0)}(1_Z,\ldots)\,\epsilon^{\sigma_Z}+\mathcal{O}\left(\epsilon^{\sigma_Z+1}\right) = \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{[Xi]}{[X1]}\mathcal{S}_n^{(0)}(1_\psi^+,\ldots,\mathcal{Q}\cdot i,\ldots)\,\epsilon^{\sigma_\psi}+\mathcal{O}\left(\epsilon^{\sigma_\psi+1}\right),$$

$$\mathcal{S}_n^{(0)}(1_\psi^+,\ldots)\,\epsilon^{\sigma_\psi}+\mathcal{O}\left(\epsilon^{\sigma_\psi+1}\right) = \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{\langle Xi\rangle}{\langle X1\rangle}\mathcal{S}_n^{(0)}(1_Z,\ldots,\mathcal{Q}^\dagger\cdot i,\ldots)\,\epsilon^{\sigma_Z-1}+\mathcal{O}\left(\epsilon^{\sigma_Z}\right).$$

The leading power of $\epsilon$ on the right-hand-side must match the leading power on the left. It is possible that cancellations among the terms on the right-hand-side may effectively increase the leading power but never decrease it. This then gives the following inequalities

$$\sigma_Z \geq \sigma_\psi \quad \text{and} \quad \sigma_\psi \geq \sigma_Z - 1\,, \tag{3.4.7}$$

for which there are only two solutions

$$\boxed{\sigma_Z = \sigma_\psi + 1 \quad \text{or} \quad \sigma_Z = \sigma_\psi\,.} \tag{3.4.8}$$

These two options have different consequences for the soft factors. For $\sigma_Z = \sigma_\psi + 1$, we have

$$\begin{aligned} 0 &= \sum_{i=2}^{n}(-1)^{L_i+P_i}[Xi]\,\mathcal{S}_n^{(0)}\left(1_\psi^+,\ldots,\mathcal{Q}\cdot i,\ldots\right),\\ \mathcal{S}_n^{(0)}\left(1_\psi^+,\ldots\right) &= \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{\langle Xi\rangle}{\langle X1\rangle}\,\mathcal{S}_n^{(0)}\left(1_Z,\ldots,\mathcal{Q}^\dagger\cdot i,\ldots\right),\end{aligned} \tag{3.4.9}$$

while for $\sigma_\phi = \sigma_\psi$, we have

$$\begin{aligned} 0 &= \sum_{i=2}^{n}(-1)^{L_i+P_i}\langle Xi\rangle\mathcal{S}_n^{(0)}\left(1_Z,\ldots,\mathcal{Q}^\dagger\cdot i,\ldots\right),\\ \mathcal{S}_n^{(0)}\left(1_Z,\ldots\right) &= \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{[Xi]}{[X1]}\mathcal{S}_n^{(0)}\left(1_\psi^+,\ldots,\mathcal{Q}\cdot i,\ldots\right).\end{aligned} \tag{3.4.10}$$

In addition there will be an infinite number of similar relations which come from matching higher powers in $\epsilon$.

**Vector multiplet.** We define the soft factors as

$$\mathcal{A}_n\left(\{\epsilon|1\rangle, |1]\}^+_\gamma, \ldots\right) \;\rightarrow\; \mathcal{S}^{(0)}_n(1^+_\gamma, \ldots)\,\epsilon^{\sigma_\gamma} + \mathcal{S}^{(1)}_n(1^+_\gamma, \ldots)\,\epsilon^{\sigma_\gamma+1} + \mathcal{O}\left(\epsilon^{\sigma_\gamma+2}\right). \tag{3.4.11}$$

The analysis of the supersymmetry Ward identities proceeds similarly to that of the chiral multiplet and results in only two options for the soft weights:

$$\boxed{\sigma_\psi = \sigma_\gamma + 1, \quad \text{or} \quad \sigma_\psi = \sigma_\gamma\,.} \tag{3.4.12}$$

The consequences for the soft factors are for $\sigma_\psi = \sigma_\gamma + 1$

$$\begin{aligned} 0 &= \sum_{i=2}^{n}(-1)^{L_i+P_i}[Xi]\mathcal{S}^{(0)}_n\left(1^+_\gamma, \ldots, \mathcal{Q}\cdot i, \ldots\right), \\ \mathcal{S}^{(0)}_n\left(1^+_\gamma, \ldots\right) &= \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{\langle Xi\rangle}{\langle X1\rangle}\mathcal{S}^{(0)}_n\left(1^+_\psi, \ldots, \mathcal{Q}^\dagger\cdot i, \ldots\right), \end{aligned} \tag{3.4.13}$$

and for $\sigma_\gamma = \sigma_\psi$

$$\begin{aligned} 0 &= \sum_{i=2}^{n}(-1)^{L_i+P_i}\langle Xi\rangle\mathcal{S}^{(0)}_n\left(1^+_\psi, \ldots, \mathcal{Q}^\dagger\cdot i, \ldots\right), \\ \mathcal{S}^{(0)}_n\left(1^+_\psi, \ldots\right) &= \sum_{i=2}^{n}(-1)^{L_i+P_i+1}\frac{[Xi]}{[X1]}\mathcal{S}^{(0)}_n\left(1^+_\gamma, \ldots, \mathcal{Q}\cdot i, \ldots\right). \end{aligned} \tag{3.4.14}$$

Note that we have made no assumptions about the sign of $\sigma$, so the relations derived here are totally general. Also, the supersymmetry Ward identities hold at all orders in perturbation theory, so the relations among the soft behaviors remain true at loop-level.

**Extended supersymmetry.** Relations between the soft weights of particles in the same massless supermultiplets in extended supersymmetry follow directly from the $\mathcal{N} = 1$ results above, since the supersymmetry Ward identities take the same form for each pair of $(s, s+\frac{1}{2})$-multiplets. In particular, the soft weights of the boson ($\sigma_B$) and fermion ($\sigma_F$) in a $(s, s+\frac{1}{2})$-multiplet are related as

$$\begin{cases} \sigma_B = \sigma_F + 1 \quad \text{or} \quad \sigma_B = \sigma_F \quad \text{for } s \text{ integer}\,, \\ \sigma_B = \sigma_F - 1 \quad \text{or} \quad \sigma_B = \sigma_F \quad \text{for } s \text{ half-integer}\,. \end{cases} \tag{3.4.15}$$

These relations will be useful in later applications in this chapter. For now, we make a small aside and demonstrate the application of (3.4.15) to extended supergravity.

### 3.4.3 Example: Low-Energy Theorems in Supergravity

It is well-known that gravitons have a universal soft behavior [92]: when the soft limit (3.4.5) is applied to a single graviton, the amplitude diverges as $1/\epsilon^3$, i.e. the soft weight is $\sigma_2 = -3$. (In this section, we use a subscript on the soft weight to indicate the spin of the particle.) Applying (3.4.15) shows that the gravitino can have $\sigma_{3/2} = -2$ or $-3$. However, unitarity and locality constraints show [51] that amplitudes cannot be more singular than $1/\epsilon^2$ for a single soft gravitino, so it must be that $\sigma_{3/2} = -2$. This must be true in any supergravity theory.

Consider now a graviphoton in $\mathcal{N} \geq 2$ supergravity. Its supersymmetry Ward identities with the gravitino imply $\sigma_1 = -2$ or $\sigma_1 = -1$. The $\sigma_1 = -2$ behavior requires the graviphoton, and by supersymmetry also the gravitino, to interact with a pair of electrically charged particles via a dimensionless coupling; however, for the gravitino such a coupling is inconsistent with unitarity and locality [51]. So there is only one option, namely $\sigma_1 = -1$.

In pure $\mathcal{N} \geq 3$ supergravity, we also have spin-$\frac{1}{2}$ fermions in the graviton supermultiplet. By (3.4.15) and the previous results, they can have either $\sigma_{1/2} = -1$ or $0$. The analysis in [51] shows that $\sigma_{1/2} = -1$ requires a dimensionless coupling of the spin-$\frac{1}{2}$ particle with two other particles, for example via a Yukawa coupling. Since there are no dimensionless couplings in pure supergravity, it follows from [51] that the amplitude has to be $\mathcal{O}(\epsilon^0)$ or softer. This leaves only one option, namely that $\sigma_{1/2} = 0$ in pure supergravity.

In pure $\mathcal{N} \geq 4$ supergravity, the scalars in the supermultiplet can have $\sigma_0 = 0$ or $\sigma_0 = 1$. If we focus on the MHV sector, the supersymmetry Ward identities give

$$\mathcal{A}_n(1_Z\, 2_{\bar{Z}}\, 3_h^-\, 4_h^+ \dots n_h^+) = \frac{\langle 13\rangle^4}{\langle 23\rangle^4}\, \mathcal{A}_n(1_h^+\, 2_h^-\, 3_h^-\, 4_h^+ \dots n_h^+)\,, \tag{3.4.16}$$

where $Z$ and $\bar{Z}$ denote any pair of conjugate scalars and $h$ are gravitons. Taking line 1 soft holomorphically, $|1\rangle \to \epsilon|1\rangle$, the graviton amplitude on the RHS diverges as $1/\epsilon^3$ but the prefactor vanishes as $\epsilon^4$. It follows that the MHV amplitude vanishes as $\mathcal{O}(\epsilon)$ in the single soft-scalar limit. In other words, for MHV amplitudes $\sigma_0 = 1$. It is tempting to conclude that one must have $\sigma_0 = 1$ for all amplitudes, but that is too glib, as we now explain.

It is known that the scalar cosets of $\mathcal{N} \geq 4$ pure supergravity theories in 4d are symmetric, and therefore lead to $\sigma_0 = 1$ vanishing low-energy theorems. But at the level of the on-shell amplitudes, this conclusion does not follow from the supersymmetry Ward identities alone: as we have seen, they give $\sigma_0 = 1$ or $\sigma_0 = 0$. That analysis has to remain true at all loop-orders. In $\mathcal{N} = 4$ supergravity, for example, the anomaly of the $U(1)$ R-symmetry can be expected to affect the soft behavior at some order. Our arguments show that it cannot happen in the MHV sector, but does not rule it out beyond MHV; this is what the $\sigma_0 = 0$ accounts for. Furthermore, one can add

| helicity state | $\sigma$ |
|---|---|
| +2 graviton | $-3$ |
| +3/2 gravitino | $-2$ |
| +1 graviphoton | $-1$ |
| +1/2 fermion | $0$ |
| 0 scalar | $0$ or $+1$ |
| -1/2 fermion | $+1$ |
| -1 graviphoton | $+1$ |
| -3/2 gravitino | $+1$ |
| -2 graviton | $+1$ |

Table 3.2: Holomorphic soft weights $\sigma$ for the $\mathcal{N} = 8$ supermultiplet. Note that the soft weights in this table follow from taking the soft limit holomorphically, $|i\rangle \to \epsilon|i\rangle$ for all states, independently of the sign of their helicity. At each step in the spectrum, the soft weight either changes by 1 or not at all. Note that one could also have used the anti-holomorphic definition $|i] \to \epsilon|i]$ of taking the soft limit; in that case the soft weights would just have reversed, to start with $\sigma = -3$ for the negative helicity graviton, but no new constraints would have been obtained on the scalar soft weights. In $\mathcal{N} = 8$ supergravity, the 70 scalars are Goldstone bosons of the coset $E_{7(7)}/SU(8)$ and hence $\sigma = 1$. Including higher-derivative corrections may change this behavior to $\sigma = 0$ depending on whether the added terms are compatible with the coset structure.

higher-derivative operators to the supergravity action such that supersymmetry is preserved but the low-energy theorems are not. Indeed, string theory does this in the $\alpha'$-expansion by adding to the $\mathcal{N} = 8$ tree-level action a supersymmetrizable operator $\alpha'^3 e^{-6\phi} R^4$. This operator does not affect the soft behavior of MHV amplitudes, but it is known that it does result in non-vanishing single soft scalar limits for 6-particle NMHV amplitudes at order $\alpha'^3$ [93, 94]. The results for $\mathcal{N} = 8$ supersymmetry are summarized in Table 3.2.

## 3.5 Supersymmetric Non-linear Sigma Model

Perhaps the simplest and most familiar class of models that exhibit both linearly realized supersymmetry and interesting low-energy theorems are the *supersymmetric non-linear sigma models*. Of particular interest are the *coset* sigma models for which the target manifold is a homogeneous space $G/H$. At lowest order, the coset sigma model captures the universal low-energy behavior of the scalar Goldstone modes of a spontaneous symmetry breaking pattern $G \to H$, where $G$ and $H$ are the *isometry* and *isotropy* groups of the target manifold respectively. If the target manifold is additionally a *symmetric* space and there are no 3-point interactions, then the off-shell Ward-Takahashi identities for the spontaneously broken currents imply $\sigma = 1$ vanishing low-energy theorems for the Goldstone scalars. An interesting recent perspective on coset sigma models can be found in [95].

At leading order it is fairly straightforward to calculate the on-shell scattering amplitudes for such a model from the (two-derivative) non-linear sigma model effective action. Using the methods of on-shell recursion, the use of an effective action is unnecessary. Instead, we may assume low-energy theorems and on-shell Ward identities of the isotropy group $H$ as the on-shell data that defines the model. Using the procedure of the soft bootstrap described in Section 3.2.4, we may apply subtracted recursion to construct the contributions to the S-matrix at leading order.

A particularly simple and well-studied example of such a construction has previously been given for the $\frac{U(N)\times U(N)}{U(N)}$ coset sigma model [79, 80]. There are several nice features of this model which make it an appealing toy-model to study on-shell. As will be discussed in Section 3.6.6, at leading order ($\tilde{\Delta} = 1$ or equivalently two-derivative) the isotropy $U(N)$ symmetry allows for the construction of *flavor*-ordered partial amplitudes with only $(n-3)!$ independent amplitudes for the scattering of $n$ Goldstone scalars.

The situation is somewhat less straightforward for models describing the low-energy dynamics of the Goldstone modes of internal symmetry breaking with some amount of linearly realized supersymmetry.[12] There are several interesting consequences of this combination of symmetries. The states must form mass degenerate multiplets of the supersymmetry algebra, which in this case means that the Goldstone scalars must always transform together with additional massless spinning states. As discussed in Section 3.4.2, the low-energy theorems of each of the particles in these *Goldstone* multiplets are not independent.

It is well-known in the literature of supersymmetric field theories that to construct a supersymmetric action, the massless scalar modes must parametrize a target space manifold with *Kähler* structure for $\mathcal{N} = 1$ supersymmetry [96]. For $\mathcal{N} = 2$ supersymmetry the target space manifold must have the structure

$$\mathcal{M}_{\mathcal{N}=2} = \mathcal{M}_{\mathrm{V}} \times \mathcal{M}_{\mathrm{H}}, \tag{3.5.1}$$

where the scalars of the *vector* multiplets parametrize the *special-Kähler* manifold $\mathcal{M}_{\mathrm{V}}$ while the scalars belonging to *hyper* multiplets parametrize the *hyper-Kähler* manifold $\mathcal{M}_{\mathrm{H}}$ [97]. As a consequence, despite the obvious virtues of a flavor ordered representation, this makes studying the supersymmetrization of the $\frac{U(N)\times U(N)}{U(N)}$ coset sigma model using subtracted recursion more difficult, since even in the $\mathcal{N} = 1$ case the target manifold is not Kähler. This does not mean that the internal symmetry breaking pattern $U(N)\times U(N) \to U(N)$ is impossible in an $\mathcal{N} = 1$ supersymmetric model. Rather it means that the target space contains $\frac{U(N)\times U(N)}{U(N)}$ as a non-Kähler submanifold and includes additional directions in field space or equivalently includes additional massless *quasi-Goldstone* scalars [98]. In general there is no unique way to extend the symmetry breaking coset to a Kähler manifold, because in any given example the spectrum of quasi-Goldstone modes depends on the details of the UV physics. Correspondingly, the quasi-

[12] In this more general context *internal* symmetry includes R-symmetry. For our purposes the relevant property is that the conserved charges are Lorentz scalars and so correspond to a spectrum of spin-0 Goldstone modes.

Goldstone scalars do not satisfy the kind of universal low-energy theorems necessary for us to construct the scattering amplitudes recursively.

Instead, in this section we will study the interplay of low-energy theorems and supersymmetry by considering the simplest symmetric coset that is both Kähler and special-Kähler

$$\frac{SU(2)}{U(1)} \cong \mathbb{CP}^1 \,, \tag{3.5.2}$$

and therefore should admit both an $\mathcal{N}=1$ and $\mathcal{N}=2$ supersymmetrization. Our assumption here is that the target manifold *is* the coset manifold and therefore the massless spectrum should contain only two real scalar degrees of freedom, both Goldstone modes. They form a single complex scalar field $Z, \overline{Z}$ which carries a conserved charge associated with the isotropy $U(1)$. These properties uniquely determine the Goldstone multiplets as an $\mathcal{N}=1$ chiral and $\mathcal{N}=2$ vector multiplet respectively.

The main results of this section are (1) the demonstration that both the $\mathcal{N}=1$ and $\mathcal{N}=2$ $\mathbb{CP}^1$ non-linear sigma models are constructible on-shell using recursion without the need to explicitly construct an effective action. And (2) this construction gives a new *on-shell* perspective on the relationship between the linearly realized target space isotropies of $\mathcal{M}_{\text{V}}$ and electric-magnetic duality transformations of the associated vector bosons.

### 3.5.1 $\mathcal{N}=1$ $\mathbb{CP}^1$ NLSM

The $\mathcal{N}=1$ $\mathbb{CP}^1$ non-linear sigma model is defined by the following on-shell data:

- A spectrum consisting of a massless $\mathcal{N}=1$ chiral multiplet $(Z, \bar{Z}, \psi^+, \psi^-)$.
- Scattering amplitudes satisfy $\mathcal{N}=1$ supersymmetry Ward identities.
- Scattering amplitudes satisfy isotropy $U(1)$ Ward identities under which $Z, \bar{Z}$ are charged.
- $\sigma_Z = \sigma_{\bar{Z}} = 1$ soft weight for the scalars.

Using the approach of the soft bootstrap, we begin by constructing the most general on-shell amplitudes at lowest valence that are consistent with the above data and minimize $\tilde{\Delta}$. There are no possible 3-point amplitudes consistent with the assumptions and so we must begin at 4-point. A $|Z|^4$ interaction, corresponding to $\tilde{\Delta}=0$, is consistent with $U(1)$ conservation but violates the assumed low-energy theorem. The next-to-lowest reduced dimension interactions correspond to $\tilde{\Delta}=1$ and have a unique 4-point amplitude consistent with the assumptions

$$\mathcal{A}_4(1_Z\, 2_{\bar{Z}}\, 3_Z\, 4_{\bar{Z}}) = \frac{1}{\Lambda^2} s_{13}. \tag{3.5.3}$$

Note that at 4-point, the conservation of the $U(1)$-charge for the complex scalar is automatically enforced as a consequence of the supersymmetry Ward identitites. We will see that this implies the conservation of the $U(1)$ charge for amplitudes with arbitrary number of external particles corresponding to $\tilde{\Delta} = 1$. Note that this is not automatic for higher order ($\tilde{\Delta} > 1$) corrections and must be imposed as a separate constraint. The remaining 4-point amplitudes are completely determined by supersymmetry; it is convenient to summarize the component amplitudes in a single *superamplitude* [99]

$$\mathcal{A}_4(1_{\Phi^+}2_{\Phi^-}3_{\Phi^+}4_{\Phi^-}) = \frac{1}{\Lambda^2}\,[13]\,\delta^{(2)}(\tilde{Q}) = \frac{1}{2\Lambda^2}\,[13]\sum_{i,j=1}^{4}\langle ij\rangle\eta_i\eta_j\,. \tag{3.5.4}$$

Here we have introduced two chiral superfields $\Phi^+$ and $\Phi^-$ that contain the positive and negative helicity fields of the $\mathcal{N}=1$ chiral multiplet as

$$\Phi^+ = \psi^+ + \eta Z\ , \qquad \Phi^- = \bar{Z} - \eta\,\psi^-. \tag{3.5.5}$$

$\eta$ is the Grassmann coordinate of $\mathcal{N}=1$ on-shell superspace and $\eta_i$ denotes the $\eta$-coordinate of the $i^{\text{th}}$ superfield. We can obtain all the component amplitudes by projecting out components of the superfield. For example, the all-fermion amplitude can be derived as follows

$$\mathcal{A}_4(1^+_\psi 2^-_\psi 3^+_\psi 4^-_\psi) = \frac{\partial}{\partial\eta_2}\frac{\partial}{\partial\eta_4}\mathcal{A}_4(1_{\Phi^+}2_{\Phi^-}3_{\Phi^+}4_{\Phi^-}) = -\frac{1}{\Lambda^2}\,[13]\,\langle 24\rangle. \tag{3.5.6}$$

It is useful to note that the expression (3.5.4) is manifestly local. It follows that all component amplitudes are free of factorization singularities, indicating the absence of 3-point interactions in this theory. Note also that the pure fermion sector is exactly the NJL model detected by the soft bootstrap in Section 3.3.2.

Next, we use these 4-point amplitudes to recursively construct $n$-point amplitudes. Following the discussion in Section 3.4, we note that the soft weight of the fermion must be either $\sigma_\psi = 0$ or $\sigma_\psi = 1$. Making the conservative choice $\sigma_\psi = 0$, we evaluate the constructibility criterion on the above on-shell data,

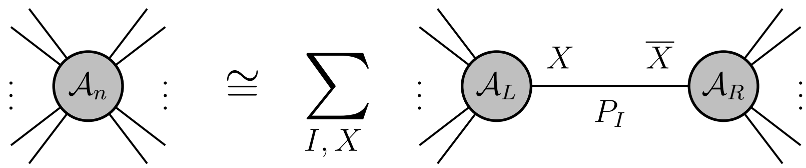
$$4 < 2n_s + n_f, \tag{3.5.7}$$

where $n_f$ is the number of external fermion states of the $n$-point amplitude and $n_s = n - n_f$ is the number of external scalar states. For $n > 4$, this condition is satisfied for all $n$-point amplitudes. We find that recursively constructing the 6-point amplitudes yields an $a_i$-independent expression. All the 6-point amplitudes can be found in Appendix B.1.

If however we make the stronger assumption $\sigma_\psi = 1$, the recursively constructed 6-point amplitude is $a_i$-dependent and therefore fails the consistency checks. As a result we conclude that the true soft weight of the fermion of our theory is $\sigma_\psi = 0$ and this is sufficient to construct the

S-matrix at leading order from the 4-point seed amplitudes (3.5.4).

The recursive constructibility of the S-matrix has non-trivial consequences for the possible conserved additive quantum numbers. In a recursive model the only non-zero amplitudes are those which can be constructed by gluing together lower-point on-shell amplitudes




where the states $X, \bar{X}$ on either side of the factorization channel $I$ have CP conjugate quantum numbers. As discussed further in Appendix C, if an additive quantum number is conserved by all seed amplitudes then it must be conserved by all recursively constructible amplitudes.

For example, in the present context the seed amplitudes conserve two *independent* $U(1)$ charges:

| | $U(1)_A$ | $U(1)_B$ |
|---|---|---|
| $Z$ | $q_A$ | 0 |
| $\bar{Z}$ | $-q_A$ | 0 |
| $\psi^+$ | 0 | $q_B$ |
| $\psi^-$ | 0 | $-q_B$ |
| $\eta$ | $-q_A$ | $q_B$ |
| $\Phi^+$ | 0 | $q_B$ |
| $\Phi^-$ | $-q_A$ | 0 |

We know to expect the existence of an isotropy $U(1)$ under which the scalars are charged, but from our on-shell construction it is unclear whether this should be $U(1)_A$ or a combination of $U(1)_A$ and $U(1)_B$. We have presented the charges as two independent R-symmetries but more correctly we should consider them as a single *global* $U(1)$ and a $U(1)_R$. The presence of a second conserved quantum number is not part of the definition of the $\mathbb{CP}^1$ non-linear sigma model but is instead an *emergent* or *accidental* symmetry at lowest order in the EFT. In general one would expect $U(1)_A \times U(1)_B$ to be explicitly broken to the isotropy $U(1)$ by higher dimension operators.

### 3.5.2 $\mathcal{N} = 2$ $\mathbb{CP}^1$ NLSM

The $\mathcal{N} = 2$ $\mathbb{CP}^1$ NLSM is defined by the following on-shell data:

- A spectrum consisting of a massless $\mathcal{N} = 2$ vector multiplet $(Z, \bar{Z}, \psi^{a+}, \psi_a^-, \gamma^+, \gamma^-)$, where $a = 1, 2$.

- Scattering amplitudes satisfy $\mathcal{N}=2$ supersymmetry Ward identities.
- Scattering amplitudes satisfy isotropy $U(1)$ Ward identities under which $Z, \bar{Z}$ are charged.

Note that, importantly, we do *not* impose the the soft weight of the scalars $\sigma_Z = \sigma_{\bar{Z}} = 1$. As we will explain further below, no model with the above properties *and* vanishing scalar soft limits exists.

To proceed, interactions with reduced dimension $\tilde{\Delta} = 0$ (such as Yukawa interactions) are incompatible with $\mathcal{N}=2$ supersymmetry for a single vector multiplet. Thus, the minimal value is $\tilde{\Delta} = 1$; that is of course also the value for the $\mathcal{N}=1$ model. It is curious to note that $\mathcal{N}=2$ supersymmetry is sufficient to uniquely construct the S-matrix at this order in $\tilde{\Delta}$. As we show in the following, *without* assuming vanishing scalar soft limits, the restriction of the external states to a single chiral multiplet $(Z, \bar{Z}, \psi^{1+}, \psi_1^-)$ reproduces the $\mathcal{N}=1$ $\mathbb{CP}^1$ sigma model.

As in the previous section, for $\tilde{\Delta} = 1$ the 4-point scalar amplitude takes the form (3.5.3). All 4-point component amplitudes are uniquely fixed by the 4-scalar amplitudes by the $\mathcal{N}=2$ supersymmetry Ward identities and they can be encoded compactly into superamplitudes using two chiral superfields [99]

$$\begin{aligned} \Phi^+ &= \gamma^+ + \eta_1 \psi^{1+} + \eta_2 \psi^{2+} - \eta_1 \eta_2 Z, \\ \Phi^- &= \bar{Z} + \eta_1 \psi_2^- - \eta_2 \psi_1^- - \eta_1 \eta_2 \gamma^- . \end{aligned} \tag{3.5.8}$$

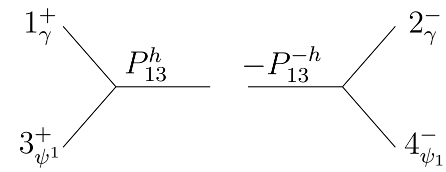
Here $\eta_1$ and $\eta_2$ are the Grassmann coordinates of $\mathcal{N}=2$ on-shell superspace. The $R$-indices on $\psi^a$ are raised and lowered using $\epsilon_{ab}$, so $\psi_2^- = \epsilon_{21}\psi^{1-} = \psi^{1-}$ and $\psi_1^- = \epsilon_{12}\psi^{2-} = -\psi^{2-}$. In terms of the superfields, the 4-point superamplitude can be expressed as

$$\mathcal{A}_4(1_{\Phi^+} 2_{\Phi^-} 3_{\Phi^+} 4_{\Phi^-}) = \frac{1}{\Lambda^2} \frac{[13]}{\langle 13 \rangle} \delta^{(4)}(\tilde{Q}) = \frac{1}{4\Lambda^2} \frac{[13]}{\langle 13 \rangle} \prod_{a=1}^{2} \sum_{i,j=1}^{4} \langle ij \rangle \eta_{ia} \eta_{ja}. \tag{3.5.9}$$

We use $\eta_{ia}$ to denote the $a^{\text{th}}$ Grassmann coordinate of the $i^{\text{th}}$ external superfield. In contrast to (3.5.4), the superamplitude (3.5.9) generates component amplitudes that are not local due to the factorization singularity at $P_{13}^2 \to 0$. For example, consider the following component amplitude

$$\mathcal{A}_4(1_\gamma^+ 2_\gamma^- 3_{\psi^1}^+ 4_{\psi_1}^-) = -\frac{\partial}{\partial \eta_{21}} \frac{\partial}{\partial \eta_{22}} \frac{\partial}{\partial \eta_{31}} \frac{\partial}{\partial \eta_{42}} \mathcal{A}_4(1_{\Phi^+} 2_{\Phi^-} 3_{\Phi^+} 4_{\Phi^-}) = -\frac{1}{\Lambda^2} \frac{[13]\,[14]\,\langle 24 \rangle}{[24]}. \tag{3.5.10}$$

Locality and unitarity imply that this 4-point amplitude must factorize into 3-point amplitudes on the singularity at $P_{13}^2 \to 0$. Denoting the helicity of the exchanged particle $h$, the amplitude factorizes as




The contribution to the residue on the singularity takes the form

$$
\begin{aligned}
P_{13}^2\mathcal{A}_4(1^+_\gamma 2^-_\gamma 3^+_{\psi^1} 4^-_{\psi_1})\Big|_{P_{13}^2=0} &= \mathcal{A}_3\big(1^+_\gamma 3^+_{\psi^1}(P_{13})_h\big)\mathcal{A}_3\big((-P_{13})_{-h}2^-_\gamma 4^-_{\psi_1}\big) \\
&= \Big(\frac{g_1}{\Lambda}[13]^{3/2-h}[1P_{13}]^{1/2+h}[3P_{13}]^{-1/2+h}\Big)\Big(\frac{g_2}{\Lambda}\langle 24\rangle^{3/2-h}\langle 2P_{13}\rangle^{1/2+h}\langle 4P_{13}\rangle^{-1/2+h}\Big) \\
&= \frac{g_1 g_2}{\Lambda^2}(-1)^{2h}\,[13]^{3/2-h}\,\langle 24\rangle^{3/2+h}\,[23]^{1/2-h}\,[14]^{1/2+h}\ , \qquad (3.5.11)
\end{aligned}
$$

with the 3-point amplitudes completely determined by Poincaré invariance and little group scaling. Comparing with the explicit form of the residue calculated from (3.5.10)

$$
P_{13}^2\mathcal{A}_4(1^+_\gamma 2^-_\gamma 3^+_{\psi^1} 4^-_{\psi_1})\Big|_{P_{13}^2=0} = \frac{1}{\Lambda^2}\,[13]\,[14]\,\langle 24\rangle^2, \qquad (3.5.12)
$$

we find that $h = 1/2$ and $g_1 g_2 = -1$. The exchanged particle of helicity $h = 1/2$ can be either $\psi^{1+}$ or $\psi^{2+}$. The locality of the $\mathcal{A}_4(1^+_{\psi^1} 2^-_{\psi_1} 3^+_{\psi^1} 4^-_{\psi_1})$ and $\mathcal{A}_4(1^+_{\psi^2} 2^-_{\psi_2} 3^+_{\psi^2} 4^-_{\psi_2})$ tells us that they do not factorize on the $(P_{13})^2 \to 0$ pole. We conclude that $\mathcal{A}_3(1^+_\gamma 2^+_{\psi_1} 3^+_{\psi_1}) = \mathcal{A}_3(1^+_\gamma 2^+_{\psi_2} 3^+_{\psi_2}) = 0$, while

$$
\mathcal{A}_3(1^+_\gamma 2^+_{\psi_1} 3^+_{\psi_2}) = \frac{g_1}{\Lambda}\,[12]\,[13]\ , \qquad \mathcal{A}_3(1^-_\gamma 2^-_{\psi_1} 3^-_{\psi_2}) = \frac{g_2}{\Lambda}\langle 12\rangle\langle 13\rangle\,. \qquad (3.5.13)
$$

We carry out a similar exercise with $\mathcal{A}_4(1^+_\gamma 2^-_\gamma 3^+_\gamma 4^-_\gamma)$ for a particle of helicity $h$ in the $P_{13}^2 \to 0$ factorization channel. Comparing with the 4-point amplitude (3.5.9) fixes $h = 0$. This could correspond to either $Z$ or $\bar{Z}$ exchange. The absence of a $P_{14}^2 \to 0$ pole in $\mathcal{A}_4(1^+_\gamma 2^-_\gamma 3_Z 4_{\bar{Z}})$ shows that $\mathcal{A}_3(1^+_\gamma 2^+_\gamma 3_{\bar{Z}}) = 0$ and

$$
\mathcal{A}_3(1^+_\gamma 2^+_\gamma 3_Z) = \frac{g_3}{\Lambda}\,[12]^2\ , \qquad \mathcal{A}_3(1^-_\gamma 2^-_\gamma 3_{\bar{Z}}) = \frac{g_4}{\Lambda}\langle 12\rangle^2\,, \qquad (3.5.14)
$$

where $g_3 g_4 = 1$. Demanding that all non-local 4-point amplitudes factorize correctly fixes $-g_1 = g_2 = g_3 = g_4 = -1$. The 3-point superamplitudes are

$$
\begin{aligned}
\mathcal{A}_3(1_{\Phi^-}2_{\Phi^-}3_{\Phi^-}) &= \delta^{(4)}(\tilde{Q}) = \frac{1}{4\Lambda}\prod_{a=1}^{2}\sum_{i,j=1}^{3}\langle ij\rangle\eta_{ia}\eta_{ja}\,, \qquad (3.5.15) \\
\mathcal{A}_3(1_{\Phi^+}2_{\Phi^+}3_{\Phi^+}) &= \frac{1}{\Lambda}\delta^{(2)}(\eta_1\,[23] + \eta_2\,[31] + \eta_3\,[12]) = \frac{1}{\Lambda}\prod_{a=1}^{2}(\eta_{1a}\,[23] + \eta_{2a}\,[31] + \eta_{3a}\,[12])\,,
\end{aligned}
$$

where $\prod_{a=1}^{2} f_a$ is defined as $f_1 f_2$.

It is interesting to observe that even though the $\mathcal{N} = 0, 1$ and $2$ $\mathbb{CP}^1$ NLSM have the pure scalar 4-point amplitude in common, in the latter case the extended supersymmetry together with locality require the presence 3-point interactions.

We are now in a position to address the constructibility of general $n$-point amplitudes. Since we are not assuming vanishing soft limits as part of our on-shell data, we are not able to make use of subtracted recursion. This is only problematic for a subset of the amplitudes in this model, at least at leading order. The unsubtracted constructibility criterion for this model reads

$$4 < n_f + 2n_v, \tag{3.5.16}$$

where $n_f$ and $n_v$ are the number of fermions and vector bosons respectively. It turns out that the amplitudes that do not satisfy this criterion can be determined from the $\mathcal{N} = 2$ supersymmetry Ward identities in terms of those that do. Remarkably, without making any strong assumptions about the structure of low-energy theorems for the scalars, which usually characterize the sigma model coset structure, the $\mathcal{N} = 2$ supersymmetry is sufficient at leading order to both construct the entire S-matrix and reproduce the amplitudes of the $\mathcal{N} = 1$ and $\mathcal{N} = 0$ models as special cases.

This same statement can be made in the perhaps more familiar language of local field theory. At this order in the EFT expansion, the S-matrix elements should be calculable from some effective action, the bosonic sector of which should be described by a two-derivative Lagrangian of the general form

$$\mathcal{L}_{\text{eff}} = P\left(|Z|^2\right)|\partial_\mu Z|^2 + Q\left(|Z|^2\right) Z\, F_+^2 + \text{h.c.} \tag{3.5.17}$$

where $P(|Z|^2)$ and $Q(|Z|^2)$ are some functions analytic around $Z \sim 0$. Insisting that the S-matrix elements satisfy the on-shell $\mathcal{N} = 2$ supersymmetry Ward identities is equivalent to requiring the existence of off-shell $\mathcal{N} = 2$ supersymmetry transformations under which the effective action is invariant. The on-shell uniqueness result is equivalent to the statement that the off-shell $\mathcal{N} = 2$ supersymmetry uniquely (up to field redefinitions) determines the form of the two-derivative effective action. In particular, the function $P(|Z|^2)$ is uniquely determined to be

$$P\left(|Z|^2\right) = \left(\frac{1}{1+|Z|^2}\right)^2, \tag{3.5.18}$$

corresponding to the Fubini-Study metric on $\mathbb{CP}^1$.

Since the entire S-matrix is determined, we can explicitly demonstrate how the presence of the vector bosons modifies the structure of the low-energy theorems from the naive vanishing soft limits suggested by the coset structure. Consider the following relation among 5-point amplitudes given by the $\mathcal{N}=2$ supersymmetry Ward identities

$$\mathcal{A}_5\left(1_\gamma^+,2_\gamma^+,3_Z,4_Z,5_{\bar{Z}}\right)=\frac{\langle 34\rangle^2}{\langle 45\rangle^2}\mathcal{A}_5\left(1_\gamma^+,2_\gamma^+,3_\gamma^+,4_Z,5_\gamma^-\right). \tag{3.5.19}$$

The amplitude on the right-hand-side satisfies (3.5.16) and therefore is constructible using unsubtracted recursion. This gives the non-constructible amplitude on the left-hand-side as

$$\mathcal{A}_5\left(1_\gamma^+,2_\gamma^+,3_Z,4_Z,5_{\bar{Z}}\right)=\frac{1}{\Lambda^3}\langle 34\rangle^2\left(\frac{[12][34]}{\langle 12\rangle\langle 34\rangle}+\frac{[23][14]}{\langle 23\rangle\langle 14\rangle}+\frac{[31][24]}{\langle 31\rangle\langle 24\rangle}\right). \tag{3.5.20}$$

The soft limits on particles 1, 2, 3 and 4 vanish, as expected. The soft limit on particle 5, however, is $\mathcal{O}(1)$, contrary to the expected soft behavior for a Goldstone mode of a symmetric coset. Explicitly

$$\mathcal{A}_5\left(1_\gamma^+,2_\gamma^+,3_Z,4_Z,5_{\bar{Z}}\right)\xrightarrow{|5]\to\epsilon|5]}\frac{1}{\Lambda^3}[12]^2+\mathcal{O}(\epsilon). \tag{3.5.21}$$

It is interesting that the coupling to the photons, required by $\mathcal{N}=2$ supersymmetry, results in non-vanishing soft scalar limits for a theory with a symmetric coset. In principle, this amplitude could have had a contact contribution of the form $\propto[12]^2$, but our calculation shows that such a term would be incompatible with $\mathcal{N}=2$ supersymmetry.

The maximal $R$-symmetry group that this model can realize is $U(2)_R=U(1)_R\times SU(2)_R$. We will now verify that the $SU(2)_R$ symmetry Ward identities hold for the seed amplitudes, the $U(1)_R$ we will address separately. To do this we choose a basis for the generators of $SU(2)_R$. The scalars and vectors both transform as $SU(2)$ singlets. The positive helicity fermion species $\psi^{1,2+}$ will transform in the fundamental representation under

$$\mathcal{T}_0=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad \mathcal{T}_+=\begin{pmatrix}0&1\\0&0\end{pmatrix},\quad \mathcal{T}_-=\begin{pmatrix}0&0\\1&0\end{pmatrix}. \tag{3.5.22}$$

The negative helicity fermions transform in the anti-fundamental with $\bar{\mathcal{T}}_i=-\mathcal{T}_i^\dagger$. This tells us that the $\mathcal{T}_0$-Ward identity is satisfied as long as the fermion species appear in pairs of (a) different helicity, same species or (b) same helicity, different species. This is true of all the non-zero amplitudes in this model. The action of $\mathcal{T}_+$ and $\mathcal{T}_-$ are

| state $i$ | $\mathcal{T}_+\cdot i$ | $\mathcal{A}_n$ prefactor | $\mathcal{T}_-\cdot i$ | $\mathcal{A}_n$ prefactor | $\mathcal{T}_0\cdot i$ | $\mathcal{A}_n$ prefactor |
|---|---|---|---|---|---|---|
| $\psi^{1+}$ | 0 | 0 | $\psi^{2+}$ | 1 | $\psi^{1+}$ | 1 |
| $\psi^{2+}$ | $\psi^{1+}$ | 1 | 0 | 0 | $\psi^{2+}$ | $-1$ |
| $\psi_1^-$ | $\psi_2^-$ | $-1$ | 0 | 0 | $\psi_1^-$ | $-1$ |
| $\psi_2^-$ | 0 | 0 | $\psi_1^-$ | $-1$ | $\psi_2^-$ | 1 |

(3.5.23)

We find that all 3-point and 4-point amplitudes in this model satisfy the $SU(2)_R$ Ward identities, for example

$$\begin{aligned}\mathcal{T}_- \cdot \mathcal{A}_4(1^+_{\psi^1}2^-_{\psi_2}3^+_{\psi^1}4^-_{\psi_1}) &= \mathcal{A}_4(1^+_{\psi^2}2^-_{\psi_2}3^+_{\psi^1}4^-_{\psi_1}) - \mathcal{A}_4(1^+_{\psi^1}2^-_{\psi_1}3^+_{\psi^1}4^-_{\psi_1}) + \mathcal{A}_4(1^+_{\psi^1}2^-_{\psi_2}3^+_{\psi^2}4^-_{\psi_1}) \\ &= -\frac{[13]}{[24]}\,(s+t+u) = 0\,.\end{aligned} \tag{3.5.24}$$

As discussed above, we conclude that at leading order the $SU(2)_R$ Ward identities are satisfied by all amplitudes in the $\mathcal{N}=2$ model.

Following the same approach as described for the $\mathcal{N}=1$ model, conservation laws satisfied by the seed amplitudes imply that the same quantities are conserved by all leading-order amplitudes if they are recursively constructible. This result extends to non-Abelian symmetries, which in the on-shell language correspond to Ward identities for non-diagonal generators; this is shown for $SU(2)$ in Appendix C. The amplitudes that are not constructible using recursion are fixed by supersymmetry in terms of those that are. Therefore, they will also respect the conservation laws and non-Abelian symmetries of the seed amplitudes.

This model also conserves a separate $U(1)_R$ charge. We know to expect the conservation of the charge associated with the $U(1)$ isotropy group. In the $\mathcal{N}=1$ case we found that the scattering amplitudes conserve an R-charge $U(1)_A$ assigned only to the complex scalar but it was consistent with the existence of $U(1)_B$ that the isotropy $U(1)$ might also assign a charge to the fermion or even to assign equal charges in the form of a global symmetry. In the present context we also have two independent $U(1)$ symmetries. The first is the $U(1) \subset SU(2)_R$ which assigns opposite charges to the fermions $\psi^{1+}$ and $\psi^{2+}$. The second assigns charges to each of the states which, up to overall normalization can be deduced from the 3- and 4-point seed amplitudes and are summarized in the following table:

| | $U(1)_R$ | $SU(2)_R$ |
|---|---|---|
| $Z$ | $-4$ | **1** |
| $\bar{Z}$ | $4$ | **1** |
| $\psi^{a+}$ | $-1$ | **2** |
| $\psi_a^-$ | $1$ | **2** |
| $\gamma^+$ | $2$ | **1** |
| $\gamma^-$ | $-2$ | **1** |
| $\eta_a$ | $3$ | **2** |
| $\Phi^+$ | $2$ | **1** |
| $\Phi^-$ | $4$ | **1** |

These are the *only* linear symmetries compatible with the seed amplitudes. The isotropy $U(1)$

must therefore be identified with some linear combination of $U(1)_R$ and $U(1) \subset SU(2)_R$. This is perhaps surprising, it tells us that the massless vector boson must also be charged under the isotropy $U(1)$. Just as for the fermions, the vector charges are *chiral* meaning that the positive and negative helicity states have opposite charges. Such charges for vectors are associated with electric-magnetic duality symmetries.

Such an extra $U(1)_R$ symmetry is possible because the maximal outer-automorphism group of the $\mathcal{N} = 2$ supersymmetry algebra is $U(2)_R$. The assignment of the associated charges is, up to normalization, fixed by the charge of the highest helicity state in the multiplet. It is interesting to observe that in the present context, knowledge of the non-vanishing 4-point amplitudes is *insufficient* to determine the $U(1)_R$ charge assignments. It is only from considering the 3-point amplitudes that we find the assignment of a non-zero chiral charge for the vector bosons unavoidable. Consider for example the amplitudes (3.5.14). Since the scalar is required to be charged under the isotropy $U(1)$, which in this case must be the $U(1)_R$ since there are no other symmetries under which the scalar is charged, we see that the vector must also be charged and satisfy $2q[\gamma^+] = -q[Z]$. The existence of fundamental 3-point interactions in this model was deduced by demanding that the singularities of the 4-point amplitudes be identified with physical factorization channels. From an on-shell point of view, it is therefore an *unavoidable* consequence of locality, unitarity and supersymmetry that the $\mathcal{M}_V$ isotropy group of an $\mathcal{N} = 2$ non-linear sigma model acts on the vector bosons as an electric-magnetic duality transformation.

The necessary existence of the fundamental 3-point amplitudes (3.5.13) and (3.5.14) has a further interesting consequence for the low-energy behavior of the vector boson. In [51] it was shown that *singular* low-energy theorems arise from the presence of certain 3-point amplitudes. In the notation used in [51] the 3-point amplitudes (3.5.13) and (3.5.14) are classified as $\mathtt{a} = 1$ in the soft limit of a positive helicity vector boson. Therefore a vector boson present in amplitudes which contain at least one of the following other particles: $Z$, $\psi^{a+}$ or $\gamma^+$ has soft weight $\sigma_\gamma = -1$. Using the general formalism developed in [51], we can write down the low-energy theorem of the vector bosons in this subclass of amplitudes

$$\mathcal{A}_{n+1}\left(s_\gamma^+, 1, 2, ..., n\right) \xrightarrow{p_s \to \epsilon p_s \text{ as } \epsilon \to 0} \sum_{k=1}^{n} \frac{[sk]}{\epsilon\langle sk\rangle} \mathcal{A}_n\left(1, 2, ..., \mathcal{F}_+ \cdot k, ..., n\right) + \mathcal{O}\left(\epsilon^0\right). \quad (3.5.25)$$

Here we are using a notation similar to [100] with the introduction of an operator $\mathcal{F}_+$ which acts on the one-particle states as

$$
\begin{array}{|c||c|c|}
\hline
\text{state } i & \mathcal{F}_+ \cdot i & \mathcal{A}_n \text{ prefactor} \\
\hline
Z & \gamma^- & 1 \\
\psi^{1+} & \psi_2^- & -1 \\
\psi^{2+} & \psi_1^- & -1 \\
\gamma^+ & \overline{Z} & -1 \\
\hline
\end{array}
\tag{3.5.26}
$$

and annihilates the states of the negative helicity multiplet. A similar operator $\mathcal{F}_-$ can be defined for the soft limit of a negative helicity vector. Using equation (3.4.13) in conjunction with the soft behavior (3.5.26) of the $n+1$-point amplitude results in the following identity for the residual $n$-point amplitudes

$$
\sum_{i=1}^{n}\sum_{j=1}^{n}(-1)^{L_i+P_i}\frac{[Xi][Yj]}{\langle Yj\rangle}\mathcal{A}_n\left(1,2,...,\mathcal{Q}_1\cdot i,...,\mathcal{F}_+\cdot j,...,n\right)=0\,, \tag{3.5.27}
$$

where here $P_i = 0$ or $1$ corresponds to the additional signs associated with the prefactors of both the supersymmetry Ward identities and the operator $\mathcal{F}_+$ given in Table 3.5.26. Note that the action of $\mathcal{Q}_1$ and $\mathcal{F}_+$ commute on all physical states, so there is no ambiguity when $i = j$ in the sums. Moreover, rearranging the order of the sums, it becomes clear that for each fixed $j$, the sum over $i$ expresses a supersymmetry Ward identity for the $n$-point amplitudes. As such, the identity (3.5.27) does not impose further constraints beyond supersymmetry.

## 3.6 Galileons

Galileon theories are scalar effective field theories (EFTs) with higher derivative self-interactions of the form

$$
\mathcal{L}=\tfrac{1}{2}(\partial\phi)^2+\sum_{n=3}^{D+1}g_n(\partial\phi)^2(\partial\partial\phi)^{n-2}\,, \tag{3.6.1}
$$

where $D$ is the spacetime dimension. The couplings $g_n$ are generally independent. The characteristic feature of these models is that despite the higher derivatives, the equations of motion are only second order. As a consequence, the Galileons have a well-defined classical field theory limit, free from Ostrogradski ghosts. This feature is strongly atypical among EFTs and make Galileons attractive for model building in cosmology and beyond. The cubic Galileon originally arose in the Dvali-Gabadadze-Porrati (DGP) model [101], but Galileons appear in other contexts too, for example in modifications of gravity [87, 102, 103]. Perhaps most significantly, Galileons emerge as subleading terms on effective actions on branes [104]. Here we focus on flat branes in Minkowski space, although other embeddings are also of interest [104–106].

A flat 3-brane placed in a $4{+}1$-dimensional Minkowski bulk will induce a spontaneous breaking of spacetime symmetry: ISO$(4,1) \to$ ISO$(3,1)$. A massless Goldstone mode $\phi$ must appear in the spectrum of the $3{+}1$-dimensional world-volume EFT which is physically identified with fluctuations of the brane into the extra dimension. The full ISO$(4,1)$ symmetry remains a symmetry of the action, and so at leading and next-to-leading order in the derivative expansion the effective action takes the form [104]

$$S = \int \mathrm{d}^4x\sqrt{-G}\left[\Lambda_2^4 + \Lambda_3^3 K[G] + \Lambda_4^2 R[G] + \Lambda_5 \mathcal{K}_{\text{GHY}}[G]\right] , \tag{3.6.2}$$

where $G$ is the pullback of the bulk metric onto the 3-brane world-volume. The leading term with coupling $\Lambda_2^4$ (the brane tension) gives the Dirac-Born-Infeld (DBI) action, while the remaining terms are built from the extrinsic curvature, Ricci curvature and higher-derivative Gibbons-Hawking-York (GHY) (for a bulk Gauss-Bonnet) terms. The $\Lambda_i$ are in general arbitrary mass scales. The resulting action is the *DBI-Galileon* [104]: its leading part is DBI and the subleading terms are cubic, quartic, and quintic in $\phi$.[13] The non-DBI interaction terms in (3.6.2) are the $4+1$-dimensional Lovelock invariants that give the characteristic second-order equations of motion.

The *Galileon models* (3.6.1) correspond to a decoupling limit in which the 3-brane tension $\Lambda_2^4 \to \infty$, but the ratios

$$g_3 = \frac{\Lambda_3^3}{\Lambda_2^6}, \quad g_4 = \frac{\Lambda_4^2}{\Lambda_2^8}, \quad g_5 = \frac{\Lambda_5}{\Lambda_2^{10}} , \tag{3.6.3}$$

are held fixed. The only part of DBI that survives is the canonical kinetic term for $\phi$.[14] The Galileons and the DBI-Galileons both enjoy a non-trivial extended shift symmetry of the form

$$\phi \to \phi + c + b_\mu x^\mu + \dots , \tag{3.6.4}$$

where $c$ is a constant, $b_\mu$ is a constant vector, and $x^\mu$ is the spacetime coordinate. The ellipses stand for possible field-dependent terms, which will not play a role for us here. These symmetries arise from the spontaneously broken symmetry generators [105]: the constant shift from the broken bulk translation and the $x^\mu$-shift from the broken Lorentz rotations.

The quartic Galileon ($g_3 = g_5 = 0$) is sometimes called the *special Galileon* [38, 77] because it has a further enhanced shift symmetry

$$\phi \to \phi + s_{\mu\nu} x^\mu x^\nu + \dots , \tag{3.6.5}$$

[13] The boundary terms $K[G]$ and $\mathcal{K}_{\text{GHY}}[G]$ are only available when the brane is considered an end-of-the-world brane and they are responsible for the odd-powered $\phi$-interactions.

[14] When decoupled from DBI, Galileons violate the null-energy condition and for that reason they have received attention as models for cosmological bounces. However, without the leading DBI terms, the Galileon theories cannot arise as the low-energy limit of a UV complete theory [107].

where the constant tensor $s_{\mu\nu}$ is symmetric and traceless [77]. This is an accidental symmetry that occurs only in the decoupling limit from DBI.

### 3.6.1 Supersymmetric Galileons

In this section, we address the question of supersymmetrization of Galileon theories in $3+1$ dimensions, both in the context of DBI-Galileons and the decoupled Galileons. Note that the fermionic superpartners will not have an extended shift symmetry, but only a shift symmetry of the form $\psi \to \psi + \xi$ with $\xi$ a constant spinor.[15]

Based on the brane construction, one expects that the quartic DBI-Galileon can be supersymmetrized, in particular, there should exist an $\mathcal{N} = 4$ supersymmetrization corresponding to the effective action for a D3-brane in 9+1-dimensional Minkowski space. It is less obvious that supersymmetry would survive the decoupling limit or if the cubic or quintic (DBI-)Galileons can be supersymmetrized. An explicit $\mathcal{N} = 1$ superfield construction of the quartic Galileon was presented in [78]. We will construct an $\mathcal{N} = 1$ quartic Galileon and comment on its uniqueness. In the literature, a supersymmetrization of the cubic Galileon was proposed, but it suffered from ghosts [108]. By a field redefinition, any cubic Galileon is equivalent to the quartic and quintic Galileon with related couplings, so we address supersymmetrization of the cubic Galileon via the quintic.[16]

Before describing our approach, we comment briefly on the super-algebra. The Poincare algebra can be extended [77,112] with the translation generator $C$ ($\delta_C\phi = 1$), the Galileon shift generator $B_\mu$ ($\delta_B\phi = x^\mu$), and the symmetric traceless generator $S_{\mu\nu}$ of the special Galileon transformations (3.6.5). Being agnostic about the origin of a Galileon extension of the super-Poincare algebra, at the minimum we might demand the closure of the extended super-translation sub-algebra with generators $P_\mu$, $Q$, $\bar{Q}$, $C$, and $B_\mu$ (plus $S_{\mu\nu}$ for the special Galileon), as well as a second set of fermionic generators $S$ and $\bar{S}$ associated with spontaneously broken supersymmetry. The latter are required by the algebra. Among the new commutator relations, we must have (schematically)

$$\begin{aligned} &[P_\mu, B_\nu] \sim \eta_{\mu\nu} C\,, \quad [P_\rho, S_{\mu\nu}] \sim \eta_{\mu\rho} B_\nu + \eta_{\nu\rho} B_\mu\,, \\ &[B_\mu, Q] \sim \sigma_\mu(\bar{Q} + \bar{S})\,, \quad [S_{\mu\nu}, Q] = 0\,. \end{aligned} \tag{3.6.6}$$

The last vanishing commutator follows from the fact that $[S_{\mu\nu}, Q]$ must be a linear-combination of fermionic generators, but there are no tensor structures available that can make it symmetric

[15]Fermions with soft behavior $\sigma = 2$ may occur in a fermionic theory whose leading interaction is quartic starting at couplings of mass dimension $-8$; this is subleading to Galileons and therefore not relevant here.

[16]In the literature, one also finds studies of conformal Galileon theories with supersymmetry, see for example [109–111]. Conformal Galileons can be thought of as the subleading terms of the effective action of branes in AdS space and they have rather different properties than the Galileons studied here. For example, the amplitudes have different soft behavior.

and traceless. Now, consider the Jacobi identity

$$[S_{\mu\nu}, \{Q, \bar{Q}\}] = \{[S_{\mu\nu}, Q], \bar{Q}\} + \{Q, [S_{\mu\nu}, \bar{Q}]\}\,. \tag{3.6.7}$$

The RHS vanishes, but using $\{Q, \bar{Q}\} \sim P$ the LHS gives a non-vanishing linear combination of $B_\mu$-generators. Therefore the algebra does not close consistently. This indicates that there is no supersymmetrization of the special Galileon that also preserves the enhanced symmetry (3.6.5). Replacing $S_{\mu\nu}$ by $B_\mu$ in the Jacobi identity (3.6.7) gives $C$ on the LHS. The RHS can match this if $\{Q, S\} \sim C$. There does not appear to be any inconsistency extending the super-translation algebra with the Galileon generators $C$ and $B_\mu$. Indeed, such an algebra follows from the scenario of bulk supersymmetry spontaneously broken to $\mathcal{N} = 1$ on the 3-brane.[17]

These algebraic arguments constrain the form of the symmetry as realized on the classical fields and are suggestive but formally problematic when extended to the quantum theory. In general, spontaneously broken symmetries do not possess well-defined Noether charges as operators on a Hilbert space.[18] As demonstrated in [116], the infinite volume improper integral of the Noether charge density operators of spontaneously broken symmetries do not converge in the weak operator topology. Furthermore, the second $S$-type supersymmetry is necessarily spontaneously broken and satisfies a current algebra with tensor central charges which cannot be integrated to a consistent charge algebra in infinite volume. (See [117] for a related discussion.) It is difficult to draw convincing conclusions from an algebra which formally does not exist.

Nonetheless we will find that the properties suggested by the algebraic arguments do indeed hold as properties of the scattering amplitudes and can be argued for in a mathematically satisfactory way. In the following subsections we will apply the technology of the soft bootstrap to assess the existence of a low-energy S-matrix with the properties expected of a supersymmetric Galileon.

### 3.6.2 Supersymmetric Galileon Bootstrap I

The first step towards supersymmetry is to combine the Galileon with another scalar to form a complex scalar $Z$ of a chiral supermultiplet. We will consider two cases in this chapter. In this section, both the real and imaginary part of $Z$ have the extended shift symmetry (3.6.4). In Sec. 3.6.3, we relax this condition.

Multi-scalar Galileon theories were constructed in [118] from the effective action of a 3-brane in

[17] The decoupling limit (3.6.3) induces an İnönü-Wigner contraction of the original ISO$(4,1)$ symmetry algebra in the direction transverse to the 3-brane. The resulting algebra $\mathfrak{Gal}(4,1)$ is a cousin of the familiar Galilean algebra of non-relativistic mechanics. In the decoupling limit (3.6.3), the extended shift symmetry (3.6.4) arises from the non-linear realization of the coset $\mathfrak{Gal}(4,1)/\text{ISO}(3,1)$. The recent work [113] extends this construction to include the supercharges. An earlier version of the algebra is in [114].

[18] The algebra constructed in [113] is of the former kind. In this case even the classical Poisson algebra will differ from the algebra realized on the fields by the appearance of central terms [115].

a bulk space with $n$ transverse directions. The actions in [118] have $n$ scalars which are Goldstone modes of each of the spontaneously broken translational symmetries and they all have extended shift symmetry (3.6.4). The models inherit $SO(n)$ symmetry from the bulk, in particular there are only even-powered interactions. This may seem to doom a quintic Galileon with more than one scalar; however, we now give evidence for the existence of a complex scalar quintic model that breaks $U(1) = SO(2)$, but still has symmetry (3.6.4) for both real scalars.

To potentially be compatible with supersymmetry, the complex scalar 5-point amplitude must be of the form $A_5(Z\bar{Z}Z\bar{Z}Z)$ (or its conjugate). The coupling has mass-dimension $-9$, and the interaction terms have 8 derivatives. Using that the 5 momenta must satisfy momentum conservation, one finds (using Mathematica to generate the polynomial basis) that there are 10 independent Lorentz-invariant contractions of 8 momenta satisfying Bose symmetry under exchanges of identical states $\{1 \leftrightarrow 3 \leftrightarrow 5\}$ and $\{2 \leftrightarrow 4\}$. Of these, 9 are polynomials of degree 4 in the Mandelstam variables, whereas the 10th is parity odd and proportional to the Levi-Civita symbol.

Imposing the Galileon symmetry in the form of the required $\sigma = 2$ soft behavior of a general linear combination of these 10 basis polynomials selects one unique answer:

$$\begin{aligned} A_5(Z\bar{Z}Z\bar{Z}Z) &= c_1\bigg(\sum_{i<j} s_{ij}^4 - \frac{1}{4}\Big(\sum_{i<j} s_{ij}^2\Big)^2\bigg) \\ &= 48c_1\Big(\epsilon_{\mu\nu\rho\sigma}p_1^\mu p_2^\nu p_3^\rho p_4^\sigma\Big)^2 . \end{aligned} \tag{3.6.8}$$

This amplitude is equal to the real 5-point Galileon, however, at higher-point these models give distinct amplitudes. For example, using soft subtracted recursion relations (which are valid by (3.2.7)) to obtain the 8-point amplitudes involves $\frac{1}{2}\binom{8}{4} = 35$ factorization diagrams for the real scalar case, while for the complex scalar case there are actually two types of 8-point amplitudes, $A_8(Z\bar{Z}Z\bar{Z}Z\bar{Z}Z\bar{Z})$ and $A_8(Z\bar{Z}Z\bar{Z}Z\bar{Z}ZZ)$. The former has 52 diagrams (of two different types) and the latter has 30 diagrams. We have computed these three 8-point amplitudes and verified that they are distinct. We conclude that this is non-trivial evidence in favor of the existence of a 5-point Galileon whose complex scalar has Galileon symmetry (3.6.4). Note that this model necessarily breaks any $U(1)$ symmetry acting on the scalars. Next we show that this quintic model is not compatible with supersymmetry.

One necessary condition for supersymmetry is the Ward identity

$$A_5(Z\bar{Z}Z\bar{\psi}\psi) = -\frac{[25]}{[24]}A_5(Z\bar{Z}Z\bar{Z}Z)\,. \tag{3.6.9}$$

The amplitude on the LHS must come from a local interaction term in the Lagrangian that arises from the supersymmetrization of the five-scalar term. So $A_5(Z\bar{Z}Z\bar{\psi}\psi)$ must be local, i.e. it cannot have any poles.[19] On the other hand, the RHS of (3.6.9) will have a pole when $[24] \to 0$ (when momenta $p_2$ and $p_4$ go collinear), unless $A_5(Z\bar{Z}Z\bar{Z}Z)$ vanishes in that limit. One can explicitly check, using the expression (3.6.8), that it does not. Hence we conclude that the quintic Galileon cannot be supersymmetrized while preserving the Galileon symmetry (3.6.4) for the complex scalar. We will relax the condition of Galileon symmetry in Sec. 3.6.3.

The only possibly non-vanishing 3-scalar amplitude in any theory of massless scalars is constant, i.e. it comes from $\phi^3$, and the resulting higher point amplitudes have singular soft limits. When a 3-particle amplitude vanishes, the associated cubic Lagrangian can be removed by a field redefinition. In particular, the cubic Galileon can be removed by a field redefinition of the form $\phi \to \phi + a(\partial\phi)^2$. This shuffles the information into 4-, 5-, and 6-point interactions. There is no independent 6th order Galileon, so this means the cubic Galileon is equivalent to a particular choice of the quartic and quintic Galileon. This remains true also when there are multiple scalars. In particular, the quintic coupling will be non-zero. From the above, we immediately conclude that the cubic Galileon cannot be supersymmetrized while preserving the Galileon symmetry (3.6.4) for the complex scalar.

Soft subtracted recursion relations with $\sigma_Z = 2$ show that there is a unique complex scalar quartic Galileon whose amplitude is

$$A_4(Z\bar{Z}Z\bar{Z}) = g_4\, stu\,. \tag{3.6.10}$$

Using the supersymmetry Ward identity, all the other 4-point amplitudes are determined in terms of this result. Thus the 4-particle sector is unique. Using the supersymmetry Ward identities, one can show that the soft behavior $\sigma_\psi$ of the fermion in the chiral multiplet is related to the scalar soft behavior as $\sigma_\psi = \sigma_Z$ OR $\sigma_\psi = \sigma_Z - 1$. The recursion relations for $A_6(Z\bar{Z}Z\bar{Z}\psi\bar{\psi})$ are valid in either case (by (3.2.7)). The condition of $a_i$-independence passes for $\sigma_\psi = 1$, but fails for $\sigma_\psi = \sigma_Z = 2$. The result that $\sigma_\psi = 1$ then proves that a supersymmetrization of the quartic Galileon must have a constant shift symmetry for the fermions.

Proceeding, the constructibility criterion (3.2.7) with $\sigma_\psi = 1$ and $\sigma_Z = 2$ shows that only amplitudes with at most a pair of fermions are constructible when based on quartic interactions with coupling dimension $[g_4] = -6$. However, at 6-point order, we can exploit the supersymmetry Ward identities to fully construct all 6-particle amplitudes in the supersymmetric quartic Galileon theory. The supersymmetry Ward identities are

$$\begin{aligned}
[25]A_6(Z\bar{Z}Z\bar{Z}Z\bar{Z}) - [26]A_6(Z\bar{Z}Z\bar{Z}\psi\bar{\psi}) + [24]A_6(Z\bar{Z}Z\bar{\psi}\psi\bar{Z}) &= 0\\
[23]A_6(Z\bar{Z}Z\bar{\psi}\psi\bar{Z}) + [25]A_6(Z\bar{Z}\psi\bar{\psi}Z\bar{Z}) - [26]A_6(Z\bar{Z}\psi\bar{\psi}\psi\bar{\psi}) &= 0\\
[31]A_6(Z\bar{\psi}\psi\bar{\psi}\psi\bar{Z}) + [35]A_6(\psi\bar{\psi}\psi\bar{\psi}Z\bar{Z}) - [36]A_6(\psi\bar{\psi}\psi\bar{\psi}\psi\bar{\psi}) &= 0\,.
\end{aligned}$$

[19]One might worry about contributions from pole diagrams involving Yukawa interactions; however, such terms would give singular soft theorems and are hence not allowed in this setting.

We use the first identity to check that the amplitudes reconstructed with soft subtracted recursion relations are compatible with supersymmetry. The second identity allows us to solve for the 4-fermion amplitude, and with this result the third identity uniquely determines the 6-fermion amplitude. There are three more independent supersymmetry Ward identities: we use them as consistency checks to make sure all the 6-point amplitudes are compatible with the supersymmetry requirements. These checks all pass.

At higher point, the constructible amplitudes with at most two fermions are not sufficient to solve the supersymmetry Ward identities. In a Lagrangian construction, there may therefore be an ambiguity starting at 8th orders in the fields in terms of independently supersymmetrizable operators which must not have any components with two fermions or less; such operators will involve so many derivatives that it is trivial that they can be compatible with the Galileon symmetry (3.6.4) for the scalars and shift symmetry for the fermions.

Notice also that the constructible amplitudes satisfy the conservation of a $U(1)_R$ charge under which only the scalar $Z$ is charged. Such a symmetry is also respected by the supersymmetrization of DBI, but given the ambiguity in the non-constructible amplitudes the strongest statement we can say is that a supersymmetric quartic Galileon *may* be consistent with such a symmetry.

In conclusion, we have found strong evidence for an $\mathcal{N} = 1$ supersymmetrization of the quartic Galileon. It is compatible with a Galileon symmetry (3.6.4) for the complex scalar and shift symmetry for the fermion. It may not be a unique supersymmetrization, as there could be independently supersymmetrizable operators starting at 8th order in fields. A superfield Lagrangian for $\mathcal{N} = 1$ quartic Galileon was presented in [78]. We find that (up to a sign in the Lagrangian) the 4 and 6-point amplitudes computed from [78] agree with ours.

The real scalar amplitudes resulting from (3.6.10) are those of the special Galileon, which has the enhanced shift symmetry (3.6.5). However, this symmetry does not carry over to the complex scalar case (i.e. it is broken by terms mixing the two scalars). This follows from using $\sigma = 3$ in the soft subtracted recursion relations: this construction $A_6(Z\bar{Z}Z\bar{Z}Z\bar{Z})$ fails $a_i$-independence. We conclude that for the quartic Galileon, the special Galileon symmetry (3.6.5) is not compatible with supersymmetry. This is also what the argument based on the algebra indicated.

### 3.6.3 Supersymmetric Galileon Bootstrap II

In this section, we provide evidence for $\mathcal{N} = 1$ supersymmetric quartic and quintic Galileon theories in which the complex scalar $Z = (\phi + i\chi)/\sqrt{2}$ bundles an honest Galileon $\phi$, who enjoys extended shift symmetry (3.6.4), with a second real scalar $\chi$, who only has a constant shift symmetry. The second scalar $\chi$ is naturally identified as an R-axion and a scenario for this type of theory is partial supersymmetry breaking.

We start by writing the most general Ansatz for the amplitudes $A_5(Z\bar{Z}Z\bar{Z}Z)$, $A_5(Z\bar{Z}Z\bar{\psi}\psi)$, $A_5(Z\bar{\psi}\psi\bar{\psi}\psi)$ and their complex conjugates. All other amplitudes must be zero for a theory compatible with supersymmetry. On this Ansatz of 122 free parameters we impose the following constraints:

- Compatibility with supersymmetry via the supersymmetry Ward identities

$$\begin{aligned} A_5(Z\bar{Z}Z\bar{Z}Z) &= -\frac{[24]}{[25]}A_5(Z\bar{Z}Z\bar{\psi}\psi) = \frac{[24]}{[35]}A_5(Z\bar{\psi}\psi\bar{\psi}\psi)\,, \\ A_5(\bar{Z}Z\bar{Z}Z\bar{Z}) &= -\frac{\langle 24\rangle}{\langle 25\rangle}A_5(\bar{Z}Z\bar{Z}\psi\bar{\psi}) = \frac{\langle 24\rangle}{\langle 35\rangle}A_5(\bar{Z}\psi\bar{\psi}\psi\bar{\psi})\,. \end{aligned} \tag{3.6.11}$$

- A shift symmetry for the complex field $Z$ in the form of $\sigma = 1$ soft behavior for the amplitudes of $Z$.

- Galileon symmetry for the real scalar field $\phi$ in the form of $\sigma = 2$ soft behavior imposed on the linear combinations of complex-scalar amplitudes,

$$A_5(\phi\cdots) = \frac{1}{\sqrt{2}}\left(A_5(Z\cdots) + A_5(\bar{Z}\cdots)\right)\,.$$

Imposing these constraints on our Ansatz left us with a 3-parameter family of solutions. Interestingly, this solution comes with a "free" $\sigma = 1$ soft behavior for the fermions, that suggests that the theory is invariant under a shift of the fermions. Moreover, the five-Galileon amplitude $A_5(\phi\phi\phi\phi\phi)$ matches the known real Galileon amplitude.

In order to further constrain the solution, we consider the 7-point amplitudes of the DBI-Galileon theory. The leading order contribution to these amplitudes is proportional to the product of the DBI coupling with mass dimension $-4$ and the the quintic Galileon coupling with mass dimension $-6$; it can be reconstructed using subtracted soft recursion relations with $\sigma_\phi = 2$, $\sigma_\chi = 1$ and $\sigma_\psi = 1$ if $2n_\chi + n_f < 8$, where $n_\chi$ is the number of $\chi$-external states and $n_f$ is the the number of fermionic external states. Demanding that the results of recursion are independent of the shift parameters $a_i$ *uniquely* fixes the parameters of our solution. The resulting scalar amplitude is

$$\begin{aligned} A_5(Z\bar{Z}Z\bar{Z}Z) = s_{24}\Big[&6s_{24}s_{25}s_{45} + \Big(4s_{12}s_{23}s_{45} + 2s_{12}s_{24}s_{34} \\ &+2s_{25}^2s_{45} + s_{24}s_{25}^2 + (2\leftrightarrow 4)\Big) + (1\leftrightarrow 5) + (3\leftrightarrow 5)\Big] - 4s_{24}^4\,, \end{aligned} \tag{3.6.12}$$

while the amplitudes with fermions can be straightforwardly obtained from the supersymmetry Ward identities (3.6.11).

To conclude this section, we find strong evidence for the existence of a supersymmetrization of the quintic Galileon. In this theory, only one of the two scalar modes enjoys the full Galileon symmetry, while the second one, an R-symmetry axion has only a shift symmetry. A very similar analysis can be carried out for the quartic Galileon. The 4-point scalar amplitude has two independent terms, $A_4(Z\bar{Z}Z\bar{Z}) = r_1\, stu + r_2\, t^3$. When $r_2 = 0$, we recover the quartic Galileon (3.6.10) which has $\sigma = 2$ for the complex scalar. Computing all constructible 6-point amplitudes in both the decoupled Galileon and DBI-Galileon places no constraints on the couplings $r_1$ and $r_2$. This is evidence that there may exist a 2-parameter family of quartic $\mathcal{N} = 1$ supersymmetric Galileons in which the complex scalar is composed of a Galileon and an R-axion.

### 3.6.4 Vector-Scalar Special Galileon

It is known that scalar Galileon theories arise in certain limits of massive gravity [86, 87] (for a review, see [88]). An on-shell massive graviton in 4d has 5 polarization states and the decoupling limit gives one real massless scalar (the Galileon) and a massless photon in addition to the massless graviton. So we expect there to be an EFT of a real Galileon scalar coupled to vector.[20] The vector couples quadratically to the scalar and was consistently truncated in [87]. Some subsequent studies have discussed the photon-scalar coupling of Galileons, see for example [119]. Here, we use soft recursion to give some definitive results about the possible scattering amplitudes in such a theory.

If the scalar has $\sigma_\phi = 2$, only the scalar amplitudes are constructible, and we are not able to say anything about the vector sector and its couplings to the scalar. If however the couplings are tuned in such a way that the cubic and quintic Galileon interactions are set to zero then in the scalar sector the soft weight of the scalar is enhanced to $\sigma_\phi = 3$, the *special Galileon* scenario. At present it is unknown whether this enhancement of symmetry can be understood in some natural way from the decoupling limit of some model of massive gravity. Moreover, it is not a priori clear if the $\sigma_\phi = 3$ enhancement can survive coupling to other particles.

We use the power of the soft bootstrap to construct the most general amplitudes consistent with the special Galileon low-energy theorem. We use the 6-point test to exclude EFTs with a special Galileon coupled non-trivially to a photon with $\sigma_\gamma > 0$. For the model with $\sigma_\phi = 3$ and $\sigma_\gamma = 0$, we find that the soft recursion 6-point test reduces the most general 6 real-parameter ansatz for the scalar and scalar-vector interactions to a 3 real-parameter family:

$$\begin{aligned}
\mathcal{A}_4(1_\phi\, 2_\phi\, 3_\phi\, 4_\phi) &= g_1 stu\,, \\
\mathcal{A}_4(1_\phi\, 2_\phi\, 2^+_\gamma\, 4^+_\gamma) &= g_2[34]^2\left(t^2+u^2+3tu\right)\,, \\
\mathcal{A}_4(1^-_\gamma\, 2_\phi\, 3_\phi\, 4^+_\gamma) &= g_1\langle 12\rangle[24]\langle 13\rangle[34]u\,, \\
\mathcal{A}_4(1_\phi\, 2_\phi\, 3^-_\gamma\, 4^-_\gamma) &= g_2^*\langle 34\rangle^2\left(t^2+u^2+3tu\right)\,.
\end{aligned} \tag{3.6.13}$$

[20] The decoupling of these interactions from the graviton is not clear [88].

The couplings of the pure vector sector are unconstrained; the most general ansatz is

$$
\begin{aligned}
&\mathcal{A}_4(1^+_\gamma\, 2^+_\gamma\, 3^+_\gamma\, 4^+_\gamma) = g_3\Big([12]^2[34]^2 s + [13]^2[24]^2 t + [14]^2[23]^2 u\Big)\,,\\
&\mathcal{A}_4(1^-_\gamma\, 2^-_\gamma\, 3^+_\gamma\, 4^+_\gamma) = g_4\langle 12\rangle^2[34]^2 s\,,\\
&\mathcal{A}_4(1^-_\gamma\, 2^-_\gamma\, 3^-_\gamma\, 4^-_\gamma) = g_3^*\Big(\langle 12\rangle^2\langle 34\rangle^2 s + \langle 13\rangle^2\langle 24\rangle^2 t + \langle 14\rangle^2\langle 23\rangle^2 u\Big)\,.
\end{aligned}
\tag{3.6.14}
$$

The most interesting feature of the above result is the relation between the coefficients of the amplitudes $\mathcal{A}_4(1_\phi\, 2_\phi\, 3_\phi\, 4_\phi)$ and $\mathcal{A}_4(1^-_\gamma\, 2_\phi\, 3_\phi\, 4^+_\gamma)$. The former is the familiar quartic Galileon, while the latter would arise from an operator of the form

$$
\mathcal{O} \sim g_1(\partial_\mu F_+^{\alpha\beta})(\partial^\mu F_-^{\dot\alpha\dot\beta})(\sigma^\nu_{\alpha\dot\alpha}\partial_\nu\phi)(\sigma^\rho_{\beta\dot\beta}\partial_\rho\phi),
\tag{3.6.15}
$$

where $F_\pm$ are as defined in and below (3.2.17)

The relation between the couplings strongly indicates the existence of a non-linear symmetry which mixes the scalar and vector modes. Describing the action of this symmetry and its consequences is left for future work.

### 3.6.5 Higher Derivative Corrections to the Special Galileon

The real quartic Galileon has low-energy theorems with $\sigma = 3$ soft weight. Being agnostic about the origin of the special Galileon, from an EFT perspective, one should write a Lagrangian with all possible operators that respect the symmetries of the theory in a derivative expansion. The authors of [120] found that among a specific subclass of Lagrangian operators, namely those with the schematic form $\partial^4\phi^4$, $\partial^6\phi^4$ and $\partial^8\phi^5$, the special Galileon is the unique choice that can give enhanced soft limits with $\sigma = 3$ soft weight. In this section, we investigate much more exhaustively the possible higher-derivative quartic and quintic operators compatible with $\sigma = 3$ soft behavior. This is done using soft-subtracted recursion relations to calculate the 6- and 7-point scattering amplitudes of the model.

Let us start our discussion with the 6-point case. The constructibility criterion (3.2.18) implies that recursion relations are valid if the coupling constant $g_6$ of the 6-point amplitude satisfies

$$
[g_6] > -20\,.
\tag{3.6.16}
$$

Given that this coupling is the product of two quartic couplings and that the leading order quartic coupling has mass dimension $-6$ recursion relations can probe contributions to the 4-point amplitude with mass dimension in the range

$$
-14 < [g_4] \leq -6\,.
\tag{3.6.17}
$$

Taking into account Bose symmetry, the most general ansatz one can write down for the 4-point matrix element of local operators is

$$\begin{aligned}\mathcal{A}_4(1_\phi 2_\phi 3_\phi 4_\phi) =& \frac{c_0}{\Lambda^6} stu \\ &+ \frac{c_1}{\Lambda^8}\left(s^4+t^4+u^4\right) \\ &+ \frac{c_2}{\Lambda^{10}}\left(s^5+t^5+u^5\right) \\ &+ \frac{1}{\Lambda^{12}}\left(c_3\left(s^6+t^6+u^6\right)+c_3' s^2t^2u^2\right)+\mathcal{O}(\Lambda^{-14})\,.\end{aligned} \tag{3.6.18}$$

The leading term with coupling $c_0/\Lambda^6$ is the usual quartic Galileon. The terms suppressed by higher powers of the the UV cutoff $\Lambda$ encode all possible higher-derivative quartic operators of the scalar field up to order $\Lambda^{-14}$.

We apply the 6-point test with $\sigma = 3$ and find that consistency requires $c_1 = c_3 = 0$ in the ansatz (3.6.18). The 4-point amplitude then becomes

$$\mathcal{A}_4(1_\phi\, 2_\phi\, 3_\phi\, 4_\phi) = \frac{c_0}{\Lambda^6} stu + \frac{c_2}{\Lambda^{10}}\left(s^5+t^5+u^5\right) + \frac{c_3'}{\Lambda^{12}} s^2t^2u^2 + \mathcal{O}(\Lambda^{-14})\,. \tag{3.6.19}$$

From this, we understand that there cannot exist an 8-derivative Lagrangian operator that preserves the special Galileon symmetry. Additionally, at 6-, 10- and 12-derivative order there exist unique quartic operators compatible with $\sigma = 3$. In Section 3.6.6, we show explicitly that the result (3.6.19) can also be obtained from an application of the BCJ double-copy.

Next we examine the possible existence of quintic operators compatible with $\sigma = 3$. We combine input from the quartic Galileon with the most general possible ansatz for the 5-point matrix elements and use the 7-point test to assess compatibility with $\sigma = 3$. The soft subtracted recursion relations at 7 points are valid if

$$[g_7] > -24\,. \tag{3.6.20}$$

Since the 7-point coupling constant is the product of a quartic (with mass dimension $-6$ or lower) and a quintic coupling, the latter must then satisfy

$$[g_5] > -18\,. \tag{3.6.21}$$

With Bose symmetry and the requirement that the ansatz for the 5-point amplitude must have soft weight $\sigma = 3$, we are left with

$$\mathcal{A}_5(1_\phi\, 2_\phi\, 3_\phi\, 4_\phi\, 5_\phi) = \frac{d_1}{\Lambda^{15}}\epsilon(1234)\sum_P(-1)^{|P|}s_{P_1P_2}s_{P_2P_3}s_{P_3P_4}s_{P_4P_5}s_{P_5P_1} \tag{3.6.22}$$

$$+\frac{1}{\Lambda^{17}}\Bigg[d_2\,\epsilon(1234)^4 + d_3\epsilon(1234)\sum_P(-1)^{|P|}s_{P_1P_2}s^2_{P_2P_3}\left(s^2_{P_2P_3}s_{P_3P_4} - s^2_{P_1P_2}s_{P_2P_4}\right)$$

$$+d_4\left(\frac{4}{5}\sum_{i<j}s^3_{ij}\sum_{i<j}s^5_{ij} + \sum_{i<j}\sum_{k\neq i,j}\left(20s^2_{ij}s^3_{ik}s^3_{jk} + 9s^4_{ij}s^2_{ik}s^2_{jk} - 2s^6_{ij}s_{ik}s_{jk}\right)\right)\Bigg] + \mathcal{O}(\Lambda^{-19})\,.$$

In the above, $\epsilon(1234) = \epsilon_{\mu\nu\rho\sigma}p_1^\mu p_2^\nu p_3^\rho p_4^\sigma$, the sum $\sum_{i<j}$ means $\sum_{i=1}^4\sum_{j=i+1}^5$, while the sum $\sum_P$ is over all permutations of $\{1,2,3,4,5\}$, $(-1)^{|P|}$ is the signature of the permutation and $P_i$ is its $i$th element. There are no contributions to the amplitude that have less than 14 derivatives. The $1/\Lambda^{14}$-term satisfies the constructibility criterion and vanishes in 3d kinematics, in agreement with the discussion of Section 3.2.4. Two of the $1/\Lambda^{17}$-terms also vanish in 3d kinematics, but this was not a priori expected since they are too high order to satisfy constructibility.

The 7-point test implies no constraints on the coefficients $d_1$, $d_2$, $d_3$ and $d_4$. This is evidence in favor of the existence of four 5-point operators that preserve the special Galileon symmetry. Next, in Section 3.6.6, we investigate whether this result can be obtained from a double-copy prescription, similar to the 4-point case.

### 3.6.6 Comparison with the Field Theory KLT Relations

The significance of the special Galileon extends well beyond the contraction limit of the 3-brane effective field theory and the decoupling limit of massive gravity. The enhancement of the soft behavior to $\sigma = 3$ (which degenerates to $\sigma = 2$ when the DBI interactions are re-introduced) or correspondingly the extension of the non-linearly realized symmetry algebra suggests that this model has a fundamental significance of its own that is at present only partially understood. Perhaps one of the deepest and least understood aspects of the special Galileon is its role in the (field theory) *KLT algebra* as the product of two copies of the $\frac{U(N)\times U(N)}{U(N)}$ non-linear sigma model. For $N = 2, 3$ this coset sigma model has been intensively studied as a phenomenological model of the lightest mesons under the name *Chiral Perturbation Theory* ($\chi$PT). Henceforth we will use this name to avoid confusion with the $\mathbb{CP}^1$ non-linear sigma model discussed in Section 3.5.

The *double-copy* relation between $\chi$PT and the special Galileon was first understood in the CHY auxilliary world-sheet formalism [121]. Specifically, it was shown in the CHY formalism that the *leading order* contribution to scattering in the special Galileon model can be obtained from the

KLT product

$$\mathcal{A}_n^{\text{sGal}} = \sum_{\alpha,\beta} \mathcal{A}_n^{\chi\text{PT}}[\alpha] S_{\text{KLT}}[\alpha|\beta] \mathcal{A}_n^{\chi\text{PT}}[\beta] \,, \tag{3.6.23}$$

where $\alpha, \beta$ index the $(n-3)!$ independent color(flavor)-orderings.[21] The KLT kernel $S_{\text{KLT}}[\alpha|\beta]$ is universal in the sense that the explicit form of the relations (3.6.23) are identical to the perhaps more familiar field theory KLT relations giving a double-copy construction of Einstein-dilaton-$B_{\mu\nu}$ gravity from two copies of Yang-Mills theory. Concretely, the first few relations have the form

$$\begin{aligned}
\mathcal{A}_4^{\text{sGal}}(1,2,3,4) &= -s_{12}\mathcal{A}_4^{\chi\text{PT}}[1,2,3,4]\,\mathcal{A}_4^{\chi\text{PT}}[1,2,4,3] \,, \\
\mathcal{A}_5^{\text{sGal}}(1,2,3,4,5) &= s_{23}s_{45}\mathcal{A}_5^{\chi\text{PT}}[1,2,3,4,5]\,\mathcal{A}_5^{\chi\text{PT}}[1,3,2,5,4] + (3 \leftrightarrow 4) \,, \\
\mathcal{A}_6^{\text{sGal}}(1,2,3,4,5,6) &= -s_{12}s_{45}\mathcal{A}_6^{\chi\text{PT}}[1,2,3,4,5,6]\Big(s_{35}\mathcal{A}_6^{\chi\text{PT}}[1,5,3,4,6,2] \\
&\qquad +(s_{34}+s_{35})\mathcal{A}_6^{\chi\text{PT}}[1,5,4,3,6,2]\Big) + \mathcal{P}(2,3,4) \,,
\end{aligned} \tag{3.6.24}$$

where $\mathcal{P}(2,3,4)$ denotes the sum of all permutations of legs 2, 3 and 4.

For the formulae (3.6.23) and (3.6.24) to even be well-defined, the color-ordered amplitudes on the right-hand-side must satisfy a number of non-trivial relations to reduce the number of independent partial amplitudes to $(n-3)!$ for the scattering of $n$ particles. The existence of a color-ordered representation is itself non-trivial and not guaranteed to be satisfied in all models with color structure [122]. In all known cases where the double-copy relations (3.6.23) give a sensible, physical output, the reduction to a reduced basis of size $(n-3)!$ is accomplished by two sets of identities among the partial amplitudes, namely the *Kleiss-Kuijf* and *fundamental Bern-Carrasco-Johansson* relations. That these identites obtain for amplitudes calculated in the leading two-derivative action of $\chi$PT was first established in [123] using semi-on-shell recursion techniques developed in [124].

Our goal in this section is to connect two (possibly discrepant) definitions of the special Galileon model:

1. The special Galileon is the most general effective field theory of a real massless scalar with $\sigma = 3$ vanishing soft limits.

2. The special Galileon is the double-copy of two copies of $\chi$PT.

What we have described above is the known fact that these definitions agree at the lowest non-trivial order. In the previous section we used soft subtracted recursion to construct the most general 4- and 5-point amplitudes consistent with the first definition up to order $\Lambda^{-12}$ and $\Lambda^{-17}$ respectively. To determine if these results agree with the second definition we must first construct the most general 4- and 5-point amplitudes in $\chi$PT compatible with the requirements of

[21] We use square brackets for the arguments of a color-ordered amplitude.

the double-copy. Here we are following the approach of [122] and making the most conservative possible assumptions. Specifically we assume that both the explicit form of the double-copy (3.6.24) *and* the relations the amplitudes must satisfy to reduce the basis of partial amplitudes to size $(n-3)!$ are *identical* to what is required at leading order.

Let us begin with the 4-point amplitudes. The relations we impose are cyclicity (C)

$$\mathcal{A}_4^{\chi\text{PT}}[1,2,3,4] = \mathcal{A}_4^{\chi\text{PT}}[2,3,4,1]\,, \tag{3.6.25}$$

Kleiss-Kuijf (KK) or $U(1)$-decoupling

$$\mathcal{A}_4^{\chi\text{PT}}[1,2,3,4] + \mathcal{A}_4^{\chi\text{PT}}[2,1,3,4] + \mathcal{A}_4^{\chi\text{PT}}[2,3,1,4] = 0\,, \tag{3.6.26}$$

and the fundamental BCJ relation

$$(-s-t)\mathcal{A}_4^{\chi\text{PT}}[1,2,3,4] - t\mathcal{A}_4^{\chi\text{PT}}[1,2,4,3] = 0\,. \tag{3.6.27}$$

Since there are no additional quantum number labels in the partial amplitudes, at each order the 4-point amplitude is determined by a single polynomial function of the available Lorentz singlets

$$\mathcal{A}_4^{\chi\text{PT}}[1,2,3,4] = F^{(0)}(s,t) + \frac{1}{\Lambda^2}F^{(2)}(s,t) + \frac{1}{\Lambda^4}F^{(4)}(s,t) + \ldots \tag{3.6.28}$$

The superscript $k$ counts both the mass dimension of the function and the number of derivatives in the underlying effective operator. In this language, the double-copy-compatibility conditions take the form

$$\begin{aligned}
&\text{C:} && F^{(k)}(s,t) = F^{(k)}(-s-t,t)\,,\\
&\text{KK:} && F^{(k)}(s,t) + F^{(k)}(s,-s-t) + F^{(k)}(-s-t,s) = 0\,,\\
&\text{BCJ:} && (-s-t)F^{(k)}(s,t) - tF^{(k)}(s,-s-t) = 0\,.
\end{aligned} \tag{3.6.29}$$

We make a general parametrization of the polynomial functions as

$$\begin{aligned}
F^{(0)}(s,t) &= c_1^{(0)},\\
F^{(2)}(s,t) &= c_1^{(2)}s + c_2^{(2)}t,\\
F^{(4)}(s,t) &= c_1^{(4)}s^2 + c_2^{(4)}st + c_3^{(4)}t^2,\\
F^{(6)}(s,t) &= c_1^{(6)}s^3 + c_2^{(6)}s^2t + c_3^{(6)}st^2 + c_4^{(6)}t^3,\\
F^{(8)}(s,t) &= c_1^{(8)}s^4 + c_2^{(8)}s^3t + c_3^{(8)}s^2t^2 + c_4^{(8)}st^3 + c_5^{(8)}t^4,
\end{aligned} \tag{3.6.30}$$

and so on. Imposing the conditions (3.6.29) gives a system of linear relations among the coefficients $c_i^{(k)}$. These are straightforward to solve and give

$$\mathcal{A}_4^{\chi\mathrm{PT}}[1,2,3,4] = \frac{g_2}{\Lambda^2}t + \frac{g_6}{\Lambda^6}t(s^2+t^2+u^2) + \frac{g_8}{\Lambda^8}t(stu) + \ldots \tag{3.6.31}$$

A few comments about this result. As expected, the leading 2-derivative contribution is compatible with the conditions (3.6.29). Surprisingly, there are no compatible contributions from 4-derivative operators, but there are unique contributions at 6- and 8-derivative order. Moreover, the structure of the result here agrees with the 4-point amplitude of *Abelian Z-theory* [125]. The Z-theory model is a top-down construction which gives open string scattering amplitudes as the field theory double-copy of Yang-Mills and a higher-derivative extension of $\chi$PT. The Z-amplitudes are by construction guaranteed to satisfy the double-copy-compatibility conditions but with Wilson coefficients $g_i$ having precise values calculated from the known string amplitudes. The method of this section can be understood as the bottom-up converse of the Z-theory construction, and at 4-point we find agreement.

To summarize, we have shown that up to 8-derivative order there is a 3-parameter family of operators that generate 4-point matrix elements compatible with the conditions required for the double-copy to be well-defined. We could continue this to higher order, but our ability to compare with the methods of Section 3.6.5 are bounded above at this order by the constructibility criterion.

To construct the associated amplitudes in the special Galileon model (according to the second definition described above) we use the first relation in (3.6.24). The result is

$$\mathcal{A}_4^{\mathrm{sGal}}(1,2,3,4) = \frac{c_1}{\Lambda^6}stu + \frac{c_2}{\Lambda^{10}}\left(s^5+t^5+u^5\right) + \frac{c_3}{\Lambda^{12}}s^2t^2u^2 + \ldots\,, \tag{3.6.32}$$

in precise agreement with the special Galileon amplitude (3.6.19).

As an additional check to the results obtained above, we calculate the 6-point amplitudes of both $\chi$PT and the special Galileon. Up to order $\mathcal{O}(\Lambda^{-6})$ the $\chi$PT amplitude can be calculated using soft subtracted recursion with (3.6.31) as input. Note that only three factorization channels contribute to this calculation because the rest do not preserve color ordering. The resulting amplitude,

$$\mathcal{A}_6^{\chi\mathrm{PT}}[1,2,3,4,5,6] = \frac{g_2^2}{\Lambda^4}\left[\frac{s_{13}s_{46}}{p_{123}^2} + \frac{s_{24}s_{15}}{p_{234}^2} + \frac{s_{35}s_{26}}{p_{345}^2} - s_{246}\right] + \mathcal{O}(\Lambda^{-8})\,, \tag{3.6.33}$$

satisfies all C, KK and BCJ constraints. Contributions subleading to the ones listed above do not satisfy the constructibility criterion (3.2.18) and cannot be calculated using soft subtracted recursion. However, we were able to uniquely determine them up to order $\mathcal{O}(\Lambda^{-10})$, by demanding that they have the correct pole structure, consistent with unitarity and locality, have $\sigma = 1$ soft weight and satisfy C, KK and BCJ conditions. The result of this calculation is listed in (B.3.2).

We are now in position to calculate the 6-point special Galileon amplitude with two different methods. We can either use the 6-point KLT relation in (3.6.24) or use soft subtracted recursion with (3.6.32) as input. The results of these calculations match perfectly up to order $\mathcal{O}(\Lambda^{-18})$, which is the furthest the recursive calculation can go.

Shifting our focus to 5-point amplitudes, we find that it is *not* possible to reproduce (3.6.22) as a double-copy of two (identical or non-identical) color-ordered scalar amplitudes, despite the perfect agreement at 4- and 6-points. Starting from a general ansatz for the scalar color-ordered amplitude, we find that the leading contribution that satisfies all C, KK and BCJ constraints is $\mathcal{O}(\Lambda^{-15})$ corresponding to a valence 5 scalar-field operator with 14 derivatives. The existence of such an operator at all is interesting since there are apparently *no odd point amplitudes in Z-theory* [125]! At this order we find that the kinematic structure of Z-theory does not coincide with the most general possible double-copy-compatible higher-derivative extension of $\chi$PT. Or perhaps said differently, just like string theory fixes the Wilson coefficients in the 4-point result (3.6.31) to take particular (non-zero) values, it appears to fix the Wilson coefficients of the odd-point amplitudes to be zero.

When we use the second relation of (3.6.24) with this result, we obtain a 5-point scalar amplitude of order $\mathcal{O}(\Lambda^{-33})$, which is significantly subleading to the amplitude (3.6.22) we calculated in the previous section for the special Galileon.

# CHAPTER 4

# Born-Infeld and Electromagnetic Duality at One-Loop

## 4.1 Review of Born-Infeld Electrodynamics

The Born-Infeld model of non-linear electrodynamics is a low-energy effective field theory of central importance in theoretical physics. Introduced long ago as an (ultimately misguided) proposed classical solution to the electron self-energy problem [126], it subsequently reappeared as the low-energy effective description of world-volume gauge fields on D-branes [127–129]. Independently of this *stringy* characterization, the Born-Infeld model has proven to be a truly exceptional example of a low-energy effective theory of non-linear electrodynamics, though perhaps at times a mysterious one.

As a classical field theory in $d = 4$ the Born-Infeld model can be described by the effective action

$$S_{\text{BI}} = -\Lambda^4 \int \mathrm{d}^4 x \left[ \sqrt{-\det \left( g_{\mu\nu} + \frac{1}{\Lambda^2} F_{\mu\nu} \right)} - 1 \right], \tag{4.1.1}$$

where $\Lambda$ is the characteristic scale in the problem. In the D-brane picture, $\Lambda$ is related to the brane tension.

Low-energy scattering of light-by-light in the Born-Infeld model can be calculated as a perturbative expansion in $1/\Lambda$. The tree-approximation to these scattering amplitudes has been a subject of interest recently in the context of modern on-shell approaches to quantum field theory. For example, in [1] two novel on-shell approaches for calculating 4d tree-level Born-Infeld amplitudes were given: by imposing *multi-chiral* low-energy theorems derived from supersymmetric relations with Goldstone fermions, and from T-duality constraints under dimensional reduction. Also very striking is the discovery in [121], in the context of the CHY formulation of the tree-level S-matrix, that the KLT formula relating Yang-Mills (YM) and gravity amplitudes also gives Born-Infeld tree amplitudes if one of the gauge theory factors is replaced with the flavor-ordered

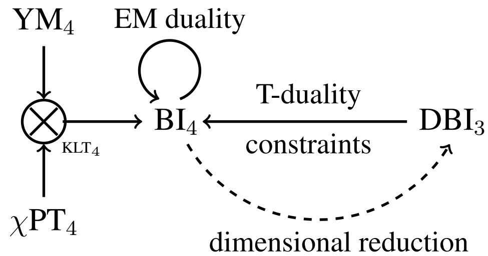



Figure 4.1: Some key-properties of BI amplitudes at tree-level, in particular the double-copy construction and 4d electromagnetic duality. The idea behind the T-duality constraint [1] is that when dimensionally reduced along one direction, a linear combination of the photon polarizations become a scalar modulus of the compactified direction,. i.e. it is the Goldstone mode of the spontaneously broken translational symmetry and as such it must have enhanced $\mathcal{O}(p^2)$ soft behavior.

amplitudes of Chiral Perturbation Theory ($\chi$PT):

$$\mathrm{BI}_d = \mathrm{YM}_d \otimes_{\mathrm{KLT}} \chi\mathrm{PT}_d. \tag{4.1.2}$$

The subscript $d$ indicates the spacetime dimensions of these theories. What all of these discoveries make clear is that there is an enormous amount of structure hidden behind the action (4.1.1) which may be leveraged to make possible previously unattainable calculations. It should also be noted that Born-Infeld plays a central role in the ever growing web of mysterious connections between gauge theories, gravity theories, and EFTs in diverse dimensions [130, 131]. Also of great relevance in this chapter, pure Born-Infeld can be defined as a consistent truncation of $\mathcal{N} > 1$ supersymmetrizations of Dirac-Born-Infeld theory.

The tree amplitudes in 4d Born-Infeld theory exhibit an important and interesting feature: they vanish unless the external states have an equal number of positive and negative helicity states. This is the on-shell manifestation of electromagnetic duality of the classical theory in 4d. In particular, the 4-particle tree amplitude[1] is

$$\mathcal{A}_4^{(\text{tree})\ \mathrm{BI}_4}(1_\gamma^+, 2_\gamma^+, 3_\gamma^-, 4_\gamma^-) = \frac{1}{\Lambda^4}[12]^2\langle 34\rangle^2\,, \tag{4.1.3}$$

while all other helicity configurations vanish. Note that the emergence of electromagnetic duality is highly non-trivial in the double-copy construction (4.1.2). Some of the key properties of the BI tree amplitudes are summarized in Figure 4.1.

The recent progress in Born-Infeld scattering has so far been restricted to tree-level amplitudes. Given the development of powerful unitarity based methods for recycling trees into loops [13],

[1]Compared to the action (4.1.1), we have rescaled $\Lambda^4 \to \Lambda^4/2$, such that the 4-point amplitude has coupling $1/\Lambda^4$.

there is every reason to believe that interesting structures are waiting for us in the loop amplitudes. In this context almost nothing is known.[2] There are good reasons for this; the calculations in Born-Infeld electrodynamics at one-loop are challenging, in ways that are importantly different from superficially similar calculations in perturbative quantum gravity. Similar to calculations at one-loop using Feynman rules derived from expanding the linearized Einstein-Hilbert action, the first computational bottleneck in Born-Infeld is given by the problem of determining the off-shell vertex factors for the interaction terms given by expanding (4.1.1)

$$S_{\rm BI} \sim \int \mathrm{d}^4 x \left[ F^2 + \frac{c_1}{\Lambda^4} F^4 + \frac{c_2}{\Lambda^8} F^6 + \ldots \right] . \tag{4.1.4}$$

As the multiplicity of external states increases, more and more terms in this expansion must be kept, and so an ever growing list of increasingly long vertex factors must be calculated. At multiplicity $n$, operators of the form $F^n$ will contribute; with vertex factors given as sums over permutations growing exponentially in $n$. Beyond the lowest multiplicity, calculating such an amplitude by hand is almost unthinkable, and even with state-of-the art computing power one soon hits a hard wall when performing such a brute force calculation. The situation here is a little different from perturbative gravity. In gravity, the vertex factors are not independent since they are not separately gauge invariant; the higher-point interactions are in principle completely determined by locality and Lorentz invariance by the three-particle ones. This can have dramatic consequences, for example in [133] all-multiplicity, rational one-loop results are obtained from the lowest multiplicity results by enforcing the correct collinear and soft limits. In Born-Infeld, however, these higher-valence operators are genuinely gauge invariant physical operators, the associated Wilson coefficients are not related by any inviolable field theory principle and must instead be fixed by imposing additional physical constraints. No analysis of soft or collinear limits could possibly determine the all-multiplicity one-loop amplitudes in Born-Infeld, unless it incorporated additional physical information beyond Lorentz invariance and locality.

The second computational bottleneck occurs when evaluating the required loop integrals. Even if the required loop integrands can be constructed, we still have to integrate the resulting expressions. Operators of the form $F^n$ are $n$-derivative operators and the associated vertex factors have $n$ powers of momentum. The resulting loop-integrands therefore involve tensors with ranks that grow larger and larger with the multiplicity. This is unlike gravity that only has two-derivative interactions. Attempting to apply traditional Passarino-Veltman reduction algorithms to such high-rank tensor expressions again quickly leads to a confrontation with the limits of computing power. Such a direct calculation is primarily limited by the fact that the method of Feynman diagrams is completely general. It therefore makes no use of any of the aforementioned properties

[2] One of the few explicit calculations is the determination of the cut-constructible part of the 4-point MHV amplitude in $\mathcal{N}=4$ $\mathrm{DBI}_4$ in [132].

that make Born-Infeld electrodynamics exceptional. For example, such an approach would be equally well-suited to calculating loop corrections in the Euler-Heisenberg effective theory [134], another well-studied example of a model of non-linear electrodynamics.

In this chapter, we initiate a study of 4d non-supersymmetric Born-Infeld theory at the loop-level. We use modern on-shell methods (supersymmetric decomposition, double-copy, T-duality…) that are specialized to the particular properties of Born-Infeld and to the objects we compute. We derive results that would be impossible to obtain with traditional methods. Specifically, we derive all-multiplicity results for the one-loop amplitudes in the *self-dual* (SD) and *next-to-self-dual* (NSD) sectors of 4d non-supersymmetric Born-Infeld:

$$\mathcal{A}_n^{\text{SD}}\left(1_\gamma^+,\, 2_\gamma^+,\, \ldots\, (n-1)_\gamma^+,\, n_\gamma^+\right) \quad \text{and} \quad \mathcal{A}_n^{\text{NSD}}\left(1_\gamma^+,\, 2_\gamma^+,\, \ldots\, (n-1)_\gamma^+,\, n_\gamma^-\right). \tag{4.1.5}$$

Any 4d cuts of these amplitudes vanish, hence to obtain them $d$-dimensional unitarity is used and the results are necessarily rational functions of the external momenta.

One motivation for these calculations is to examine the fate of electromagnetic duality at loop-level in pure Born-Infeld theory. We make some observations at the end of the chapter, but otherwise this will be the subject of a forthcoming paper.

## 4.2 Overview of Method

Our goal in this chapter is to calculate SD and NSD one-loop amplitudes in non-supersymmetric Born-Infeld in $d = 4$. As discussed in Section 4.1 instead of traditional Feynman diagrammatics we make extensive use of modern on-shell methods to construct the amplitudes. In particular, we use $d$-dimensional generalized unitarity methods [135] to construct the complete loop-integrand in a physically motivated dimensional scheme. We begin with a brief overview of unitarity methods and then describe in detail the approach taken in this chapter. In Section 4.2.1, we introduce the techniques in the familiar context of Yang-Mills theory, then adapt the methods to Born-Infeld in Section 4.2.2.

### 4.2.1 Generalized Unitarity and Supersymmetric Decomposition

The main idea of unitarity based methods [136] is to exploit that the loop integrand is a complex rational function of the loop momentum with singularity structure constrained by factorization into on-shell tree amplitudes. Here we focus specifically on one-loop order and all calculations are made in a given dimensional regularization scheme. This means that while the external momenta and polarizations are strictly $d = 4$-dimensional, the loop momentum is formally regarded as

$d = (4 - 2\epsilon)$-dimensional.

### 4-Dimensional Unitarity Methods

Expanding the loop-*integrand* around $\epsilon = 0$, the leading $\mathcal{O}(\epsilon^0)$ component has an unambiguous physical meaning related to unitarity of the S-matrix. Via the Cutkosky theorem [137], the factorization of the integrand into on-shell tree amplitudes on 4d cuts

$$l_1^2 \cdots l_k^2 \, \mathcal{I}_n[l] \Big|_{l_1^2 = \cdots = l_k^2 = 0} = \sum_{\text{states}} \mathcal{A}_{(1)}^{\text{tree}_4} \ldots \mathcal{A}_{(k)}^{\text{tree}_4}, \quad k \leq 4, \tag{4.2.1}$$

where $l_i^\mu$, for $i = 1, \cdots k$ are 4d momenta, ensures that the integrated amplitude has the correct branch cut discontinuities required by the optical theorem. A rational function with all the correct 4d cuts (and no spurious cuts) then yields the correct amplitude at $\mathcal{O}(\epsilon^0)$ after integration, up to a function with no branch cuts, i.e. a rational function. This is the idea of the *4-dimensional unitarity* approach: the cut-constructible part of the amplitude is completely fixed by the physical tree amplitudes. Due to a complete understanding of integrand reduction to a basis of master scalar integrals at one-loop this procedure can be completely automated [138]. The remaining rational function ambiguity must then be determined by imposing additional physical constraints, such as cancellation of spurious singularities in the cut-constructible part or by imposing known behavior in soft or collinear limits [133, 139]. One advantage of calculating the 4d-cut-constructible part and the rational part separately in this way is that at all stages of the calculation we make use of regularization scheme-independent, physical objects (on-shell 4d tree-amplitudes). The primary disadvantage to this approach is the relative difficulty in calculating the rational terms separately.

### $d$-Dimensional Unitarity Methods

In certain cases, the cut-constructible part vanishes and the integrated loop-amplitude is purely rational. In that case, the method outlined above for determining the rational part is not applicable. This, in particular, will be the situation for the amplitudes (4.1.5) of interest in this chapter.

A more familiar example is the SD and NSD sectors of pure Yang-Mills theory (i.e. the all-plus and all-plus-one-minus gluon amplitudes): at one-loop, any 4d cut has factors of tree amplitudes of the SD and NSD helicity configurations and those vanish [90], hence all the 4d cuts vanish. According to the discussion above, the absence of 4d cuts implies that the resulting integrand is zero at $\mathcal{O}(\epsilon^0)$ (vanishes in $d = 4$), but may have non-zero contributions at $\mathcal{O}(\epsilon)$. As a result, SD and NSD one-loop amplitudes have no branch cut discontinuities and are instead purely rational functions. These rational contributions arise from subtle $\epsilon/\epsilon$ cancellations after integration; the same mechanism gives rise to the chiral anomaly in dimensional regularization [3]. Since the SD and NSD sectors of YM and BI theory are very similar, we introduce the method here for YM , then adapt it to BI theory in the Section 4.2.2.

The method of *d-dimensional unitarity* [135] does not separate the 4d-cuts and rational terms. In the d-dimensional unitarity approach, we must first define a suitable dimensional regularization scheme in which $d$-dimensional integrand cuts have the form

$$l_1^2 l_2^2 \dots l_k^2 \mathcal{I}_n[l]\Big|_{l_1^2=\dots=l_k^2=0} = \sum_{\text{states}} \mathcal{A}_{(1)}^{\text{tree}_d} \dots \mathcal{A}_{(k)}^{\text{tree}_d} \,, \tag{4.2.2}$$

where the on-shell cut momenta $l_i$ are $d$-dimensional. The additional constraint of correct cuts in $d$-dimensions is sufficient to construct the integrand to all orders in $\epsilon$, allowing us to determine both the 4d cut-constructible and rational parts at the same time. This approach is therefore well-suited to the purely rational SD and NSD one-loop amplitudes of Yang-Mills. The difficulty of this approach is that we are forced to work with regularization scheme-dependent quantities, which are therefore non-unique, and furthermore since the cuts are in $d$-dimensions, we lose the simplicity of spinor-helicity variables.

In certain special cases, such as pure Yang-Mills and pure Born-Infeld in $d = 4$, we can maneuver around these difficulties and define a regularization scheme in which both the $d$-dimensional-cut structure is quite simple *and* we can still make use of spinor-helicity variables. This simplified implementation of $d$-dimensional unitarity is sometimes referred to as *supersymmetric decomposition* and this is what we describe next.

#### Consistent Truncation and Supersymmetric Decomposition

It is instructive to first review the concept of supersymmetric consistent truncation at tree-level. In general we say that model A is a *consistent truncation* of model B if the on-shell states of A form a subset of the on-shell states of B and (when restricted to the A-states) the S-matrices are identical at tree-level.[3] This occurs in any model in which the states of B/A (B-states that are not A-states) carry an independent charge or parity; such states can give no contribution to state-sums on factorization singularities and hence no contribution to the tree-level S-matrix elements with all external A-states. A simple example of this occurs in any model containing both Bosonic and Fermionic states; since the quantity $(-1)^F$ is conserved we can always construct a consistent truncation by restricting to the Bosonic sector. If there are additional conserved quantities in the Bosonic sector, then it may be possible to give a further truncation.

As a relevant example, consider $\mathcal{N} = 2$ super Yang-Mills (without matter hypermultiplets) in $d = 4$. The spectrum consists of a massless vector multiplet containing a gauge boson $g^{\pm}$, two

[3] This is equivalent to the statement that solutions to the classical equations of motion for model A are also solutions to the equations of motion of model B with the fields in B/A turned off.

Weyl fermions $\psi^{\pm}_{1,2}$ and a complex scalar $\phi$, $\overline{\phi}$. Restricting to the Bosonic sector gives a consistent truncation, the resulting model is non-supersymmetric and describes Yang-Mills coupled to a massless (adjoint) complex scalar. In this model there is an additional *global* symmetry, descended from R-symmetry, under which the states are charged as

$$Q[g^{\pm}] = 0, \qquad Q[\phi] = 1, \qquad Q[\overline{\phi}] = -1. \tag{4.2.3}$$

Consequently, we can define a further truncation to the purely *gluonic sector*, the resulting model is precisely pure non-supersymmetric Yang-Mills. The statement of consistent truncation in this example is then

$$\mathcal{A}_n^{\text{(tree) } \mathcal{N}=2 \text{ SYM}} [1_g, \dots n_g] = \mathcal{A}_n^{\text{(tree) YM+Adj}} [1_g, \dots n_g] = \mathcal{A}_n^{\text{(tree) YM}} [1_g, \dots n_g] \, . \tag{4.2.4}$$

Since gluonic amplitudes in $\mathcal{N} = 2$ SYM in the SD and NSD helicity sectors vanish at all orders of perturbation theory, these same helicity sectors must likewise vanish in tree-level non-supersymmetric Yang-Mills.

The notion of consistent truncation in the form of equalities such as (4.2.4) does not continue to hold at loop-level. We can, however, make use of supersymmetric truncations at one-loop to form a *supersymmetric decomposition*. Let us illustrate this in the context of Yang-Mills. At one-loop, all states in the model generically run in every loop, for $\mathcal{N} = 0, 1$ and 2 SYM we can schematically represent the contributions to purely gluonic amplitudes as

$$\begin{aligned}
\mathcal{A}_n^{\text{(1-loop) YM}} [1_g \dots n_g] &= \mathcal{A}_n^{[V]} [1_g \dots n_g] \\
\mathcal{A}_n^{\text{(1-loop) } \mathcal{N}=1 \text{ SYM}} [1_g \dots n_g] &= \mathcal{A}_n^{[V]} [1_g \dots n_g] + \mathcal{A}_n^{[F]} [1_g \dots , n_g] \\
\mathcal{A}_n^{\text{(1-loop) } \mathcal{N}=2 \text{ SYM}} [1_g \dots n_g] &= \mathcal{A}_n^{[V]} [1_g \dots n_g] + 2\mathcal{A}_n^{[F]} [1_g \dots n_g] + \mathcal{A}_n^{[S]} [1_g \dots n_g] \, ,
\end{aligned} \tag{4.2.5}$$

where $V$, $F$, and $S$ represent contributions from vector bosons, Weyl fermions, and complex scalars, respectively. The contributions on the right-hand-side have no invariant physical meaning, even in the context of a Feynman diagram expansion, as a grouping of terms they depend on the choice of regularization scheme. One can, however, give invariant physical meaning to these expressions on 4d-unitarity cuts: the decomposition reflects the contributions to the state sums. Note that it is the existence of the same conservation laws that allowed us to construct consistent truncations at tree-level that make this decomposition sensible. In particular, due to (4.2.3), there are no mixed scalar/gluon contributions to 4d cuts. If the amplitudes are calculated in the Four Dimensional Helicity (FDH) or similar schemes, in which the one-to-one correspondence between the (external) 4-dimensional helicity states and the (internal) $d$-dimensional states is preserved [140] then the relations (4.2.5) are well-defined on $d$-dimensional cuts.

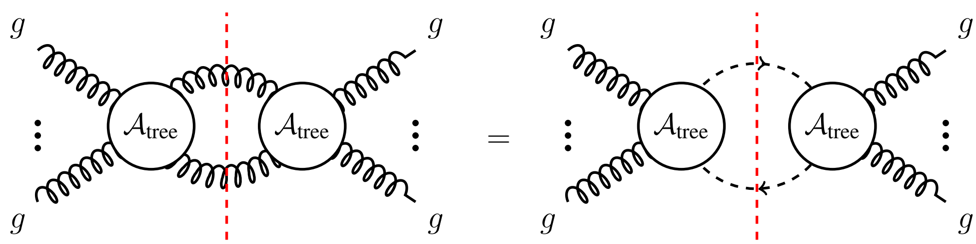
The notion of a *supersymmetric decomposition* is a rearrangement of (4.2.5) such that one-loop amplitudes in non-supersymmetric Yang-Mills can be given as sums over contributions from $\mathcal{N} = 1, 2$ vector multiplets and adjoint scalars

$$\begin{aligned}&\mathcal{A}_n^{\text{(1-loop) YM}}\left[1_g \ldots n_g\right] \\ &\quad = -\mathcal{A}_n^{\text{(1-loop) }\mathcal{N}=2\text{ SYM}}\left[1_g \ldots n_g\right] + 2\mathcal{A}_n^{\text{(1-loop) }\mathcal{N}=1\text{ SYM}}\left[1_g \ldots n_g\right] + \mathcal{A}_n^{[S]}\left[1_g \ldots n_g\right].\end{aligned} \tag{4.2.6}$$

Next, we assume that our regularization scheme is supersymmetric (for example FDH [141]), and therefore the one-loop amplitudes satisfy the same supersymmetry Ward identities as the tree-level amplitudes.[4] This dramatically simplifies in the SD and NSD sectors, since the contributions from the $\mathcal{N} > 0$ components vanish. In these sectors the supersymmetric decomposition simplifies to

$$\mathcal{A}_n^{\text{(1-loop) YM}}\left[1_g^+ \ldots (n-1)_g^+,\, n_g^{\pm}\right] = \mathcal{A}_n^{[S]}\left[1_g^+ \ldots (n-1)_g^+,\, n_g^{\pm}\right]. \tag{4.2.7}$$

We refer to this as the *scalar-loop representation* of the one-loop amplitude. Again, in the context of $d$-dimensional unitarity we can interpret this statement unambiguously as a statement about the $d$-dimensional unitarity cuts of the loop-integrand.




As a consequence, the complete one-loop integrand can be reconstructed by requiring the correct $d$-dimensional unitarity cuts into $d$-dimensional tree-amplitudes of the form

$$\mathcal{A}_n^{\text{(tree)}}\left[1_\phi, 2_g \ldots (n-1)_g, n_{\overline{\phi}}\right]. \tag{4.2.8}$$

Here only the momenta of the scalars are $d$-dimensional, while the momenta and polarizations of the gluons are 4-dimensional.

We rewrite the $d$-dimensional momenta in terms of 4-dimensional momenta as

$$l^\mu = l^\mu_{[4]} + l^\mu_{[-2\epsilon]}. \tag{4.2.9}$$

Due to the orthogonality of 4-dimensional and $(-2\epsilon)$-dimensional subspaces, we can rewrite the

[4]In a non-supersymmetric scheme such as conventional dimensional regularization (CDR) the result of the loop integrals will typically not satisfy the supersymmetry Ward identities. Supersymmetry must be restored by adding finite local counterterms which modify the rational part of the one-loop amplitudes.

various Lorentz singlets that appear in the amplitude as

$$q \cdot l = q \cdot l_{[4]}, \qquad l^2 = l_{[4]}^2 + l_{[-2\epsilon]}^2 \equiv l_{[4]}^2 + \mu^2, \tag{4.2.10}$$

where $q^\mu$ is any 4-dimensional vector and $\mu^2 \equiv l_{[-2\epsilon]}^2$. Using these relations we find that we can rewrite all $d$-dimensional amplitudes (4.2.8) as 4-dimensional amplitudes with a massive scalar of mass $\mu^2$.

Up to this point we have not explicitly defined the regularization scheme, we have only made use of some assumed general properties. It is important to emphasize that physical observables are independent of the choice of regularization scheme. In this chapter, the calculation we describe is made in a particular version of dimensional regularization that has certain convenient properties, but the physical conclusions should be independent of this choice, we discuss this further in the Discussion section.

We shall define the massive scalar amplitudes directly in 4d, requiring all of the standard tree-level properties of Lorentz invariance, locality and unitarity, in addition to the requirement

$$\mathcal{A}_n^{\text{tree}}\left[1_\phi, 2_g, \ldots, (n-1)_g, n_{\overline{\phi}}\right] \xrightarrow{\mu^2 \to 0} \mathcal{A}_n^{\text{tree }(\mathcal{N}=2)}\left[1_\phi, 2_g, \ldots, (n-1)_g, n_{\overline{\phi}}\right]. \tag{4.2.11}$$

Even though the 4d cuts vanish in the SD and NSD amplitudes of consideration, the relations (4.2.5) make sense for all helicity amplitudes, and for those with non-vanishing 4d cuts the $\mathcal{A}_n^{[S]}$ cuts must be equal to products of tree-amplitudes of $\mathcal{N} = 2$ SYM. The problem of constructing the integrand in the scalar loop representation then has two parts:

1. Define a model of a massive adjoint scalar coupled to Yang-Mills which reduces to the Bosonic sector of $\mathcal{N} = 2$ SYM in the massless limit.

2. Construct a complex rational function of 4d momenta with correct cuts into the massive scalar tree amplitudes and no spurious cuts.

The required massless limit (4.2.11) is not sufficient to determine the massive scalar model described in Step 1. In addition to the *minimal* coupling,[5] we could also add generic terms to the scalar potential or higher-derivative couplings, for example we might consider a model described by the action

$$S[A_\mu, \phi, \overline{\phi}] = S_{\text{minimal}}[A_\mu, \phi, \overline{\phi}] + \int \mathrm{d}^4 x \left[\frac{\mu^2}{\Lambda_1^4}|\phi|^6 + \frac{\mu^2}{\Lambda_2^4}|\phi|^2 \mathrm{Tr}[F^2]\right], \tag{4.2.12}$$

[5]This includes the $|\phi|^4$ term in the scalar potential required to satisfy the requirement of $\mathcal{N} = 2$ supersymmetry in the massless limit.

where $\Lambda_1$ and $\Lambda_2$ are independent mass scales. Such a model clearly satisfies the correct massless limit. The presence of independent dimensionful parameters however makes this physically unacceptable, these would appear in the integrand we construct according to Step 2, and consequently the integrated amplitude. To ensure the absence of such spurious parameters we impose:

3. The result we calculate should agree with the parametric dependence on couplings expected from a full Feynman diagram calculation, therefore an acceptable massive scalar extension of Yang-Mills theory should depend only on the dimensionless Yang-Mills coupling $g_{\text{YM}}$.

By this simple argument all such higher dimension couplings must be absent, and we find that the conditions (1-3) uniquely pick out the minimally coupled massive adjoint scalar with the supersymmetric scalar potential. Such tree amplitudes can be generated efficiently by using massive BCFW recursion, which is reviewed in Section 4.3.3.1.

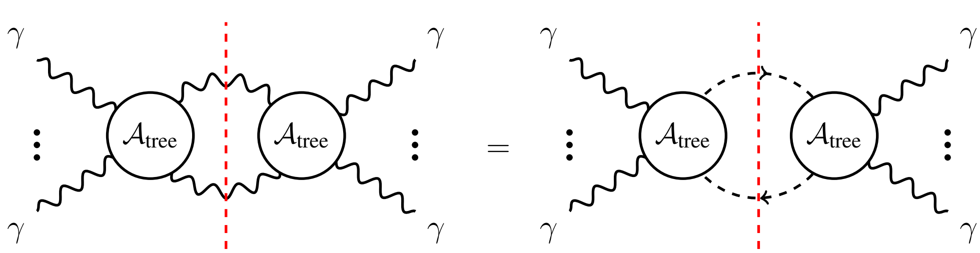
The strategy described above has been used successfully to calculate all-multiplicity one-loop amplitudes in the SD and NSD sectors of pure Yang-Mills [13]. It has also been implemented in pure Einstein gravity [133] and also recently Einstein Yang-Mills [142]. The purpose of this chapter is to implement this approach in non-supersymmetric Born-Infeld electrodynamics in $d = 4$. In the following subsection we will describe the novelties that appear in this model compared to Yang-Mills.

### 4.2.2 Massive Scalar Extension of Born-Infeld

Almost everything we described in Section 4.2.1 for pure Yang-Mills in $d = 4$ applies to pure Born-Infeld in $d = 4$. At tree-level, non-supersymmetric Born-Infeld is a consistent truncation of $\mathcal{N} = 2$ super Born-Infeld. Consequently, the SD and NSD amplitudes vanish at tree-level. Moreover, in a supersymmetric regularization scheme, the SD and NSD one-loop amplitudes have a scalar-loop representation

$$\mathcal{A}_n^{(\text{1-loop})\ \text{BI}_4}\left(1_\gamma^+,\ldots,(n-1)_\gamma^+,n_\gamma^\pm\right) = \mathcal{A}_n^{[S]}\left(1_\gamma^+,\ldots,(n-1)_\gamma^+,n_\gamma^\pm\right). \tag{4.2.13}$$

These one-loop amplitudes have no $d = 4$ cuts, so are purely rational. We compute the integrand using $d$-dimensional unitarity in which the cuts factor into tree amplitudes with two massive scalars coupled to the Born-Infeld photons.



Here the massive scalar model should reduce to $\mathcal{N} = 2$ super Born-Infeld in the massless limit, analogously to (4.2.11). Since there are independent gauge-invariant local operators coupling the Born-Infeld photon and a massive scalar which vanish in the massless limit, this is not sufficient to determine the massive model. Unlike Yang-Mills, we can construct an infinite number of such operators *without* introducing spurious dimensionful parameters. In other words, the analogue of conditions (1)-(3) above are *not sufficient* to uniquely pick out a specific model.

To proceed, additional physical constraints must be applied to uniquely define the massive scalar extension of Born-Infeld. In the remainder of this section, we describe the model, which we call mDBI$_4$ (massive DBI in 4d), and argue from two points of view why it is an appropriate definition. In Section 4.3 we then calculate the mDBI$_4$ tree amplitudes

$$\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-2)_\gamma^+, (n-1)_\gamma^\pm, n_{\bar{\phi}}\right), \tag{4.2.14}$$

needed for the unitarity cuts, where the complex scalar has mass $\mu^2 \equiv l_{[-2\epsilon]}^2$ in $d = 4$. As stated, these tree amplitudes must satisfy

$$\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma, \ldots, (n-1)_\gamma, n_{\bar{\phi}}\right) \xrightarrow{\mu^2 \to 0} \mathcal{A}_n^{\mathcal{N}=2\ \mathrm{BI}_4}\left(1_\phi, 2_\gamma, \ldots, (n-1)_\gamma, n_{\bar{\phi}}\right). \tag{4.2.15}$$

The two approaches to define mDBI$_4$ are dimensional reduction and the double-copy; we now describe each in turn.

**Dimensional Reduction and Supersymmetry**

We define mDBI$_4$ as the dimensional reduction of pure Born-Infeld from $d = 6$ (BI$_6$). Specifically we take 6d tree-amplitudes with momenta and polarizations in the configuration described in Table 4.1, i.e. the photon momenta and polarizations lie in a 4d subspace for lines $2, 3, \ldots, n-1$ while lines $1$ and $n$ have genuinely 6d momenta but polarizations orthogonal to the 4d subspace, so in the 4d setting they are scalars. This is an appropriate definition because the amplitudes (4.2.14) arise from $d$-dimensional cuts of a loop-integrand in a supersymmetric regularization scheme.

As in the previous subsection, it is instructive to first describe the case of pure Yang-Mills. In any scheme, on 4d cuts the integrand factors into tree-amplitudes of YM$_4$, which by virtue of being a consistent truncation of $\mathcal{N} = 2$ SYM$_4$ satisfy the supersymmetry Ward identities for 8 supercharges. On $d$-dimensional cuts, however, we would generically expect the action of the supersymmetry algebra to be explicitly broken. To construct a supersymmetric regularization scheme, we want to define a dimensional continuation from $d = 4$ in which the action of the 8 supercharges of $\mathcal{N} = 2$ is unbroken.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\vec{p}_{1,n}$ | x | x | x | x | x |
| $\vec{\epsilon}_{1,n}$ | | | | x | x |
| $\vec{p}_{2,3,\ldots,n-1}$ | x | x | x | | |
| $\vec{\epsilon}_{2,3,\ldots,n-1}$ | x | x | x | | |

Table 4.1: Kinematic configuration of momenta and polarizations of $\text{BI}_6$ defining $\text{mDBI}_4$ and for $\text{YM}_6$ defining $(\text{YM} + \text{mAdj})_4$.

A natural way to do this is to recognize that the Yang-Mills-scalar tree amplitudes (4.2.8) can be obtained from pure Yang-Mills in $d = 6$ ($\text{YM}_6$) with momenta and polarizations in the configuration given in Table 4.1. Since $\text{YM}_6$ is a consistent truncation of $\mathcal{N} = (1, 0)$ $\text{SYM}_6$, the $\text{YM}_6$ tree amplitudes must satisfy the full set of $\mathcal{N} = (1, 0)$ supersymmetry Ward identities. It therefore follows that in the configuration given in Table 4.1, the 6d amplitudes written in a 4d language, must satisfy (some version of) the supersymmetry Ward identities for 8 supercharges. We should therefore expect a regularization scheme with a scalar-loop representation (4.2.7), with massive scalar amplitudes defined by this dimensional reduction from 6d, to preserve (some version of) the full $\mathcal{N} = 2$ supersymmetry on *d-dimensional cuts*, and it is therefore a supersymmetric scheme. This definition of the Yang-Mills-scalar amplitudes satisfies the criteria we gave in the previous subsection of absence of spurious parametric dependence. The massive scalar extension of 4d Yang-Mills theory defined this way will be denoted $(\text{YM} + \text{mAdj})_4$; as it turns out, it will be useful in our amplitude constructions.

The same argument applies essentially verbatim to Born-Infeld. $\text{BI}_6$ is a consistent truncation of $\mathcal{N} = (1, 0)$ super Born-Infeld ($\text{SBI}_6$), so the tree-amplitudes of $\text{mDBI}_4$ defined by the configuration given in Table 4.1 must preserve the action of 8 supercharges. Hence the SD and NSD one-loop integrands of $\text{BI}_4$ in the scalar loop representation (4.2.1) preserve the action of $\mathcal{N} = 2$ supersymmetry on $d$-dimensional cuts, and therefore define a scheme that we expect to be supersymmetric. We do not have a formal proof of this statement.

**BCJ Double-Copy**

A complimentary argument, with the same conclusion, is given by considering the BCJ double copy. It was shown in [121], in the context of the CHY formalism [143, 144], that the field theory KLT formulae which give gravity tree amplitudes as the double-copy of gauge theory tree amplitudes also give Born-Infeld if one of the gauge theory factors is replaced by Chiral Perturbation Theory ($\chi$PT). $\chi$PT is a non-linear sigma model with target space $\frac{SU(N)\times SU(N)}{SU(N)}$.

This double-copy statement applies at tree-level in $d$-dimensions

$$\mathrm{BI}_d = \mathrm{YM}_d \otimes_{\mathrm{KLT}} \chi\mathrm{PT}_d \,. \tag{4.2.16}$$

It has been conjectured by BCJ that the double-copy could be extended to loop integrands [145]. This remains a conjecture, though it has been successfully applied in many examples and represents the current state of the art for high loop order calculations in maximal supergravity [146]. In this spirit we conjecture that the tree-level double-copy construction of Born-Infeld extends to a complete loop-level double copy following BCJ.

In this chapter we do not make use of explicit color-kinematics dual BCJ integrands. Rather, we proceed by assuming that such a representation of the $\mathrm{BI}_4$ integrand exists in a supersymmetric regularization scheme which admits a scalar-loop representation (4.2.13). Then on $d$-dimensional cuts, the integrand factors into tree amplitudes in a model coupling Born-Infeld photons to a massive scalar. Furthermore, these tree amplitudes should be given by the tree-level double-copy of $\mathrm{YM}_4$ coupled to a massive scalar and $\chi\mathrm{PT}_4$ coupled to a massive scalar. The existence of such double-copy compatible massive scalar models is quite non-trivial.

We now want to show that the proposed definition of $\mathrm{mDBI}_4$ is indeed generated by the tree-level double copy. The key to this is that the KLT product is valid in $d$-dimensions, it therefore commutes with dimensional reduction[6] in the sense described by the configuration in Table 4.1:

$$\begin{tikzcd}
 & \mathrm{YM}_6 \arrow[rr, "\text{Dimensional Reduction}"] \arrow[dl] & & (\mathrm{YM{+}mAdj})_4 \arrow[dr] & \\
\bigotimes_{\mathrm{KLT}} \arrow[r] & \mathrm{BI}_6 \arrow[rr, "\text{Dimensional Reduction}"] & & \mathrm{mDBI}_4 & \bigotimes_{\mathrm{KLT}} \arrow[l] \\
 & \chi\mathrm{PT}_6 \arrow[rr, "\text{Dimensional Reduction}"] \arrow[ul] & & \mathrm{m}\chi\mathrm{PT}_4 \arrow[ur] &
\end{tikzcd}$$

Since both Yang-Mills and $\chi$PT satisfy the conditions necessary for the double-copy to be well-defined in $d$-dimensions, we can begin with these models in $d = 6$. As illustrated in the diagram above we have two choices, either take the 6d double-copy first and then dimensionally reduce to 4d, or dimensionally reduce to the 4d massive scalar models first and then take the 4d double-copy; it is clear these choices will agree. In the first case, the validity of the $d$-dimensional double copy gives precisely the definition of $\mathrm{mDBI}_4$ given above, the second case gives us exactly the massive scalar double-copy we expect on $d$-dimensional cuts if the loop BCJ conjecture is correct.

[6] The dimensional reduction of $\chi\mathrm{PT}_6$ to $d = 4$ is defined by the momentum configuration in Table 4.1.

The advantage of working in the 4d formulation is that we can take advantage of the 4d spinor helicity formalism.

## 4.3 Calculating mDBI$_4$ Tree Amplitudes

### 4.3.1 General Structure

As described in the previous section, the input required for constructing the (N)SD loop integrands using $d$-dimensional unitarity are tree amplitudes in some model (which we call mDBI$_4$) describing a massless Born-Infeld photon coupled to a massive complex scalar. We need two types of tree amplitudes:

- mDBI$_4$ NSD amplitudes: These are of the form $\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^+, n_{\overline{\phi}}\right)$ and will be used to calculate BI$_4$ SD and NSD amplitudes in Sections 4.4.2 and 4.4.3 respectively.
- mDBI$_4$ MHV amplitudes: These are of the form $\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^-, n_{\overline{\phi}}\right)$ and will be used to calculate BI$_4$ NSD amplitudes in Section 4.4.3.

First we will give a general parametrization of such tree amplitudes, then in the following section we will fix all ambiguities using two complimentary approaches.

The analytic structure of the mDBI$_4$ amplitudes have the general form of a rational function of external kinematic data and can be split into contributions

$$\begin{aligned}
&\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^\pm, n_{\overline{\phi}}\right) \\
&= \mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^\pm, n_{\overline{\phi}}\right)\Big|_{\text{factoring}} + \mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^\pm, n_{\overline{\phi}}\right)\Big|_{\text{contact}}.
\end{aligned} \tag{4.3.1}$$

The *factoring* terms contain all kinematic singularities, which are required to be simple poles on invariant masses of subsets of external momenta, and have residues given by sums of products of lower point amplitudes. In this sense the factoring terms are recursively determined by amplitudes at lower multiplicity. In EFTs (such as mDBI$_4$) the resulting rational function is incompletely determined by factorization, and there is some remaining polynomial ambiguity. These ambiguities are contained in the *contact* contribution, which encodes all independent local operators compatible with the assumed properties of the model. We can give a general parametrization of these contact contributions for mDBI$_4$ through a combination of dimensional analysis, little group scaling and analysis of the massless limit.

In $d = 4$ the amplitudes have mass dimension $[\mathcal{A}_n] = 4 - n$, this includes both dimensionful coupling constants and kinematic dependence. The contact contribution is then a sum over terms of the schematic form

$$\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^\pm, n_{\overline{\phi}}\right)\Big|_{\text{contact}} \sim \frac{1}{\Lambda^m} F_n^\pm\left(\{|i\rangle, |i]\}, p_{[4]}^{1,n}, \mu^2\right), \tag{4.3.2}$$

where $[\Lambda] = 1$ is the dimensionful scale appearing in the Born-Infeld action (4.1.1) and $[F_n] - m = 4 - n$. Since this is a contact contribution $F_n$ must be a polynomial in the Lorentz invariant spinor contractions and the mass of the scalar $\mu^2$. These polynomials must have the correct little group scaling dictated by their helicity configurations. This sets a lower-bound on the mass dimension of $F_n^\pm$ since we must have

$$\begin{aligned}
F_n^+\left(\{|i\rangle, |i]\}, p_{[4]}^{1,n}, \mu^2\right) &\sim |2]^2|3]^2 \ldots |n-1]^2 G_n^+\left(\{|i\rangle, |i]\}, p_{[4]}^{1,n}, \mu^2\right) \\
F_n^-\left(\{|i\rangle, |i]\}, p_{[4]}^{1,n}, \mu^2\right) &\sim |2]^2|3]^2 \ldots |n-1\rangle^2 G_n^-\left(\{|i\rangle, |i]\}, p_{[4]}^{1,n}, \mu^2\right).
\end{aligned} \tag{4.3.3}$$

Here $G^\pm$ are again polynomials in helicity spinors, but with zero little group weight. Since $[G^\pm] \geq 0$ we must have $[F_n^\pm] \geq n - 2$.

Next we impose that the complete mDBI$_4$ amplitudes should agree with $\mathcal{N} = 2$ BI$_4$ in the limit $\mu^2 \to 0$. This constraint is quite powerful due to the conservation of a $U(1)_R$ duality charge in $\mathcal{N} = 2$ BI$_4$. Up to an arbitrary normalization, the states of the $\mathcal{N} = 2$ massless vector multiplet can be assigned the following additive quantum numbers

$$Q[\gamma^\pm] = \pm 1, \quad Q[\psi_{1,2}^\pm] = \pm 1/2, \quad Q[\phi] = Q[\overline{\phi}] = 0. \tag{4.3.4}$$

It is straightforward to show that these charges are conserved at tree-level since they are conserved by the leading $n = 4$ interactions and the entire tree-level S-matrix is constructible by on-shell subtracted recursion [53]. Note that this $U(1)_R$ is *not* a subgroup of the $SU(2)_R$ symmetry group under which the fermions $\psi_A$ transform as a doublet. It is an independent symmetry which enhances the full R-symmetry group of $\mathcal{N} = 2$ BI$_4$ to $U(2)_R$. The analogous enhancement of R-symmetry in maximally supersymmetric Born-Infeld was first discussed in [147]. As a consequence of the conservation of the duality charges (4.3.4), in the NSD and MHV sectors of mDBI$_4$ the massless limits are given by

$$\begin{aligned}
&\mathcal{A}_4^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_{\overline{\phi}}\right) \xrightarrow{\mu^2 \to 0} 0, \\
&\mathcal{A}_4^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^-, 4_{\overline{\phi}}\right) \xrightarrow{\mu^2 \to 0} -\langle 3|p_1|2]^2, \\
&\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^\pm, n_{\overline{\phi}}\right) \xrightarrow{\mu^2 \to 0} 0, \quad n > 4.
\end{aligned} \tag{4.3.5}$$

Due to the different singularity structure, the factoring and contact terms cannot interfere in this limit, and so the contact terms must vanish independently. For this to happen the contact terms must be proportional to some positive power of $\mu^2$, which further increases the minimal dimension to $[F_n^{\pm}] \geq n$. The contact terms must then have the schematic form

$$\begin{aligned}
\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^+, n_{\overline{\phi}}\right)\Big|_{\text{contact}} &\sim \frac{\mu^2}{\Lambda^{2n-4}}|2]^2|3]^2 \ldots |n-1]^2 + \mathcal{O}\left(\frac{1}{\Lambda^{2n-3}}\right) \\
\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^-, n_{\overline{\phi}}\right)\Big|_{\text{contact}} &\sim \frac{\mu^2}{\Lambda^{2n-4}}|2]^2|3]^2 \ldots |n-1\rangle^2 + \mathcal{O}\left(\frac{1}{\Lambda^{2n-3}}\right).
\end{aligned} \tag{4.3.6}$$

It is easy to see that in the $(n-1)^-$ (MHV) case no contact term of this leading mass dimension can exist since there is no non-vanishing way to contract the angle spinors.

Next we recall our discussion from Section 4.2, such contact contributions should not introduce any spurious dimensionful parameters which might appear in the final integrated amplitude. We should not consider contributions with more inverse powers of $\Lambda$ at a fixed multiplicity $n$. In Appendix D we give a short proof that at each multiplicity $n$ there is a unique contact term, the final result can be parametrized as

$$\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^+, n_{\overline{\phi}}\right)\Big|_{\text{contact}} = \frac{c_n\mu^2}{\Lambda^{2n-4}}\left([23]^2[45]^2 \ldots [n-2, n-1]^2 + \ldots\right), \tag{4.3.7}$$

where $+\ldots$ denotes the sum over all ways of partitioning the set $\{2, \ldots, n-1\}$ into subsets of length 2. Such local matrix elements can be generated from local operators of the form

$$\mathcal{L}_{\text{mDBI}_4} \supset \frac{c_{2n}\mu^2}{\Lambda^{4n-4}}|\phi|^2\left(F_{\alpha\beta}^+ F^{+\alpha\beta}\right)^{n-1} \quad + \quad \text{h.c.} \tag{4.3.8}$$

In subsequent sections the $\Lambda$ dependence of the scattering amplitudes will be suppressed, they can trivially be restored by dimensional analysis.

The remarkable result (which we will verify using two complimentary approaches in the following sections, the first presented below and the second described in Appendix 4.3.3) is that if we define mDBI$_4$ as the dimensional reduction of BI$_6$ as described above, then $c_n = 0$ for $n > 4$. The complete tree amplitudes are then completely fixed by recursive factorization into the fundamental 4-point mDBI$_4$ amplitudes.

### 4.3.2 T-Duality and Low-Energy Theorems

One of the most important and remarkable properties of D-branes (of which Born-Infeld and related models provide the low-energy effective description) is their behaviour under T-duality [148]. Though this is a non-perturbative stringy property, a useful remnant remains even in the

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\vec{p}_{1,n}$ | x | x | | x | x |
| $\vec{\epsilon}_{1,n}$ | | | | x | x |
| $\vec{p}_{2,3,\ldots,n-2}$ | x | x | | | |
| $\vec{\epsilon}_{2,3,\ldots,n-2}$ | x | x | | | |
| $\vec{p}_{n-1}$ | x | x | | | |
| $\vec{\epsilon}_{n-1}$ | | | x | | |

Table 4.2: Kinematic configuration of momenta and polarizations of $\mathrm{BI}_6$ defining the 3d dimensional reduction of $\mathrm{mDBI}_4$. The 3-direction will be T-dualized, mapping the polarization of the photon labeled $n-1$ to a brane modulus.

tree-level scattering amplitudes of pure Born-Infeld. We will consider the configuration of momenta and polarizations described in Table 4.2.

At tree-level all internal momenta are linear combinations of external momenta, and so in this configuration the amplitudes are independent of the 3-direction in momentum space. This means that the tree-amplitudes are invariant under compactification of the spatial 3-direction on $S^1$. T-duality in this context is the statement that a space-filling D5-brane on $\mathbb{R}^{4+1} \times S^1$ with the radius of $S^1$ given by $R$, is equivalent to a codimension-1 D4-brane on $\mathbb{R}^{4+1} \times S^1$, where $S^1$ is the transverse dimension with radius $\sim 1/R$. In the full string theory, T-duality relates infinite towers of KK and winding modes. In this low-energy EFT containing only the massless states as on-shell degrees of freedom, the only non-trivial mapping is between photons polarized in the compact direction on the D5-brane and the brane modulus of the D4-brane

$$|\gamma^{\top}(\vec{p})\rangle \leftrightarrow |\Phi(\vec{p})\rangle. \tag{4.3.9}$$

Since the tree-level amplitudes in Table 4.2 are independent of the compactification, they must remain invariant in the limit $R \to 0$. In the T-dual configuration this corresponds to the decompactification limit in which we have a D4 brane embedded in $\mathbb{R}^{5+1}$. In this limit, the spontaneous symmetry breaking pattern in the T-dual frame jumps discontinuously

$$\frac{\mathrm{ISO}(4,1) \times \mathrm{SO}(2)}{\mathrm{ISO}(4,1)} \xrightarrow{R\to 0} \frac{\mathrm{ISO}(5,1)}{\mathrm{ISO}(4,1)}. \tag{4.3.10}$$

The brane modulus is then identified as the Goldstone mode of both the translation symmetry in the 3-direction *and* the Lorentz transformations mixing the 3- and world-volume directions. In the physical scattering amplitudes this manifests as *enhanced* soft theorems for the brane modulus

$$\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-2)_\gamma^+, (n-1)_\Phi, n_{\bar{\phi}}\right) \sim \mathcal{O}\left(p_{n-1}^2\right), \quad \text{as} \quad p_{n-1} \to 0, \tag{4.3.11}$$

where the momenta and polarizations are as given in Table 4.2. In this section we will use this result to fix the contact term ambiguities of the mDBI$_4$ amplitudes. This momentum configuration is an effective further dimensional reduction from 4d to 3d and so we will write the explicit form of the amplitudes in 3d language. In our conventions, the dimensional reduction map takes an especially simple form

$$4d \to 3d: \qquad \langle ij\rangle \to \langle ij\rangle, \qquad [ij] \to \langle ij\rangle, \tag{4.3.12}$$

which we will then further simplify (for purely Bosonic amplitudes this means rewriting all helicity spinor contractions as Mandelstam invariants). To apply these results to the Ansatz form of the mDBI$_4$ amplitudes described above, which are in the helicity basis, we must relate the transverse polarization $\gamma^\top$ to a linear combination of helicity states. In our conventions the correct linear combination is found to be

$$|\gamma^\top(\vec{p})\rangle = |\gamma^+(\vec{p})\rangle - |\gamma^-(\vec{p})\rangle, \tag{4.3.13}$$

which for the helicity amplitudes means

$$\begin{aligned}&\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-2)_\gamma^+, (n-1)_\gamma^\top, n_{\overline{\phi}}\right) = \\ &\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-2)_\gamma^+, (n-1)_\gamma^+, n_{\overline{\phi}}\right) - \mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-2)_\gamma^+, (n-1)_\gamma^-, n_{\overline{\phi}}\right).\end{aligned} \tag{4.3.14}$$

The method used in this section will be to form this linear combination of Ansatze, apply the dimensional reduction map and then take the soft limit $p_{n-1} \to 0$. Compatibility with T-duality then requires that the $\mathcal{O}(p_{n-1})$ terms cancel amongst themselves, this requirement uniquely fixes the $c_n$ coefficients.

#### 4.3.2.1 Explicit Examples of T-duality Constraints

We will begin with the 4-point amplitudes in mDBI$_4$. As described above the MHV amplitude is uniquely fixed by the $\mu^2 \to 0$ limit, while the NSD amplitudes are fixed up to an overall coefficient

$$\mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_{\overline{\phi}}\right) = c_4\mu^2[23]^2. \tag{4.3.15}$$

By taking the appropriate linear combination according to (4.3.13) we can form an amplitude for which particle 3 is polarized in the direction *transverse* to a particular 2d subspace

$$\begin{aligned}&\mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^\top, 4_{\overline{\phi}}\right) \\ &\quad = \mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_{\overline{\phi}}\right) - \mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^-, 4_{\overline{\phi}}\right) \\ &\quad = c_4\mu^2[23]^2 + \langle 3|p_1|2]^2.\end{aligned} \tag{4.3.16}$$

We then apply the dimensional reduction map, after reduction to 3d the various spinor contractions reduce to

$$\begin{aligned}
&[23]^2 \to s_{23} \\
&\langle 3|p_1|2]^2 \to \mathrm{Tr}\left[p_3 \cdot p_1 \cdot p_2 \cdot p_1\right] = 2\left(2(p_1 \cdot p_3)(p_1 \cdot p_2) - p_1^2(p_2 \cdot p_3)\right). \qquad (4.3.17)
\end{aligned}$$

Applying this gives

$$\mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^\top, 4_{\overline{\phi}}\right) \xrightarrow{3d} 2(c_4+1)\mu^2(p_2 \cdot p_3) + 4(p_1 \cdot p_3)(p_4 \cdot p_3). \qquad (4.3.18)$$

In the limit where $p_3 \to 0$ we can see that the first term vanishes at $\mathcal{O}(p_3)$ while the second term vanishes at $\mathcal{O}(p_3^2)$, the T-duality constraint then forces us to choose $c_4 = -1$. This result can also be obtained in a completely different way by using a massive version of the KLT relations (4.3.50).

At 6-point and higher it is necessary to define the soft degree more precisely. Let's quickly review the rigorous definition of a soft limit (see [51] for more details). We evaluate our amplitude on a one-parameter family of momenta of the form

$$\hat{p}_5(\epsilon) = \epsilon p_5, \quad \hat{p}_i(\epsilon) = p_i + \epsilon q_i, \quad i \neq 5. \qquad (4.3.19)$$

The deformed momenta should satisfy momentum conservation and the on-shell conditions for all values of $\epsilon \in \mathbb{C}$, which requires

$$p_5^2 = 0, \quad p_i \cdot q_i = 0, \quad q_i^2 = 0, \quad \sum_{i\neq 5} p_i = 0, \quad p_5 + \sum_{i \neq 5} q_i = 0. \qquad (4.3.20)$$

At leading order in the $\epsilon$-expansion the $q_i$ momenta do not appear. After dimensional reduction our amplitudes are trivially at least $\mathcal{O}(\epsilon)$, our goal is then to show that these leading terms are actually zero and that therefore the leading term in the expansion is $\mathcal{O}(\epsilon^2)$. For this purpose, taking the soft limit is equivalent to taking $p_i,\ i \neq 5$ to satisfy 5-particle momentum conservation, and $p_5$ as an unrelated null vector. We should bare this in mind when making algebraic manipulations involving conservation of momentum.

Let's now proceed with the calculation of the 6-point soft limit. We begin with a general Ansatze which has the correct factorization properties and generally parametrized contact terms, and make a dimensional reduction

$$\begin{aligned}
&\mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_{\overline{\phi}}\right) \\
&\xrightarrow{3d+\mathrm{soft}} \frac{(\mu^2)^2 s_{23}s_{45}}{s_{123}+\mu^2} + \frac{(\mu^2)^2 s_{24}s_{35}}{s_{124}+\mu^2} + \frac{(\mu^2)^2 s_{25}s_{34}}{s_{12}+\mu^2} + c_6\mu^2\left(s_{23}s_{45} + s_{24}s_{35} + s_{25}s_{34}\right), \quad (4.3.21)
\end{aligned}$$

also,

$$\begin{aligned}
&\mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^-, 6_{\overline{\phi}}\right)\\
&\xrightarrow{3d+\text{soft}} \frac{\mu^2}{2}\left[\frac{s_{23}\left(2(p_5\cdot p_6)(s_{46}+\mu^2)+\mu^2 s_{45}\right)}{s_{123}+\mu^2}+\frac{s_{34}\left(2(p_5\cdot p_1)(s_{12}+\mu^2)+\mu^2 s_{25}\right)}{s_{12}+\mu^2}\right.\\
&\qquad\qquad \left.+\frac{s_{34}\left(4(p_5\cdot p_{34})(p_2\cdot p_{34})+2\mu^2(p_2\cdot p_5)\right)}{s_{126}}\right]+\mathcal{P}\left(2,3,4\right). \qquad (4.3.22)
\end{aligned}$$

Taking the difference we find that the $(\mu^2)^2$ terms cancel and the remaining terms are purely local

$$\begin{aligned}
&\mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^\top, 6_{\overline{\phi}}\right)\\
&= \mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_{\overline{\phi}}\right)-\mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^-, 6_{\overline{\phi}}\right)\\
&\xrightarrow{3d+\text{soft}} \frac{1}{2}c_6\mu^2 s_{23}s_{45}-\mu^2 s_{23}(p_5\cdot p_6)-\mu^2 s_{34}(p_1\cdot p_5)-2\mu^2(p_5\cdot p_{16})(p_2\cdot p_{16})\\
&\qquad\qquad +\mu^2 s_{16}(p_2\cdot p_5)+\mathcal{P}\left(2,3,4\right)\\
&\qquad = c_6\mu^2\left(s_{23}s_{45}+s_{24}s_{35}+s_{25}s_{34}\right)-2\mu^2 s_{12}(p_5\cdot p_{16})+4\mu^2 s_{12}(p_5\cdot p_{16})\\
&\qquad\quad -2\mu^2 s_{12}(p_5\cdot p_{16})\\
&\qquad = c_6\mu^2\left(s_{23}s_{45}+s_{24}s_{35}+s_{25}s_{34}\right). \qquad (4.3.23)
\end{aligned}$$

Somewhat miraculously all of the terms cancel except for the unknown contact term. Since this is manifestly $\mathcal{O}(p_5)$, we must choose $c_6 = 0$ to satisfy the constraint of T-duality. Again, this same conclusion can also be reached after a rather lengthy numerical calculation involving the massive KLT relations (4.3.55). In Appendix E we give the explicit calculation of $c_8$, again we confirm the result of the numerical KLT calculation. In the next subsection we will give an explicit all-multiplicity proof that the T-duality constraints require $c_n = 0$ for $n > 4$.

#### 4.3.2.2 Small Mass Expansion and the Absence of Contact Terms

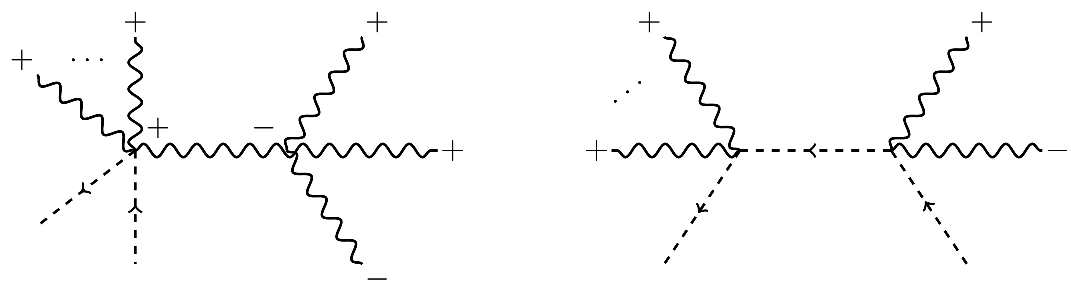
That the 6-point dimensional reduction and soft limit calculation gave $c_6 = 0$ is somewhat remarkable, and could not easily have been anticipated without a detailed calculation. For $n \geq 8$ the conclusion that $c_n = 0$ is less mysterious and can be argued on general grounds by considering the structure of the mDBI$_4$ amplitudes *as an expansion around* the $\mu^2 \to 0$ limit. In Appendix D we show that there is a unique contact term at each multiplicity of the form

$$\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^+, n_{\overline{\phi}}\right)\Big|_{\text{contact}} = c_n\mu^2\left([23]^2[45]^2\ldots[n-2,n-1]^2+\ldots\right). \quad (4.3.24)$$

Dimensionally reducing to 3d this becomes

$$\xrightarrow{3d} c_n\mu^2\left(s_{23}s_{45}\ldots s_{n-2,n-1}+\ldots\right), \qquad (4.3.25)$$

which is manifestly $\mathcal{O}(p_{n-1})$ in the soft limit of particle $n-1$. If $c_n \neq 0$ then this term must cancel against some term in the factoring part of the Ansatz to give the correct $\mathcal{O}(p_{n-1}^2)$ soft limit. To show that this can never happen we expand in the limit $\mu^2 \to 0$. The contact terms clearly always contribute at $\mathcal{O}(\mu^2)$. Since $\mu^2$ is a free parameter (corresponding to our choice of momenta in the 4 and 5 directions from the 6d perspective), the T-duality constraints should apply order-by-order in the expansion. For a non-trivial cancellation between the contact and factoring terms to occur, the factoring terms must give a contribution at $\mathcal{O}(\mu^2)$. If such a contribution exists then we must be able to identify a factorization channel for which the product of the leading small mass behavior on both sides is $\mathcal{O}(\mu^2)$. Since negative and odd powers of $\mu$ do not appear, one half of the factorization diagram must be $\mathcal{O}(\mu^0)$. At each multiplicity there are only two possible factorization channels which can give such a contribution:




both of which have the form of a lower-point NSD amplitude glued to an $\mathcal{O}(\mu^0)$ 4-point amplitude. For $n = 8$, the $\mathcal{O}(\mu^2)$ contribution to the NSD amplitude arises solely from the contact term which we explicitly verified (by two different methods) was absent. So we conclude there cannot be an $\mathcal{O}(\mu^2)$ contribution to the $n = 8$ MHV amplitude and hence no contact term. We can continue in this way and make an inductive argument that the absence of contact terms at $n-2$-point implies the absence of contact terms at $n$-point. Together with the explicit $n = 6$ case, we find that all higher point contact terms are zero in mDBI$_4$, the amplitudes are (almost) as simple as possible. We will leverage this simplicity in the following section to construct all-multiplicity one-loop integrands for the SD and NSD sectors of BI$_4$.

### 4.3.3 Alternative Approach to Contact Terms: Massive KLT Relations

As discussed in Section 4.2.2, the tree-level amplitudes of Born-Infeld in $d$-dimensions are given by the KLT product

$$\text{BI}_d = \text{YM}_d \otimes_{\text{KLT}} \chi\text{PT}_d, \tag{4.3.26}$$

where $\chi\text{PT}_d$ denotes the $\frac{SU(N)\times SU(N)}{SU(N)}$ non-linear sigma model in $d$-dimensions. Beginning with $d = 6$ we can (formally) calculate tree amplitudes in BI$_6$ from the tree amplitudes for YM$_6$ and $\chi$PT$_6$ using the dimension independent form of the KLT product. Since we do not require the completely general 6d Born-Infeld amplitudes, only the configuration in Figure 4.1, we can dimensionally reduce the 6d KLT relations into a form of *massive* KLT relations by separating the 4d and extra-dimensional components of the momenta. This amounts to taking the dimension independent form the KLT relations and making the replacements

$$s_{1i} \to s_{1i} + \mu^2, \quad s_{nj} \to s_{nj} + \mu^2, \tag{4.3.27}$$

where $i \neq n$ and $j \neq 1$ (Note that we are defining our Mandelstam invariants as $s_{ij} \equiv (p_i + p_j)^2$). Using this prescription the needed KLT relations

$$\text{mDBI}_4 = \text{YM+mAdj}_4 \otimes_{\text{KLT}} \text{m}\chi\text{PT}_4, \tag{4.3.28}$$

up to $n = 8$ take the explicit form [133]

$$\mathcal{A}_4^{\text{mDBI}_4}\left(1_\phi, 2_\gamma, 3_\gamma, 4_{\overline{\phi}}\right) = (s_{12} + \mu^2)\mathcal{A}_4^{\text{YM+mAdj}_4}[1_\phi, 2_g, 3_g, 4_{\overline{\phi}}]\mathcal{A}_4^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 2, \mathbf{4}, 3], \tag{4.3.29}$$

$$\begin{aligned}
&\mathcal{A}_6^{\text{mDBI}_4}\left(1_\phi, 2_\gamma, 3_\gamma, 4_\gamma, 5_\gamma, 6_{\overline{\phi}}\right) \\
&\quad = (s_{12} + \mu^2)s_{45}\mathcal{A}_6^{\text{YM+mAdj}_4}[1_\phi, 2_g, 3_g, 4_g, 5_g, 6_{\overline{\phi}}] \\
&\qquad \times \left(s_{35}\mathcal{A}_6^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 5, 3, 4, \mathbf{6}, 2] + (s_{34} + s_{35})\mathcal{A}_6^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 5, 4, 3, \mathbf{6}, 2]\right) \\
&\qquad + \mathcal{P}(2, 3, 4),
\end{aligned} \tag{4.3.30}$$

$$\begin{aligned}
&\mathcal{A}_8^{\text{mDBI}_4}\left(1_\phi, 2_\gamma, 3_\gamma, 4_\gamma, 5_\gamma, 6_\gamma, 7_\gamma, 8_{\overline{\phi}}\right) \\
&\quad = (s_{12} + \mu^2)s_{67}\mathcal{A}_8^{\text{YM+mAdj}_4}[1_\phi, 2_g, 3_g, 4_g, 5_g, 6_g, 7_g, 8_{\overline{\phi}}] \\
&\qquad \times \Big[(s_{13} + \mu^2)s_{14}\Big(s_{57}\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 5, 6, \mathbf{8}, 2, 3, 4] \\
&\qquad\qquad\qquad\qquad + (s_{57} + s_{56})\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 6, 5, \mathbf{8}, 2, 3, 4]\Big) \\
&\qquad\quad + (s_{13} + \mu^2)(s_{14} + s_{34} + \mu^2)\Big(s_{57}\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 5, 6, \mathbf{8}, 2, 4, 3] \\
&\qquad\qquad\qquad\qquad + (s_{57} + s_{56})\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 6, 5, \mathbf{8}, 2, 4, 3]\Big) \\
&\qquad\quad + (s_{14} + \mu^2)(s_{13} + s_{23} + \mu^2)\Big(s_{57}\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 5, 6, \mathbf{8}, 3, 2, 4] \\
&\qquad\qquad\qquad\qquad + (s_{57} + s_{56})\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 6, 5, \mathbf{8}, 3, 2, 4]\Big) \\
&\qquad\quad + (s_{13} + s_{23} + \mu^2)(s_{14} + s_{24} + \mu^2)\Big(s_{57}\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 5, 6, \mathbf{8}, 3, 4, 2] \\
&\qquad\qquad\qquad\qquad + (s_{57} + s_{56})\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 6, 5, \mathbf{8}, 3, 4, 2]\Big) \\
&\qquad\quad + (s_{13} + \mu^2)(s_{14} + s_{24} + s_{34} + \mu^2)\Big(s_{57}\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 5, 6, \mathbf{8}, 4, 2, 3] \\
&\qquad\qquad\qquad\qquad + (s_{57} + s_{56})\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1}, 7, 6, 5, \mathbf{8}, 4, 2, 3]\Big)
\end{aligned}$$

$$+ (s_{13}+s_{23}+\mu^2)(s_{14}+s_{34}+s_{24}+\mu^2)\Big(s_{57}\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1},7,5,6,\mathbf{8},4,3,2]$$

$$+(s_{57}+s_{56})\mathcal{A}_8^{\text{m}\chi\text{PT}_4}[\mathbf{1},7,6,5,\mathbf{8},4,3,2]\Big)\Big]$$

$$+\mathcal{P}\,(2,3,4,5,6)\,. \tag{4.3.31}$$

In the m$\chi$PT amplitudes bolded momenta denote massive particles.

Note that these expressions differ by an overall sign from the expressions given in [71] due to our conventions for the Mandelstam invariants. Below we will describe the calculation of both YM + $\text{mAdj}_4$ and m$\chi\text{PT}_4$ amplitudes and then give the result of the double copy.

#### 4.3.3.1 YM+mAdj$_4$ from Massive BCFW

The needed tree-level amplitudes of YM+mAdj can be calculated using BCFW recursion from 3-point input. Since this model should have only marginal couplings between the gluons and massive adjoint scalar, the tree-level amplitudes are completely fixed by gauge invariance. This approach was first used in [149], below we give a brief review.

The seed amplitudes for the recursion are

$$\mathcal{A}_3^{\text{YM+mAdj}_4}[1_\phi,2_g^+,3_{\overline{\phi}}] = -\frac{[2|p_1|q\rangle}{\langle 2q\rangle}, \qquad \mathcal{A}_3^{\text{YM+mAdj}_4}[1_\phi,2_g^-,3_{\overline{\phi}}] = \frac{[\tilde{q}|p_1|2\rangle}{[\tilde{q}\;2]}, \tag{4.3.32}$$

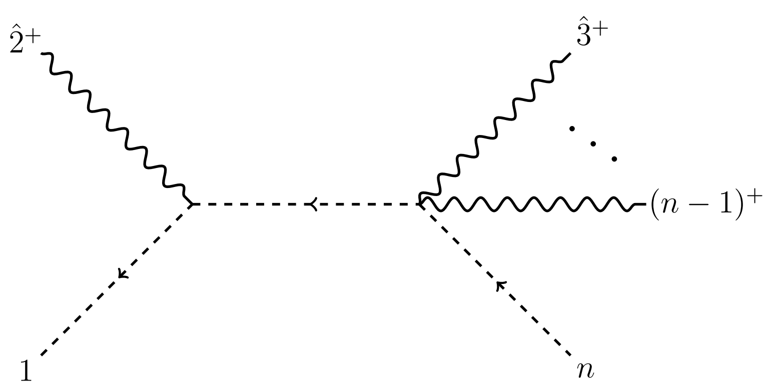
where $|q\rangle$ and $|\tilde{q}]$ are arbitrary. We want to calculate NSD amplitudes

$$\mathcal{A}_n^{\text{YM+mAdj}_4}[1_\phi,2_g^+,3_g^+\ldots,(n-1)_g^+,n_{\overline{\phi}}], \tag{4.3.33}$$

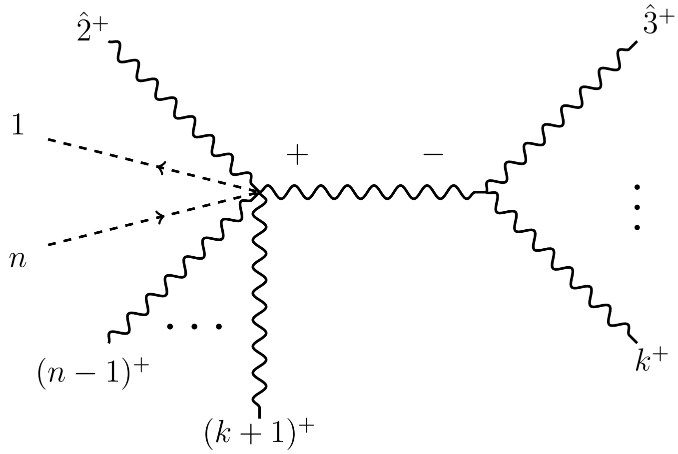
using a BCFW shift

$$|\hat{2}\rangle = |2\rangle - z|3\rangle, \qquad |\hat{3}] = |3] + z|2]. \tag{4.3.34}$$

With the given color-ordering (and the fact that the shifted lines must sit on opposite sides of the factorization diagram) there are two types of factorization channel which could contribute:



and




Interestingly, the second diagram never contributes. The argument for this is has two parts, first we consider diagrams with $k > 4$. In this case the right-hand amplitude is of the form $\mathcal{A}_{k-1}^{\text{YM+mAdj}_4}[-,+,\ldots,+]$ which vanishes at tree-level. For the case $k = 4$ the right-hand amplitude is simply the pure Yang-Mills 3-point amplitude[7]

$$\mathcal{A}_3^{\text{YM+mAdj}_4}\left[(-\hat{p}_{34})_g^-, \hat{3}_g^+, 4_g^+\right] = \frac{[\hat{3}4]^3}{[4,-\hat{p}_{34}][-\hat{p}_{34},\hat{3}]}. \tag{4.3.35}$$

On the factorization channel we have $[\hat{3}4] = 0$ and therefore this amplitude vanishes. So we see that only a single factorization channel contributes at each recursive step. Explicitly the BCFW recursion relation takes the form

$$\begin{aligned}&\mathcal{A}_n^{\text{YM+mAdj}_4}\left[1_\phi, 2_g^+, 3_g^+, \ldots, (n-1)_g^+, n_{\overline{\phi}}\right]\\&\quad= \frac{\mathcal{A}_3^{\text{YM+mAdj}_4}[1_\phi, \hat{2}_g^+, (-\hat{p}_{12})_{\overline{\phi}}]\mathcal{A}_{n-1}^{\text{YM+mAdj}_4}[(\hat{p}_{12})_\phi, \hat{3}_g^+, 4_g^+, \ldots, (n-2)_g^+, (n-1)_{\overline{\phi}}]}{s_{12}+\mu^2}.\end{aligned} \tag{4.3.36}$$

We will now use this to calculate the amplitudes up to $n = 8$. Here (and subsequently) we will use the convenient shorthand notation

$$p_{1,k} \equiv p_{12\ldots k}, \qquad D_n \equiv \langle 23\rangle\langle 34\rangle \ldots \langle n{-}2, n{-}1\rangle(s_{12}{+}\mu^2)(s_{123}{+}\mu^2)\ldots(s_{12\ldots n-2}{+}\mu^2). \tag{4.3.37}$$

[7]Here and subsequently, we use the convention $|-p] = i|p]$ and $|-p\rangle = i|p\rangle$. This is because the prescription for dimensional reduction to $3d$ we use in Section 4.3.2 requires that we treat the angle and square spinors "democratically". A consequence of this convention choice is that the Parke-Taylor amplitudes acquire an additional factor of $-1$ for an even number of external states.

At 4-point we need both the NSD and MHV amplitudes

$$\mathcal{A}_4^{\text{YM+mAdj}_4}[1_\phi, 2_g^+, 3_g^+, 4_{\overline{\phi}}] = -\frac{\mu^2[23]}{\langle 23\rangle(s_{12}+\mu^2)}, \tag{4.3.38}$$

and

$$\mathcal{A}_4^{\text{YM+mAdj}_4}[1_\phi, 2_g^+, 3_g^-, 4_{\overline{\phi}}] = -\frac{\langle 3|p_1|2]^2}{s_{23}(s_{12}+\mu^2)}. \tag{4.3.39}$$

At 6-point we will only need amplitudes in the NSD sector

$$\mathcal{A}_6^{\text{YM+mAdj}_4}\left[1_\phi, 2_g^+, 3_g^+, 4_g^+, 5_g^+, 6_{\overline{\phi}}\right] = -\frac{\mu^2[2|p_1\cdot p_{23}\cdot p_{45}\cdot p_6|5]}{D_6}. \tag{4.3.40}$$

Similarly at 8-point

$$\begin{aligned}
&\mathcal{A}_8^{\text{YM+mAdj}_4}\left[1_\phi, 2_g^+, 3_g^+, 4_g^+, 5_g^+, 6_g^+, 7_g^+, 8_{\overline{\phi}}\right] \\
&\quad= \frac{1}{D_8}\Big[-(\mu^2)^3[2|p_1\cdot p_{23}\cdot p_{67}\cdot p_8|7] + (\mu^2)^2[2|p_1\cdot p_{23}\cdot p_{4,8}\cdot p_{5,8}\cdot p_{67}\cdot p_8|7] \\
&\qquad\quad + (\mu^2)^2[2|p_1\cdot p_{23}\cdot p_{5,8}\cdot p_{6,8}\cdot p_{67}\cdot p_8|7] \\
&\qquad\quad -\mu^2[2|p_1\cdot p_{23}\cdot p_{4,8}\cdot p_{5,8}\cdot p_{5,8}\cdot p_{6,8}\cdot p_{67}\cdot p_8|7]\Big]\,.
\end{aligned} \tag{4.3.41}$$

All multiplicity results for these amplitudes have been calculated in [150], but we will not need explicit expressions beyond 8-points.

#### 4.3.3.2 m$\chi$PT$_4$ from Soft Limits and Dimensional Reduction

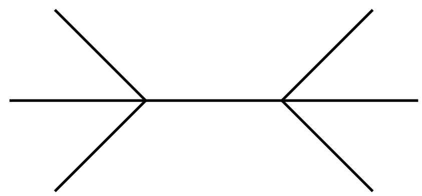
The needed tree level amplitudes for $\chi$PT$_d$ can be calculated using the *soft bootstrap* approach [1, 38, 53, 76, 79, 81, 95, 120, 151–154]. While it is certainly possible to setup formal recursion relations analogous to the BCFW recursion used above (this is the so-called *subtracted recursion* [79, 80]), in practice since this is such a simple model there is a more efficient approach. We note that locality is manifest in the $\chi$PT amplitudes, and so we can treat the contact terms of lower-point amplitudes as “vertex rules”, gluing them together in a diagrammatic expansion. This will automatically generate expressions with the correct factorization properties (which can be verified straightforwardly *post hoc* by computing residues), the remaining ambiguity is contained in the contact terms. These ambiguous contributions can then be determined by imposing the Adler zero, that is, single soft limit which vanish at $\mathcal{O}(p)$ [37].

We start with the flavor-ordered 4-point amplitude

$$\mathcal{A}_4^{\chi\text{PT}_d}[1,2,3,4] = s_{13}. \tag{4.3.42}$$

With the dimensionful coupling suppressed, the $\chi\mathrm{PT}_d$ tree-amplitudes take a dimension independent form. Similar to the definition of $\mathrm{mDBI}_4$ we define the model $\mathrm{m}\chi\mathrm{PT}_4$ as the tree amplitudes of $\chi\mathrm{PT}_6$ with momenta in the configuration given in Figure 4.1. Operationally these amplitudes are calculated using the replacement rules (4.3.27), on the $\chi\mathrm{PT}_d$ amplitudes, similar to the way we derived the massive KLT relations above.

Now we turn to the explicit calculation of the 6-point $\chi\mathrm{PT}_d$ amplitude. In this case the factoring part of the amplitude corresponds to diagrams with a unique topology



There are three inequivalent cyclic permutations of the external labels $[1,2,3,4,5,6]$, so the factoring part of the six point amplitude has the form

$$\mathcal{A}_6^{\chi\mathrm{PT}_d}[1,2,3,4,5,6]\bigg|_{\text{factoring}} = \frac{s_{13}s_{46}}{s_{123}} + \frac{s_{24}s_{51}}{s_{234}} + \frac{s_{35}s_{62}}{s_{345}}. \tag{4.3.43}$$

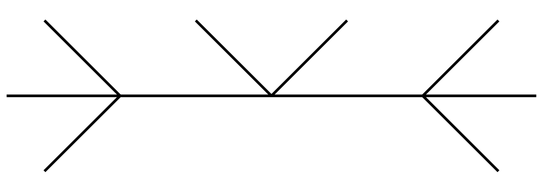
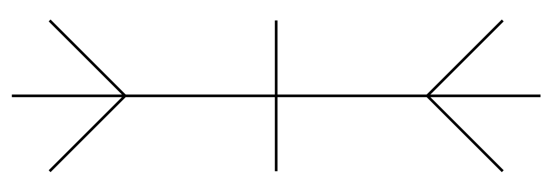
This differs from the full answer by a possible contact term. Such a contact contribution is fixed by demanding that the amplitude vanishes in the soft limit of each particle. It is straightforward to verify that the following expression satisfies all of the aforementioned properties

$$\mathcal{A}_6^{\chi\mathrm{PT}_d}[1,2,3,4,5,6] = \frac{s_{13}s_{46}}{s_{123}} + \frac{s_{24}s_{51}}{s_{234}} + \frac{s_{35}s_{62}}{s_{345}} - s_{135}. \tag{4.3.44}$$

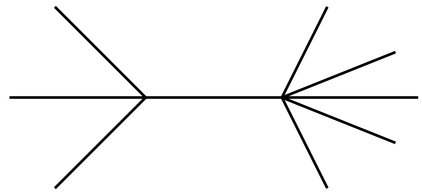
We can then convert this into an $\mathrm{m}\chi\mathrm{PT}_4$ amplitude with particles 1 and 5 massive for later use in the KLT product

$$\mathcal{A}_6^{\mathrm{m}\chi\mathrm{PT}_4}[\mathbf{1},2,3,4,\mathbf{5},6] = \frac{(s_{13}+\mu^2)s_{46}}{s_{123}+\mu^2} + \frac{s_{24}s_{51}}{s_{234}} + \frac{(s_{35}+\mu^2)s_{62}}{s_{345}+\mu^2} - s_{135}. \tag{4.3.45}$$

For $n=8$ there are three distinct factorization topologies we need to consider, two constructed from 4-point vertices

and one from a 4-point and a 6-point vertex



It is straightforward to write down the factoring part of this amplitude

$$\mathcal{A}_8^{\chi\mathrm{PT}_d}[1,2,3,4,5,6,7,8]\Big|_{\text{factoring}} = \frac{s_{13}s_{1235}s_{68}}{s_{123}s_{678}} + \frac{1}{2}\left(\frac{s_{13}s_{48}s_{57}}{s_{123}s_{567}}\right) - \frac{s_{13}s_{468}}{s_{123}} + \mathcal{C}\left(1,2,3,4,5,6,7,8\right). \tag{4.3.46}$$

where $\mathcal{C}$ denotes the sum over all *cyclic* permutations. The contact terms we need to add can be found straightforwardly by taking soft limits, the result is

$$\mathcal{A}_8^{\chi\mathrm{PT}_d}[1,2,3,4,5,6,7,8] = \left[\frac{s_{13}s_{1235}s_{68}}{s_{123}s_{678}} + \frac{1}{2}\left(\frac{s_{13}s_{48}s_{57}}{s_{123}s_{567}}\right) - \frac{s_{13}s_{468}}{s_{123}} + \mathcal{C}\left(1,2,3,4,5,6,7,8\right)\right] + s_{2468}. \tag{4.3.47}$$

Constructing the m$\chi$PT$_4$ amplitude with particle 1 and 5 massive gives

$$\begin{aligned}
&\mathcal{A}_8^{\mathrm{m}\chi\mathrm{PT}_4}[\mathbf{1},2,3,4,\mathbf{5},6,7,8] \\
&= \frac{(s_{13}+\mu^2)s_{1235}s_{68}}{(s_{123}+\mu^2)s_{678}} + \frac{(s_{13}+\mu^2)s_{48}(s_{57}+\mu^2)}{(s_{123}+\mu^2)(s_{567}+\mu^2)} - \frac{(s_{13}+\mu^2)s_{468}}{s_{123}+\mu^2} + \frac{s_{24}s_{2346}(s_{71}+\mu^2)}{s_{234}(s_{781}+\mu^2)} \\
&\quad + \frac{s_{24}s_{51}s_{68}}{s_{234}s_{678}} - \frac{s_{24}s_{571}}{s_{234}} + \frac{(s_{35}+\mu^2)(s_{3457}+\mu^2)s_{82}}{(s_{345}+\mu^2)(s_{812}+\mu^2)} + \frac{(s_{35}+\mu^2)s_{62}(s_{71}+\mu^2)}{(s_{345}+\mu^2)(s_{781}+\mu^2)} \\
&\quad - \frac{(s_{35}+\mu^2)s_{682}}{s_{345}+\mu^2} + \frac{s_{46}(s_{4568}+\mu^2)(s_{13}+\mu^2)}{(s_{456}+\mu^2)(s_{123}+\mu^2)} + \frac{s_{46}s_{73}s_{82}}{(s_{456}+\mu^2)(s_{812}+\mu^2)} - \frac{s_{46}(s_{713}+\mu^2)}{s_{456}+\mu^2} \\
&\quad + \frac{(s_{57}+\mu^2)s_{5671}s_{24}}{(s_{567}+\mu^2)s_{234}} - \frac{(s_{57}+\mu^2)s_{824}}{s_{567}+\mu^2} + \frac{s_{68}(s_{6781}+\mu^2)(s_{35}+\mu^2)}{s_{678}(s_{345}+\mu^2)} - \frac{s_{68}s_{135}}{s_{678}} \\
&\quad + \frac{(s_{71}+\mu^2)(s_{7812}+\mu^2)s_{46}}{(s_{781}+\mu^2)(s_{456}+\mu^2)} - \frac{(s_{71}+\mu^2)s_{246}}{s_{781}+\mu^2} + \frac{s_{82}(s_{8123}+\mu^2)(s_{57}+\mu^2)}{(s_{812}+\mu^2)(s_{567}+\mu^2)} \\
&\quad - \frac{s_{82}(s_{357}+\mu^2)}{s_{812}+\mu^2} + s_{2468}.
\end{aligned} \tag{4.3.48}$$

Simple closed form expressions for all $\chi$PT$_d$ amplitudes are not known, but this procedure is simple enough that it can be implemented efficiently to calculate amplitudes up to the desired multiplicity. As in the previous section we will only need explicit expressions up to $n = 8$.

#### 4.3.3.3 Result of Double Copy

We can begin with the calculation of the 4-point amplitudes of $\mathrm{mDBI}_4$, which are simple enough to be evaluated by hand without difficulty

$$\begin{aligned}\mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_{\overline{\phi}}\right) &= (s_{12}+\mu^2)\mathcal{A}_4^{\text{YM+mAdj}}[1_\phi, 2_g^+, 3_g^+, 4_{\overline{\phi}}]\mathcal{A}_4^{\mathrm{m}\chi\mathrm{PT}}[\mathbf{1}, 2, \mathbf{4}, 3] \\ &= (s_{12}+\mu^2)\left[-\frac{\mu^2[23]}{\langle 23\rangle(s_{12}+\mu^2)}\right][s_{23}] \\ &= -\mu^2[23]^2, \end{aligned} \tag{4.3.49}$$

and

$$\begin{aligned}\mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^-, 4_{\overline{\phi}}\right) &= (s_{12}+\mu^2)\mathcal{A}_4^{\text{YM+mAdj}}[1_\phi, 2_g^+, 3_g^-, 4_{\overline{\phi}}]\mathcal{A}_4^{\mathrm{m}\chi\mathrm{PT}}[\mathbf{1}, 2, \mathbf{4}, 3] \\ &= (s_{12}+\mu^2)\left[-\frac{\langle 3|p_1|2]^2}{s_{23}(s_{12}+\mu^2)}\right][s_{23}] \\ &= -\langle 3|p_1|2]^2. \end{aligned} \tag{4.3.50}$$

We will also need the 4-point pure Born-Infeld amplitude. This can also be calculated with the (massless) KLT product using the 4-point Parke-Taylor gluon amplitude

$$\begin{aligned}\mathcal{A}_4^{\mathrm{mDBI}_4}\left(1_\gamma^+, 2_\gamma^+, 3_\gamma^-, 4_\gamma^-\right) &= s_{12}\left[-\frac{[12]^3}{[23][34][41]}\right][s_{23}] \\ &= [12]^2\langle 34\rangle^2. \end{aligned} \tag{4.3.51}$$

Notice that due to our convention choice (see comments in footnote 7), the Parke-Taylor amplitude above has an additional factor of $-1$.

Simplifying the massive KLT relations algebraically beyond 4-point is a daunting task. Fortunately it is straightforward to construct a general Ansatz for the higher-multiplicity amplitudes. Beginning with the NSD 6-point amplitude we know the answer should have the form

$$\begin{aligned}&\mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_{\overline{\phi}}\right) \\ &\quad = \frac{1}{4}\left[\frac{(\mu^2)^2[23]^2[45]^2}{s_{123}+\mu^2} + \mathcal{P}\left(2,3,4,5\right)\right] + c_6\mu^2\left([23]^2[45]^2 + [24]^2[35]^2 + [25]^2[34]^2\right). \end{aligned} \tag{4.3.52}$$

This expression has the correct factorization singularities consistent with the known 4-point amplitudes, and a polynomial ambiguity parametrized by a single coefficient $c_6$, as discussed above. To determine the coefficient $c_6$ we numerically evaluate the KLT sum (4.3.30) on several sets of randomly generated kinematic variables and compare with a numerical evaluation of the Ansatz.

For more than one choice of kinematics this overconstrains the problem and allows us to both verify the validity of the Ansatz and determine the value of the coefficient. Doing so we find that the Ansatz is valid and $c_6 = 0$; the amplitude is simply

$$\mathcal{A}_6^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_{\overline{\phi}}\right) = \frac{1}{4}\left[\frac{(\mu^2)^2[23]^2[45]^2}{s_{123}+\mu^2}\right] + \mathcal{P}\left(2,3,4,5\right). \tag{4.3.53}$$

Next we calculate the MHV 6-point amplitude. As discussed in Section 4.3.1, in this case there are *no* contact terms consistent with little group scaling and Bose symmetry. There is then no ambiguity in the answer, the result of gluing together the 4-point amplitudes on factorization channels is the unique correct result. We find

$$\begin{aligned}\mathcal{A}_6^{\mathrm{mDBI}_4}&\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^-, 6_{\overline{\phi}}\right)\\ &= \frac{\mu^2}{2}\left[\frac{[23]^2\langle 5|p_6|4]^2}{s_{123}+\mu^2} + \frac{[34]^2\langle 5|p_1|2]^2}{s_{125}+\mu^2} + \frac{[34]^2\langle 5|p_{34}|2]^2}{s_{126}}\right] + \mathcal{P}\left(2,3,4\right).\end{aligned} \tag{4.3.54}$$

At 8-point the method is the same, we begin with the calculation of the NSD amplitude. Using the result $c_6 = 0$, we should use an Ansatz of the form

$$\begin{aligned}&\mathcal{A}_8^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+, 7_\gamma^+, 8_{\overline{\phi}}\right)\\ &= -\frac{1}{8}\left[\frac{(\mu^2)^3[23]^2[45]^2[67]^2}{(s_{123}+\mu^2)(s_{678}+\mu^2)}\right] + c_8\mu^2[23]^2[45]^2[67]^2 + \mathcal{P}\left(2,3,4,5,6,7\right).\end{aligned} \tag{4.3.55}$$

Explicit numerical evaluation of the massive KLT relations reveals the surprising result that $c_8 = 0$ also! Finally, as above the MHV 8-point amplitude is completely fixed by factorization

$$\begin{aligned}\mathcal{A}_8^{\mathrm{mDBI}_4}&\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+, 7_\gamma^-, 8_{\overline{\phi}}\right)\\ = -\frac{(\mu^2)^2}{4}&\left[\frac{[23]^2[45]^2\langle 7|p_8|6]^2}{(s_{123}+\mu^2)(s_{678}+\mu^2)} + \frac{[23]^2[45]^2\langle 7|p_{123}|6]^2}{(s_{123}+\mu^2)(s_{458}+\mu^2)} + \frac{[23]^2[45]^2\langle 7|p_1|6]^2}{(s_{167}+\mu^2)(s_{458}+\mu^2)}\right.\\ &\left. + \frac{[34]^2[56]^2\langle 7|p_{34}|2]^2}{s_{347}(s_{568}+\mu^2)} + \frac{[23]^2[56]^2\langle 7|p_{56}|4]^2}{s_{567}(s_{123}+\mu^2)}\right] + \mathcal{P}\left(2,3,4,5,6\right).\end{aligned} \tag{4.3.56}$$

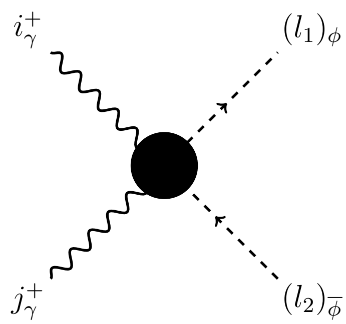
## 4.4 All Multiplicity Rational One-Loop Amplitudes

### 4.4.1 Diagrammatic Rules for Constructing Loop Integrands

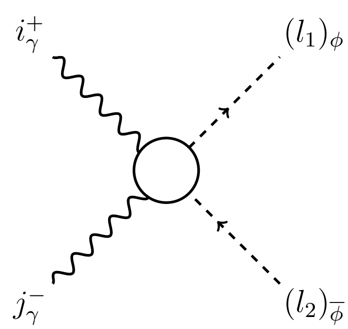
With the results in the previous section, and the discussion in Section 4.2, we have in principle obtained a complete understanding of the structure of the $d$-dimensional unitarity cuts of SD and NSD $\mathrm{BI}_4$ one-loop integrands. Our goal is now to use this to engineer the explicit form of the integrands and then integrate them to obtain the full amplitudes. Ordinarily, gluing together on-shell tree-amplitudes into full loop integrands is a delicate business. Constructing expressions with the correct cuts in one channel may give *polluting* contributions to another channel. Separating these contributions and building up loop integrands in a systematic way has been a subject of intense study over the past several decades [136].

Fortunately for us, the mDBI$_4$ tree amplitudes are of sufficiently simple form that it is straightforward to construct integrands with all of the correct cuts using a set of *diagrammatic rules*. There are two properties that allow us to do this; first, locality is manifest in the mDBI$_4$ amplitudes, and second, due to the absence of contact terms above $n = 4$ the number of *elementary* vertex rules is strictly finite. Notice how much simpler this is than calculating loop diagrams directly from ordinary Feynman rules! If we were calculating loop amplitudes in Born-Infeld the old-fashioned way we would need to calculate new (and increasingly complicated) Feynman vertex rules at each multiplicity.

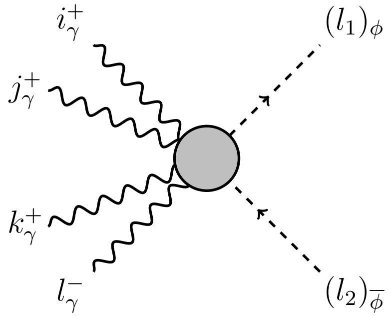
Since we are constructing loop integrands in the scalar loop representation (4.2.1) we will construct a diagrammatic representation in which each diagram consists of a scalar loop *decorated* with any of the following vertex factors:




$$= \quad -\mu^2[ij]^2$$

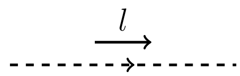



$$= \quad -\langle j|l_1|i]^2$$




$$= \frac{\mu^2[k|p_{ij}|l\rangle^2[ij]^2}{s_{ijl}} + \mathcal{C}(i,j,k)$$

Here $+\,\mathcal{C}(i,j,k)$ denotes the sum over cyclic permutations, all of the momenta are defined to be out-going with photon lines on-shell, while the scalar lines are off-shell. These vertex rules can be glued together on scalar lines in the usual way with the standard massive scalar propagator



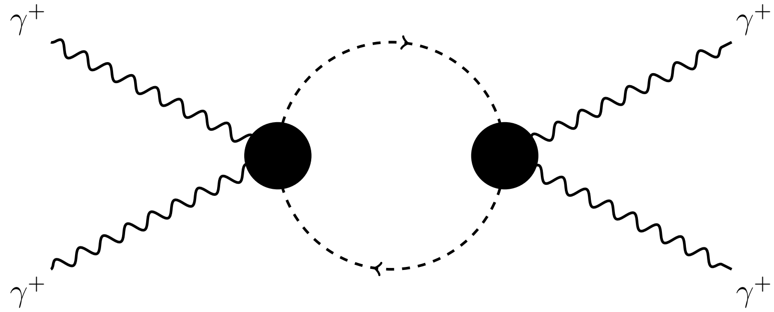

$$= \frac{1}{l^2+\mu^2}$$

These diagrammatic rules can be justified *post hoc*, by verifying that the resulting loop integrands have the correct massive scalar cuts. These are not Feynman rules in the usual sense, and have not been derived from a Lagrangian. This is especially clear in the 6-point vertex rule (denoted with a gray blob), which is a non-local expression; the poles encode factorization singularities into Born-Infeld photons. Due to the helicity selection rules of $\mathrm{BI}_4$ at tree-level arising from supersymmetric truncation, no further photonic singularities can appear in amplitudes with at most a single negative helicity external state.

In the following sections we will give explicit examples of the applications of these diagrammatic rules to 4- and 6-point SD and NSD loop integrands, and then present explicit expressions for the all-multiplicity results together with the integrated expressions at $\mathcal{O}(\epsilon^0)$.

### 4.4.2 Self-Dual Sector

In the self-dual sector, since there are only positive helicity external states, at each multiplicity there is only a single topologically distinct diagram and it is constructed solely from black vertices. Beginning with $n=4$, the diagram has the form:




There are three non-trivial permutations of the external labels. The integrand is then

$$\mathcal{I}_4^{\text{SD}}[l;\mu^2] = \frac{1}{2}\left[\frac{(\mu^2)^2[12]^2[34]^2}{[l^2+\mu^2]\,[(l-p_{12})^2+\mu^2]} + \mathcal{P}\left(2,3,4\right)\right], \tag{4.4.1}$$

where the factor of $\frac{1}{2}$ compensates for the equivalent permutations in $\mathcal{P}\left(2,3,4\right)$ that are summed over.

We now explicitly verify that the diagrammatic rules of Section 4.4.1 yield an integrand that satisfies the cut conditions. Since the integrand has only one distinct two-particle cut (all others are related by label permutations), we choose to consider the $p_{12}$-cut. When the on-shell conditions $l^2 = -\mu^2$ and $(l-p_{12})^2 = -\mu^2$ are imposed, the integrand yields

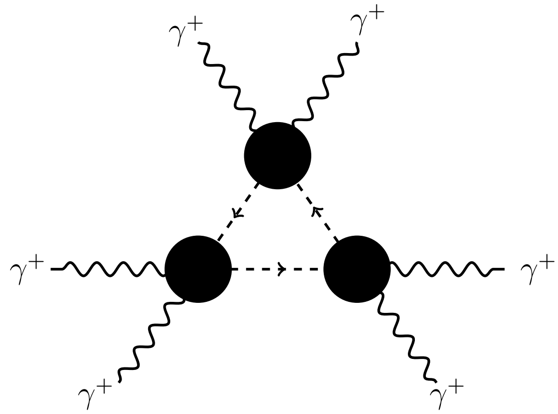
$$\begin{aligned}
&\left[l^2+\mu^2\right]\left[(l-p_{12})^2+\mu^2\right]\mathcal{I}_4^{\text{SD}}[l;\mu^2]\Big|_{p_{12}\text{-cut}} \\
&\qquad\qquad = \mathcal{A}_4\left(1_\gamma^+,2_\gamma^+,-l_\phi,(l-p_{12})_{\bar{\phi}}\right)\mathcal{A}_4\left(l_{\bar{\phi}},(p_{12}-l)_\phi,3_\gamma^+,4_\gamma^+\right) \\
&\qquad\qquad = (\mu^2)^2[12]^2[34]^2
\end{aligned} \tag{4.4.2}$$

as expected. The NSD amplitudes above are given in (4.3.49).

Using the general result for rational loop integrals (F.1.14) gives

$$\begin{aligned}
&\mathcal{A}_4^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+,2_\gamma^+,3_\gamma^+,4_\gamma^+\right) \\
&\quad = \frac{1}{2}\int\frac{\mathrm{d}^4 l}{(2\pi)^4}\int\frac{\mathrm{d}^{-2\epsilon}\mu}{(2\pi)^{-2\epsilon}}\left[\frac{(\mu^2)^2[12]^2[34]^2}{[l^2+\mu^2]\,[(l-p_{12})^2+\mu^2]}+\mathcal{P}\left(2,3,4\right)\right] \\
&\quad = [12]^2[34]^2 I_2^{d=4-2\epsilon}[(\mu^2)^2;p_{12}] + [13]^2[24]^2 I_2^{d=4-2\epsilon}[(\mu^2)^2;p_{13}] \\
&\qquad + [14]^2[23]^2 I_2^{d=4-2\epsilon}[(\mu^2)^2;p_{14}] \\
&\quad = -\frac{i}{960\pi^2}\left([12]^2[34]^2 s_{12}^2 + [13]^2[24]^2 s_{13}^2 + [14]^2[23]^2 s_{14}^2\right) + \mathcal{O}(\epsilon).
\end{aligned} \tag{4.4.3}$$

Similarly for $n=6$ there is a unique topologically distinct class of diagram:




The integrand is then given by

$$\mathcal{I}_6^{\text{SD}}[l;\mu^2] = -\frac{1}{4}\left[\frac{(\mu^2)^3[12]^2[34]^2[56]^2}{[l^2+\mu^2]\left[(l-p_{34})^2+\mu^2\right]\left[(l+p_{12})^2+\mu^2\right]} + \mathcal{P}\left(2,3,4,5,6\right)\right]. \tag{4.4.4}$$

The integrand has only one distinct cut into tree-level amplitudes. Consider for example the integrand on the $p_{12}$-cut,

$$\begin{aligned}
&\left[l^2+\mu^2\right]\left[(l+p_{12})^2+\mu^2\right]\mathcal{I}_6^{\text{SD}}[l;\mu^2]\Big|_{p_{12}\text{-cut}} \\
&\quad = \mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_\phi, -(l+p_{12})_{\bar\phi}\right)\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{12})_\phi, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+\right) \\
&\qquad\quad + \mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_{\bar\phi}, -(l+p_{12})_{\phi}\right)\mathcal{A}_6\left(-l_{\phi}, (l+p_{12})_{\bar\phi}, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+\right) \\
&\quad = 2\mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_\phi, -(l+p_{12})_{\bar\phi}\right)\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{12})_\phi, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+\right).
\end{aligned} \tag{4.4.5}$$

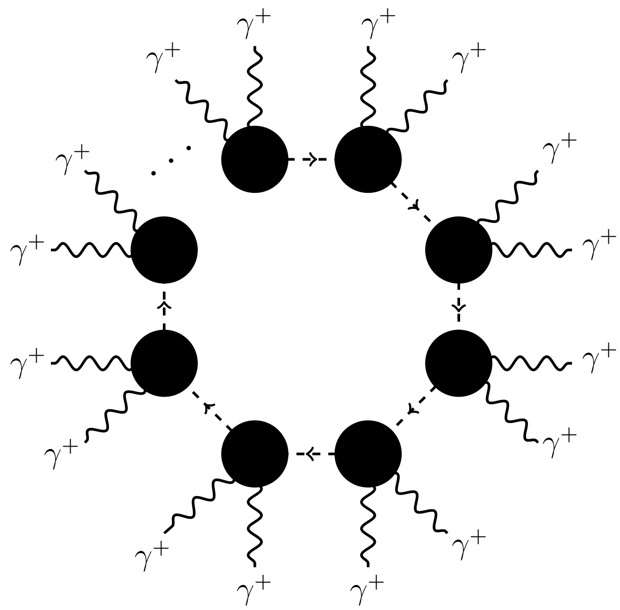
where the amplitudes are given in (4.3.49) and (4.3.53) and the form of the 6-point amplitude (4.3.53) makes it apparent that there are no local contributions to two-scalar cuts.

The factor of 2 in (4.4.5) is multiplied by $\frac{1}{8}$ (which compensates for the equivalent permutations in $\mathcal{P}\left(2,3,4,5,6\right)$ that are summed over). This matches the factor of $\frac{1}{4}$ in the integrand and hence verifies the rules of Section 4.4.1.

Integrating this using the formula (F.1.14) gives

$$\begin{aligned}
&\mathcal{A}_6^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+\right) \\
&\quad = \frac{1}{4}\left[\frac{i}{2880\pi^2}[12]^2[34]^2[56]^2\left(s_{12}^2+s_{34}^2+s_{56}^2+s_{12}s_{34}+s_{12}s_{56}+s_{34}s_{56}\right)\right. \\
&\qquad\qquad \left. + \mathcal{P}\left(2,3,4,5,6\right)\right] + \mathcal{O}(\epsilon).
\end{aligned} \tag{4.4.6}$$

The generalization to all multiplicity in the SD sector is now clear. There is always a single topologically distinct diagram with a corresponding scalar rational integral:




The complete integrand is then

$$\begin{aligned}
&\mathcal{I}_{2n}^{\mathrm{SD}}[l;\mu^2] \\
&= \left(\frac{1}{2}\right)^{n-1}\left([12]^2[34]^2\ldots[2n-1,2n]^2\frac{(-\mu^2)^n}{\prod_{i=1}^n\left[\left(l-\sum_{j=1}^{2i}p_j\right)^2+\mu^2\right]}+\mathcal{P}(2,3,\ldots,2n)\right) .
\end{aligned} \tag{4.4.7}$$

Using the result of equation (F.1.14), we find that the integrated amplitude is

$$\boxed{\begin{aligned}
&\mathcal{A}_{2n}^{\mathrm{BI}_4\ \text{1-loop}}\left(1_\gamma^+,2_\gamma^+,\ldots,2n_\gamma^+\right) \\
&\quad=\frac{i}{32\pi^2}\left(-\frac{1}{2}\right)^{n-1}\frac{1}{n(n+1)(n+2)(n+3)} \\
&\quad\times\left[[12]^2[34]^2\ldots[2n-1,2n]^2\left(\sum_{i<j}^n\sum_{k<l}^n a_{ijkl}\left(\sum_{m=2i+1}^{2j}p_m\right)^2\left(\sum_{m=2k+1}^{2l}p_m\right)^2\right)\right. \\
&\quad\left.+\mathcal{P}(2,3,\ldots,2n)\right]+\mathcal{O}(\epsilon)\,,
\end{aligned}} \tag{4.4.8}$$

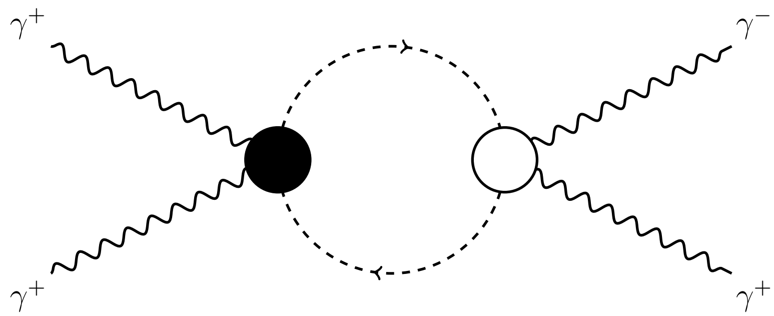
with

$$a_{ijkl}=\begin{cases}1 & \text{if all } i,j,k,l \text{ are different}\\ 2 & \text{if exactly 2 of } i,j,k,l \text{ are identical}\\ 4 & \text{if } i=k \text{ and } j=l\end{cases} . \tag{4.4.9}$$

It is straightforward to check that this result matches the results of the explicit calculations for the cases of $n = 2$ and $n = 3$, presented above.

### 4.4.3 Next-to-Self-Dual Sector

In the NSD sector the diagrams have a similar structure, consisting a single scalar loop decorated with the vertex factors. The novelty here is the appearance of a single negative helicity photon, and so each diagram contains either a single white or gray vertex. At 4-point there is only a single topologically distinct class of diagram, and contains both a black and white vertex[8]:




There are three non-trivial permutations of the external labels. Consider a single such permuation corresponding to momenta $p_1$ and $p_2$ flowing out of the black vertex, the corresponding integrand has the form

$$\mathcal{I}_4^{\text{NSD}}\left[l;\mu^2\right]\Big|_{12} = \frac{\mu^2[12]^2\langle 4|l|3]^2}{[l^2+\mu^2]\,[(l-p_{12})^2+\mu^2]}. \tag{4.4.10}$$

We now verify that the diagrammatic rules of Section 4.4.1 give an integrand with the right cuts in the NSD sector. There is only one distinct two-particle cut. As expected, the contribution to the integrand (4.4.10) on the $p_{12}$-cut is

$$\begin{aligned}\left[l^2+\mu^2\right]\left[(l-p_{12})^2+\mu^2\right]\mathcal{I}_4^{\text{SD}}[l;\mu^2]\big|_{p_{12}\text{-cut}} \\ &= \mathcal{A}_4\left(1_\gamma^+,2_\gamma^+,-l_\phi,(l-p_{12})_{\bar{\phi}}\right)\mathcal{A}_4\left(l_{\bar{\phi}},(p_{12}-l)_\phi,3_\gamma^+,4_\gamma^-\right) \\ &= \mu^2[12]^2\langle 4|l|3]^2,\end{aligned} \tag{4.4.11}$$

where the amplitudes are given in (4.3.49) and (4.3.50).

Unlike all of the integrals in the SD sector, this is a rational tensor integral. The explicit value of

[8]Note that there is no *tadpole* diagram with a single gray vertex since this contributes a scaleless integral which vanishes in dimensional regularization.

an integral of this form is in (F.2.6), this gives

$$\begin{aligned}\int \frac{\mathrm{d}^4 l}{(2\pi)^4} \int \frac{\mathrm{d}^{-2\epsilon}\mu}{(2\pi)^{-2\epsilon}} &\left[\frac{\mu^2[12]^2\langle 4|l|3]^2}{[l^2+\mu^2]\left[(l-p_{12})^2+\mu^2\right]}\right] \\ &= [12]^2 I_2^{d=4-2\epsilon}[\mu^2\langle 4|l|3]^2; p_{12}] \\ &= \frac{-i}{1920\pi^2}[12]^2\langle 4|\sigma_\mu|3]\langle 4|\sigma_\nu|3]\left[g^{\mu\nu}s_{12}^2 - 6p_{12}^\mu p_{12}^\nu s_{12}\right] + \mathcal{O}(\epsilon) \\ &= 0 + \mathcal{O}(\epsilon). \end{aligned} \tag{4.4.12}$$

Since the remaining channels are simple permutations of this one we conclude

$$\mathcal{A}_4^{\text{BI}_4\ \ \text{1-loop}}\left(1_\gamma^+, 2_\gamma^+, 3_\gamma^+, 4_\gamma^-\right) = 0 + \mathcal{O}(\epsilon). \tag{4.4.13}$$

Beginning at 6-point there are two distinct classes of diagrams, corresponding to diagrams containing a single white or gray vertex. Note that the 6-point integrand also has two distinct cuts. For instance, take the integrand on the $p_{56}$-cut,

$$\begin{aligned}&\left[l^2+\mu^2\right]\left[(l+p_{56})^2+\mu^2\right]\mathcal{I}_6^{\text{SD}}[l;\mu^2]\Big|_{p_{56}\text{-cut}} \\ &\quad= \mathcal{A}_4\left(5_\gamma^+, 6_\gamma^-, l_\phi, -(l+p_{56})_{\bar\phi}\right)\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{56})_\phi, 1_\gamma^+, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+\right) \\ &\qquad\quad + \mathcal{A}_4\left(5_\gamma^+, 6_\gamma^-, l_{\bar\phi}, -(l+p_{56})_\phi\right)\mathcal{A}_6\left(-l_\phi, (l+p_{56})_{\bar\phi}, 1_\gamma^+, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+\right) \\ &\quad= 2\mathcal{A}_4\left(5_\gamma^+, 6_\gamma^-, l_\phi, -(l+p_{56})_{\bar\phi}\right)\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{56})_\phi, 1_\gamma^+, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+\right).\end{aligned} \tag{4.4.14}$$

where the explicit forms of the amplitudes are given in (4.3.53) and (4.3.50). This generalises to any $p_{i6}$-cut, where $i \neq 6$.

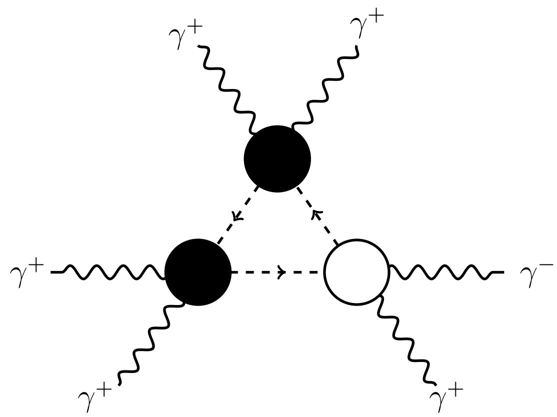
As a representative of the other class of cuts, consider the $p_{12}$-cut (which generalises to all $p_{ij}$-cuts where $i, j \neq 6$.),

$$\begin{aligned}&\left[l^2+\mu^2\right]\left[(l+p_{12})^2+\mu^2\right]\mathcal{I}_6^{\text{SD}}[l;\mu^2]\Big|_{p_{12}\text{-cut}} \\ &\quad= \mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_\phi, -(l+p_{12})_{\bar\phi}\right)\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{12})_\phi, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^-\right) \\ &\qquad\quad + \mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_{\bar\phi}, -(l+p_{12})_\phi\right)\mathcal{A}_6\left(-l_\phi, (l+p_{12})_{\bar\phi}, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^-\right) \\ &\quad= 2\mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_\phi, -(l+p_{12})_{\bar\phi}\right)\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{12})_\phi, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^-\right)\end{aligned} \tag{4.4.15}$$

where the amplitudes are given in (4.3.49) and (4.3.54). Note that there are two kinds of contributions to $\mathcal{A}_6\left(-l_{\bar\phi}, (l+p_{12})_\phi, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^-\right)$: one factorizes on an internal scalar and the other factorizes on an internal photon,

$$\mathcal{A}_6\left(-l_{\bar{\phi}},(l+p_{12})_\phi,3_\gamma^+,4_\gamma^+,5_\gamma^+,6_\gamma^-\right)=\mathcal{A}_6^{\text{scalar}}\left(-l_{\bar{\phi}},(l+p_{12})_\phi,3_\gamma^+,4_\gamma^+,5_\gamma^+,6_\gamma^-\right)$$
$$+\mathcal{A}_6^{\text{photon}}\left(-l_{\bar{\phi}},(l+p_{12})_\phi,3_\gamma^+,4_\gamma^+,5_\gamma^+,6_\gamma^-\right). \quad (4.4.16)$$

The first class of contributing diagrams is similar to the 4-point calculation and takes the form:




Summing over all permutations of the external labels gives the following contribution to the integrand

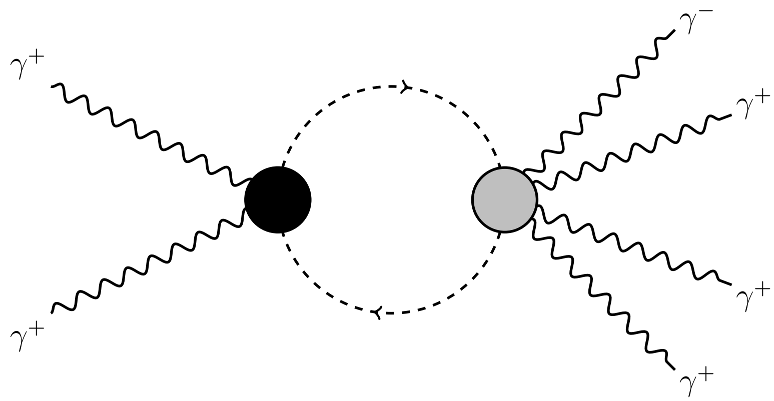
$$\mathcal{I}_6^{\text{NSD}}[l;\mu^2]\Big|_{\text{white}}=\frac{1}{4}\left[\frac{-(\mu^2)^2[12]^2[34]^2\langle 6|l|5]^2}{[l^2+\mu^2]\left[(l-p_{12})^2+\mu^2\right]\left[(l+p_{56})^2+\mu^2\right]}+\mathcal{P}\left(1,2,3,4,5\right)\right]. \quad (4.4.17)$$

This contribution has the correct $i6$-cuts (4.4.14). On a $p_{12}$-cut, (4.4.17) produces

$$\left[l^2+\mu^2\right]\left[(l+p_{12})^2+\mu^2\right]\mathcal{I}_6^{\text{SD}}[l;\mu^2]\Big|_{p_{ij}\text{-cut}}$$
$$=2\mathcal{A}_4\left(1_\gamma^+,2_\gamma^+,l_\phi,-(l+p_{12})_{\bar{\phi}}\right)\mathcal{A}_6^{\text{scalar}}\left(-l_{\bar{\phi}},(l+p_{12})_\phi,3_\gamma^+,4_\gamma^+,5_\gamma^+,6_\gamma^-\right). \quad (4.4.18)$$

The rest of the 6-point MHV amplitude is accounted for by the second class of diagrams.

The contributions from diagrams containing a single gray vertex:




which contributes the following to the integrand

$$\mathcal{I}_6^{\text{NSD}}[l;\mu^2]\Big|_{\text{gray}} = \frac{1}{2}\left[\frac{-(\mu^2)^2[12]^2[34]^2\langle 6|p_{12}|5]^2}{s_{125}\,[l^2+\mu^2]\,[(l-p_{12})^2+\mu^2]} + \mathcal{P}\left(1,2,3,4,5\right)\right]. \tag{4.4.19}$$

Here the $p_{12}$-cut yields

$$\begin{aligned}&\left[l^2+\mu^2\right]\left[(l+p_{12})^2+\mu^2\right]\mathcal{I}_6^{\text{SD}}[l;\mu^2]\Big|_{p_{12}\text{-cut}}\\ &\quad= 2\mathcal{A}_4\left(1_\gamma^+, 2_\gamma^+, l_\phi, -(l+p_{12})_{\bar{\phi}}\right)\mathcal{A}_6^{\text{photon}}\left(-l_{\bar{\phi}}, (l+p_{12})_\phi, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^-\right).\end{aligned} \tag{4.4.20}$$

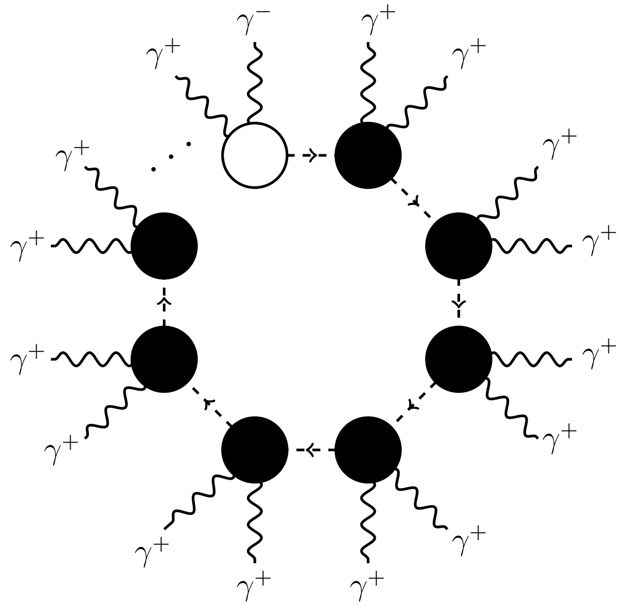
Thus the combined contributions to the integrand from both diagrams (4.4.17) and (4.4.19) is verified to have the correct cuts.

The integration of (4.4.17) and (4.4.19) can be carried out straightforwardly using the general results (F.1.14) and (F.2.6)

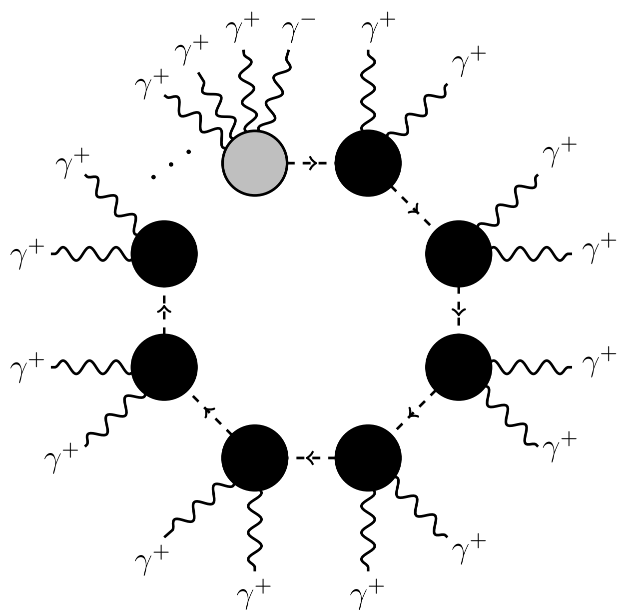
$$\begin{aligned}&\mathcal{A}_6^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^-\right)\\ &\quad= \frac{-i}{23040\pi^2}[12]^2[34]^2\langle 6|p_{125}|5]^2\left(s_{56}+3s_{12}+3s_{34}-6\frac{s_{12}^2}{s_{125}}\right) + \mathcal{P}\left(1,2,3,4,5\right) + \mathcal{O}(\epsilon).\end{aligned} \tag{4.4.21}$$

Unlike the cases we have seen so far, this expression is non-local. The factorization poles in the amplitude can be traced back to the non-local gray vertex factor and the associated set of gray loop diagrams. Calculating residues on these poles yields a 4-point SD amplitude times a Born-Infeld tree.

Finally we consider the all-multiplicity result in the NSD sector. Similar to the NSD 6-point example, there will be local contributions from diagrams containing a single white vertex:




as well as non-local contributions from diagrams containing a single gray vertex:




The explicit contributions to the integrand are, respectively

$$\mathcal{I}_{2n}^{\mathrm{NSD}}[l;\mu^2]\bigg|_{\text{white}} = -\left(-\frac{1}{2}\right)^{n-1}[12]^2\ldots[2n-3\;2n-2]^2[2n-1|l|2n\rangle^2$$
$$\times\frac{(\mu^2)^{n-1}}{\prod_{i=1}^{n}\left[\left(l-\sum_{j=1}^{2i}p_j\right)^2+\mu^2\right]}+\mathcal{P}(1,2,\ldots,2n-1), \tag{4.4.22}$$

and

$$\mathcal{I}_{2n}^{\text{NSD}}[l;\mu^2]\Big|_{\text{gray}}$$
$$= -\left(-\frac{1}{2}\right)^{n-1} \frac{[12]^2 \dots [2n-3\ 2n-2]^2 [2n-1|p_{2n}+p_{2n-2}+p_{2n-3}|2n\rangle^2}{s_{2n,2n-2,2n-3}}$$
$$\times \frac{(\mu^2)^{n-1}}{\prod_{i=1}^{n-2}\left[\left(l-\sum_{j=1}^{2i} p_j\right)^2+\mu^2\right]\left(l-\sum_{j=1}^{2n} p_j\right)^2} + \mathcal{P}(1,2,\dots,2n-1)\,. \quad (4.4.23)$$

Integrating these contributions separately using (F.1.14) and (F.2.6) gives the result

$$\boxed{\begin{aligned}&\mathcal{A}_{2n}^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+,2_\gamma^+,\dots,(2n-1)_\gamma^+,2n_\gamma^-\right) = \\ &\mathcal{A}_{2n}^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+,2_\gamma^+,\dots,(2n-1)_\gamma^+,2n_\gamma^-\right)\Big|_{\text{white}} + \mathcal{A}_{2n}^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+,2_\gamma^+,\dots,(2n-1)_\gamma^+,2n_\gamma^-\right)\Big|_{\text{gray}},\end{aligned}}$$
(4.4.24)

where

$$\boxed{\begin{aligned}&\mathcal{A}_{2n}^{\text{BI}_4\ \text{1-loop}}\left(1_\gamma^+,2_\gamma^+,\dots,(2n-1)_\gamma^+,2n_\gamma^-\right)\Big|_{\text{white}} \\ &\quad= \frac{-i}{16\pi^2}\left(-\frac{1}{2}\right)^{n-1}\frac{1}{(n-1)n(n+1)(n+2)(n+3)}[12]^2\dots[2n-3\ 2n-2]^2 \\ &\quad\times \sum_{i<j}^{n}\left(\sum_{m=2i+1}^{2j} p_m\right)^2\left[\sum_{k<l}^{n} 2\, a_{ijkl}\left(\sum_{m=1}^{2k}[2n-1|p_m|2n\rangle\right)\left(\sum_{m=1}^{2l}[2n-1|p_m|2n\rangle\right)\right. \\ &\quad\quad\left.+\sum_{k=1}^{n} b_{ijk}\left(\sum_{m=1}^{2k}[2n-1|p_m|2n\rangle\right)^2\right] + \mathcal{P}(1,2,\dots,2n-1) + \mathcal{O}(\epsilon),\end{aligned}}$$
(4.4.25)

with

$$b_{ijk} = \begin{cases} 2 & \text{if } i\neq k \text{ and } j \neq k \\ 6 & \text{if } i=k \text{ or } j=k \end{cases}. \quad (4.4.26)$$

And also

$$\boxed{\begin{aligned}&\mathcal{A}_{2n}^{\text{BI}_4\text{ 1-loop}}\left(1_\gamma^+,2_\gamma^+,\dots,(2n-1)_\gamma^+,2n_\gamma^-\right)\Big|_{\text{gray}}\\&=\frac{i}{32\pi^2}\frac{(n-2)!}{(n+2)!}\left(-\frac{1}{2}\right)^{n-1}\frac{[12]^2\dots[2n-3\;2n-2]^2[2n-1|p_{2n-2}+p_{2n-3}|2n\rangle^2}{s_{2n,2n-2,2n-3}}\\&\times\left[\sum_{i<j}^{n-2}\sum_{k<l}^{n-2}a_{ijkl}\left(\sum_{m=2i+1}^{2j}p_m\right)^2\left(\sum_{m=2k+1}^{2l}p_m\right)^2+4\sum_{i\le j}^{n-2}\left(\sum_{m=1}^{2i}p_m\right)^2\left(\sum_{m=1}^{2j}p_m\right)^2\right.\\&\left.+2\sum_{i=1}^{n-2}\sum_{k<l}^{n-2}a_{i(n-1)kl}\left(\sum_{m=1}^{2i}p_m\right)^2\left(\sum_{m=2k+1}^{2l}p_m\right)^2\right]+\mathcal{P}(1,2,\dots,2n-1)+\mathcal{O}(\epsilon).\end{aligned}} \tag{4.4.27}$$

It is easy to check that these generic result match the cases of $n=2$ and $n=3$ that were presented above.

As we have already discussed for the 6-particle case, the NSD $(2n)$-particle amplitudes we calculate have poles that can be traced back to the associated poles of the gray vertex factors for $n\ge 3$. These poles are located at $s_{i,j,2n}=0$, for $i<j\le 2n-1$, and the associated residues are products of the tree 4-particle amplitude and a SD $(2n-2)$-particle amplitude of the form (4.4.8). Let us now demonstrate this factorization explicitly. Consider for example the residue of (4.4.24) at $s_{2n-2,2n-1,2n}=0$,

$$\begin{aligned}&\operatorname*{Res}_{p_f^2=0}\mathcal{A}_{2n}^{\text{BI}_4\text{ 1-loop}}\left(1_\gamma^+,2_\gamma^+,\dots,(2n-1)_\gamma^+,2n_\gamma^-\right)\\&=2\frac{1}{32\pi^2}\frac{(n-2)!}{(n+2)!}\left(-\frac{1}{2}\right)^{n-1}[12]^2\dots[2n-5\;2n-4]^2[2n-2\;2n-1]^2[2n-3|p_f|2n\rangle^2\\&\times\left[\sum_{i<j}^{n-2}\sum_{k<l}^{n-2}a_{ijkl}\left(\sum_{m=2i+1}^{2j}p_m\right)^2\left(\sum_{m=2k+1}^{2l}p_m\right)^2+4\sum_{i\le j}^{n-2}\left(\sum_{m=1}^{2i}p_m\right)^2\left(\sum_{m=1}^{2j}p_m\right)^2\right.\\&\left.+2\sum_{i=1}^{n-2}\sum_{k<l}^{n-2}a_{i(n-1)kl}\left(\sum_{m=1}^{2i}p_m\right)^2\left(\sum_{m=2k+1}^{2l}p_m\right)^2\right]+\mathcal{P}(1,2,\dots,2n-3)+\mathcal{O}(\epsilon)\,,\end{aligned} \tag{4.4.28}$$

where $p_f=p_{2n-2}+p_{2n-1}+p_{2n}$ is the momentum on the factorization channel. Notice that not all permutations listed in (4.4.27) contribute to the residue while the additional factor of 2 in the right-hand side comes from the trivial permutation $2n-2\leftrightarrow 2n-1$. Now, on the factorization channel

$$[2n-3|p_f|2n\rangle=-[2n-3,p_f]\langle p_f,2n\rangle=-i[2n-3,p_f]\langle -p_f,2n\rangle\,. \tag{4.4.29}$$

Also, we can use momentum conservation to write

$$\sum_{m=1}^{2i} p_m = -p_f - \sum_{m=2i+1}^{2n-3} p_m = -\sum_{m=2i+1}^{2n-2} \tilde{p}_m \,, \tag{4.4.30}$$

where we have defined

$$\tilde{p}_m = \begin{cases} p_m & \text{if } m \leq 2n-3 \\ p_f & \text{if } m = 2n-2 \end{cases} \tag{4.4.31}$$

With this definition we can write the above residue as

$$\begin{aligned}
&\operatorname*{Res}_{p_f^2=0} \mathcal{A}_{2n}^{\mathrm{BI}_4 \text{ 1-loop}} \left(1_\gamma^+, 2_\gamma^+, \ldots, (2n-1)_\gamma^+, 2n_\gamma^-\right) = \left([2n-2, 2n-1]^2 \langle -p_f, 2n\rangle^2\right) \\
&\times \Bigg[ \frac{1}{32\pi^2} \frac{(n-2)!}{(n+2)!} \left(-\frac{1}{2}\right)^{n-2} [12]^2 \ldots [2n-5, 2n-4]^2 [2n-3, p_f]^2 \\
&\qquad \times \sum_{i<j}^{n-1} \sum_{k<l}^{n-1} a_{ijkl} \left(\sum_{m=2i+1}^{2j} \tilde{p}_m\right)^2 \left(\sum_{m=2k+1}^{2l} \tilde{p}_m\right)^2 + \mathcal{P}(1, 2, \ldots, 2n-3) + \mathcal{O}(\epsilon) \Bigg] \,,
\end{aligned} \tag{4.4.32}$$

which clearly shows its factorized form. More precisely, we can write

$$\begin{aligned}
&\operatorname*{Res}_{p_f^2=0} \mathcal{A}_{2n}^{\mathrm{BI}_4 \text{ 1-loop}} \left(1_\gamma^+, \ldots, (2n-1)_\gamma^+, 2n_\gamma^-\right) \\
&= \mathcal{A}_{2n-2}^{\mathrm{BI}_4 \text{ 1-loop}} \left(1_\gamma^+, \ldots, (2n-3)_\gamma^+, (p_f)_\gamma^+\right) \times \mathcal{A}_4^{\mathrm{BI}_4} \left((-p_f)_\gamma^-, (2n-2)_\gamma^+, (2n-1)_\gamma^+, (2n)_\gamma^-\right) .
\end{aligned} \tag{4.4.33}$$

The fact that the pole terms of the NSD 1-loop amplitude factorize to a SD 1-loop and a tree-level MHV amplitude at all multiplicities means that if we choose to remove the SD amplitudes by introducing finite local counter-terms, then the NSD amplitudes become local and can also be set to zero with the introduction of further finite local counter-terms. The consequences will be discussed in the next section.

## 4.5 Quantum Electromagnetic Duality

The main results of this chapter are (4.4.8) and (4.4.24), explicit expressions for the SD and NSD amplitudes at one-loop, that would have been impossible to obtain by using traditional Feynman diagrammatics. As expected, they are finite and at $\mathcal{O}(\epsilon^0)$ given by rational functions. For the SD and NSD sectors, these properties follow from the property of $\mathrm{BI}_4$ being a consistent truncation of a supersymmetric model at tree-level. More generally however, we expect both of these properties to obtain in all helicity sectors *except* the duality-conserving sector

$$\mathcal{A}_{2n}^{\mathrm{BI}_4}\left(1_\gamma^+,\ldots,n_\gamma^+,(n+1)_\gamma^-,\ldots,(2n)_\gamma^-\right). \tag{4.5.1}$$

As a consequence of an electromagnetic duality symmetry, these amplitudes which conserve a chiral charge for the photon are the only non-vanishing amplitudes at tree-level [155, 156]. At one-loop, only amplitudes in the duality-conserving sector can have non-vanishing 4d cuts and consequently non-rational functional dependence.

The methods of this chapter do not directly extend to calculations at one-loop beyond the SD and NSD sectors. In a sense then we have explored only a small fraction of the structure of Born-Infeld at one-loop. At higher multiplicity the majority of non-duality-conserving sectors, which are expected to be rational, cannot be calculated by constructing integrands from massive scalar cuts. In the duality-conserving sector, the cut-constructible parts can be obtained using the non-vanishing 4d cuts.

It is important to note that the explicit results (4.4.8) and (4.4.24) were obtained in a particular version of dimensional regularization. Specifically, we imposed that the tree-amplitudes appearing in $d$-dimensional cuts should satisfy the low-energy theorem described in Section 4.3.2. Physically this is equivalent to requiring that the low-energy consequences of T-duality are preserved by the dimensional regulator. While this choice of regularization scheme greatly simplifies the analysis, physical observables must be independent of this choice. It would be an interesting and useful consistency check to re-calculate the one-loop amplitudes in the SD and NSD sectors (and beyond) using a different regularization scheme. For example, one might consider an alternate non-supersymmetric dimensional scheme (such as the 't Hooft-Veltman scheme) or more ambitiously a *non-dimensional* scheme such as Passarino-Veltman. Though potentially much more complicated, the latter case has the virtue of being defined intrinsically in $d=4$ and may therefore avoid explicitly breaking the electromagnetic duality symmetry. Such questions are important, but are outside of the scope of this chapter.

Having explicit forms for two infinite classes of duality-violating one-loop amplitudes, we are in a position to make an interesting observation about the fate of electromagnetic duality at the one-loop quantum-level (see [157] for recent discussion). Recall that electromagnetic duality is not a symmetry in the usual sense. In the standard covariant approach to perturbative quantization, the (effective) quantum theory of Born-Infeld electrodynamics is defined by a path integral

$$e^{i\Gamma[J]}=\int[DA]\,e^{iS[A]+i\int A\wedge J}, \tag{4.5.2}$$

where $S$ is the manifestly Lorentz-invariant effective action (4.1.1), and the path integral measure includes appropriate gauge fixing terms. Curiously, electromagnetic duality is a symmetry only on the *saddle points* of this integral. This is equivalent to the familiar statement that electromagnetic duality is a symmetry of the classical equations of motion, *but not* a symmetry of the off-shell effective action [158]. The practical consequences of this observation is that the standard Feynman rules (for example in $R_\xi$-gauge) derived from the action (4.1.1) do not manifest the conservation of duality charge at each vertex. At tree-level, duality violating scattering amplitudes are seen to vanish only after summing over all relevant Feynman graphs.

One approach to realizing electromagnetic duality off-shell was given by Deser and Teitelboim who proposed a modified transformation of the covariant action (4.1.1) which acts *non-locally* on the gauge potential [159, 160]. The consequences of such non-local symmetries on perturbative scattering amplitudes are unclear. Another road is to maintain the standard local form of the duality transformation, but replace (4.1.1) with a classically equivalent non-covariant action. This was the approach of Schwarz and Sen [161, 162] and also the first-order (or phase space path integral) approach of Deser and Teitelboim [159, 160]. While such non-covariant actions do indeed manifest the conservation of duality charge vertex-by-vertex in the Feynman diagram expansion, we have simply traded one hard problem for another since now it is not clear that the loop-level scattering amplitudes we calculate are Lorentz invariant (see [163] for related discussion). In summary, it would appear to not be possible to define electromagnetic duality as a local, off-shell symmetry of the action while preserving manifest Lorentz covariance. In retrospect we should not expect such a thing to be possible, if it were then the standard Noether procedure would allow us to construct a local Lorentz covariant current operator for the duality charge. But since this charge is carried by massless spin-1 states, such an object is forbidden by the Weinberg-Witten theorem [74]. Given this state of affairs, it is unclear if it is possible to define a quantization of Born-Infeld electrodynamics that preserves electromagnetic duality *in addition to* the standard properties of Lorentz invariance, locality and unitarity. In the specific context considered in this chapter we would like to know if it is possible to define an S-matrix at loop-level which respects the helicity selection rules associated with the conservation of duality charge.

The approach we took, constructing local loop integrands consistent with $d$-dimensional unitarity and 4-dimensional Lorentz invariance, would appear to preserve all of the expected properties manifestly, with the exception of duality invariance. Determining if our explicit results are consistent with the existence of such a duality-respecting quantization is a little subtle. It is too naive to simply observe that the duality-violating one-loop amplitudes (4.4.8) and (4.4.24) are non-zero. Similar to $U(1)$ symmetries acting on chiral fermions, duality rotations act as chiral rotations on states of spin-1, and are therefore only defined in exactly 4-dimensions. Our explicit results however were obtained in a dimensional regularization scheme which explicitly breaks the symmetry. To determine if a genuine anomaly is present, we must first recall that the classical action used to define the full quantum theory as a path integral (4.5.2) is ambiguous up to the addition of *finite local counterterms*. If a consistent set of local, Lorentz-invariant counterterms can be added to the action such that their contribution *cancels* the explicitly calculated rational one-loop amplitudes, then there is no anomaly and the symmetry is preserved. In a related context, recent explicit calculations in $\mathcal{N} = 4$ supergravity in $d = 4$ have revealed that the conventional understanding of the physical consequences of chiral anomalies may be modified in the context of duality symmetries [164, 165].

For the one-loop Born-Infeld amplitudes considered in this chapter, in the SD sector the expressions (4.4.8) are manifestly local and Lorentz-invariant, and so can be consistently cancelled by local counterterms. In the NSD sector the expressions (4.4.24) are non-local, here we must sum over both contact contributions from independent local operators and factoring contributions containing both counterterms and tree-level Born-Infeld vertices. The condition that these non-local contributions can be removed with finite local counterterms requires that our explicit results (4.4.24) have the singularity and factorization properties of tree-amplitudes, and we verified this explicitly at the end of Section 4.4.3.

These results give an infinite number of non-trivial checks on the preservation of duality under quantization, but do not constitute a proof. Extending the results of this chapter to the remaining duality-violating sectors and beyond is therefore essential to understanding the ultimate fate of electromagnetic duality in quantum Born-Infeld.

# CHAPTER 5

# The Black Hole Weak Gravity Conjecture with Multiple Charges

## 5.1 Multi-Charge Generalization of the Weak Gravity Conjecture

The purpose of this chapter is to generalize the discussion in Section 1.2.2 to the universality class of models for which the low-energy matter spectrum consists of $N$ $U(1)$ gauge fields. We consider black hole solutions with general electric and magnetic charges.

The two-derivative approximation to the EFT has many accidental symmetries, including an $O(N)$ global flavor symmetry, parity and $U(N)$ electromagnetic duality symmetry. We do not assume that any of these symmetries are preserved in the UV, and instead analyze the most general possible EFT with the assumed low-energy spectrum

$$\begin{aligned} S = \int \mathrm{d}^4x\sqrt{-g}\Big[&\frac{M_{\mathrm{Pl}}^2}{4}R - \frac{1}{4}F^i_{\mu\nu}F^{i\,\mu\nu} + a_{ijk}F^i_{\mu\nu}F^{j\nu\rho}F^{k\mu}_{\rho} + b_{ijk}F^i_{\mu\nu}F^{j\nu\rho}\tilde{F}^{k\mu}_{\rho} \\ &+ \alpha_{ijkl}\,F^i_{\mu\nu}F^{j\,\mu\nu}F^k_{\rho\sigma}F^{l\,\rho\sigma} + \beta_{ijkl}\,F^i_{\mu\nu}\tilde{F}^{j\,\mu\nu}F^k_{\rho\sigma}\tilde{F}^{l\,\rho\sigma} \\ &+ \gamma_{ij}\,F^i_{\mu\nu}F^j_{\sigma\rho}W^{\mu\nu\sigma\rho} + \chi_{ijkl}\,\tilde{F}^i_{\mu\nu}F^{j\,\mu\nu}F^k_{\rho\sigma}F^{l\,\rho\sigma} + \omega_{ij}\,F^i_{\mu\nu}\tilde{F}^j_{\sigma\rho}W^{\mu\nu\sigma\rho}\Big]. \end{aligned} \tag{5.1.1}$$

In [166] it was shown that the kinematic condition for a large extremal black hole with multiple charges to decay is a non-trivial generalization of the single charge version of the WGC. In general, if a set of *light* states $|\vec{q}_i, m_i\rangle$ are available with masses $m_i$ and charge vectors $\vec{q}_i$, then the possible charge-to-mass ratio vectors of the associated multi-particle states $|\vec{q}_1, m_1\rangle^{\otimes N_1} \otimes |\vec{q}_2, m_2\rangle^{\otimes N_2} \otimes ...$ are given by

$$\vec{z} \in \left\{ \frac{M_{\mathrm{Pl}}\sum_i N_i\vec{q}_i}{\sum_i |N_i| m_i}, \quad N_i \in \mathbb{Z} \right\}. \tag{5.1.2}$$

Here $N_i < 0$ corresponds to contributions from CP conjugate states. This set describes the convex hull of the charge-to-mass vectors $\vec{z}_i = \vec{q}_i/m_i$. The condition that the decay of asymptotically large extremal black holes be allowed is given by the *convex hull condition* [166]:

**Weak Gravity Conjecture (Multiple Charges):** *In a UV complete model of quantum gravity, the convex hull of the set of charge-to-mass vectors*

$$\vec{z}_i \equiv \frac{M_{\text{Pl}}}{m_i}\begin{pmatrix}\vec{q}_i \\ \vec{p}_i\end{pmatrix}, \tag{5.1.3}$$

*for every charged state in the spectrum, with mass $m$, electric charges $\vec{q} = (q^1, q^2...)$ and magnetic charges $\vec{p} = (p^1, p^2, ...)$, must enclose the unit ball $|\vec{z}|^2 \leq 1$.*

As in the single charge case, to show that a given model does not satisfy this condition requires complete knowledge of the spectrum of charged states. It is however possible to show that this condition is satisfied with only partial knowledge of the spectrum since the convex hull of a subset of vectors always forms a subregion of the full convex hull. This condition has been previously analyzed from several perspectives [167], considering contributions from the *particle regime*. In this chapter, we will describe the general conditions on the Wilson coefficients $\{a_{ijk}, b_{ijk}, \alpha_{ijkl}, \beta_{ijkl}, \gamma_{ij}, \chi_{ijkl}, \omega_{ij}\}$ under which the convex hull condition is satisfied by contributions from the *black hole regime*.

## 5.2 Extremality Shift

In this section we will determine the effect of higher-derivative operators on the extremality bound using the method developed in [46]. In the case of multiple charges, this amounts to delineating the space of allowed charge combinations $Q = \sqrt{q_1^2 + p_1^2 + ...}$ for a given mass $m$. We use the presence of a naked singularity, or absence of an event horizon, to rule out charge configurations at a given mass; such combinations of charge and mass will be called *superextremal*.

In pure Einstein-Maxwell theory, the superextremal black holes have $Q/m > 1$. We refer to such an inequality as the *extremality bound*. This requirement derives from the positivity of the discriminant of the function $1/g_{rr}$, which itself comes from the requirement that that function should have a zero (i.e. the event horizon). We will see that the higher-derivative corrections have the effect of shifting the right-hand side of this bound by factors proportional to the Wilson coefficients and suppressed by factors of $1/Q$. Generically, $n$-derivative operators will contribute a term in the extremality bound that is proportional to $1/Q^{n-2}$.

This approach is necessarily first-order in the EFT coefficients; if we were to compute the shift to second-order in the four-derivative coefficients, we would need also to consider the first-order effect of six-derivative operators, as these contribute at the same order in $1/Q$. This means that at each step we eliminate all terms that are beyond leading-order in the four-derivative coefficients.

### 5.2.1 No Correction from Three-Derivative Operators

When $N \geq 3$ the leading effective interactions are given by three-derivative operators:

$$S_3 = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{M_{\mathrm{Pl}}^2}{4}R - \frac{1}{4}F^i_{\mu\nu}F^{i\,\mu\nu} + a_{ijk}F^i_{\mu\nu}F^{j\nu\rho}F^{k\mu}_{\rho} + b_{ijk}F^i_{\mu\nu}F^{j\nu\rho}\tilde{F}^{k\mu}_{\rho}\right], \tag{5.2.1}$$

where the dual field strength tensor is defined as

$$\tilde{F}^{i\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F^i_{\rho\sigma}\,. \tag{5.2.2}$$

From the index structure of the three-derivative operators (alternatively from the structure of the corresponding local matrix elements given in Appendix G) one can show that both $a_{ijk}$ and $b_{ijk}$ are totally antisymmetric.

We analyze solutions to the equations of motion:

$$\begin{aligned}
\nabla_\mu F^{i\mu\nu} &= -6a_{ijk}\nabla_\mu\left(F^{j\nu\rho}F^{k\mu}_{\rho}\right) - 6b_{ijk}\nabla_\mu\left(F^{j\nu}_{\alpha}\tilde{F}^{k\mu\alpha}\right), \\
R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} &= \frac{2}{M_{\mathrm{Pl}}^2}\left[F^i_{\mu\rho}F^{i\rho}_{\nu} - \frac{1}{4}g_{\mu\nu}F^i_{\rho\sigma}F^{i\rho\sigma}\right. \\
&\quad \left. + 2\,a_{ijk}\left[F^i_{\alpha\mu}F^{j\rho}_{\nu}F^{k\alpha}_{\rho} - \frac{1}{2}g_{\mu\nu}F^i_{\rho\sigma}F^{j\sigma\alpha}F^{k\rho}_{\alpha}\right] + 2b_{ijk}F^i_{\mu\rho}F^j_{\nu\sigma}\tilde{F}^{k\rho\sigma}\right].
\end{aligned} \tag{5.2.3}$$

By an elementary spurion analysis it is clear that there can be no modification of the extremality bound at $\mathcal{O}(a,b)$. Promoting $a_{ijk}$ and $b_{ijk}$ to background fields transforming as totally anti-symmetric tensors of the (explicitly broken) flavor symmetry group $O(N)$, at leading order the extremality shift can depend only on invariants of the form $a_{ijk}q^iq^jq^k$ or $a_{ijk}q^iq^jp^k$, which vanish. At next-to-leading order there could be contributions of the form $a_{ijk}a_{klm}q^ip^jq^lp^m$, which do not obviously vanish for similarly trivial reasons. If present such contributions would appear at the same order, $\mathcal{O}\left(1/Q^2\right)$ as the leading-order contributions from the four-derivative operators.

Interestingly these $\mathcal{O}(a^2, ab, b^2)$ corrections also vanish. To show this, we evaluate the right-hand-side of (5.2.3) on a spherically symmetric ansatz,

$$ds^2 = g_{tt}(r)\,dt^2 + g_{rr}(r)\,dr^2 + r^2d\Omega^2, \qquad F^{i\,tr}(r), \qquad F^{i\,\theta\phi}(r), \tag{5.2.4}$$

with the remaining components of the field strength tensors set to zero. The higher-derivative terms are seen to vanish due to the structure of the index contractions. The equations of motion for the non-zero components $g_{tt},\ g_{rr},\ F^{itr},\ F^{i\theta\phi}$ are *identical* to the equations of motion of two-derivative Einstein-Maxwell. The Reissner–Nordström black hole remains the unique spherically symmetric solution to the higher-derivative equations of motion with a given charge and mass.

It is interesting to note that the above argument fails if the solution is only axisymmetric, as in the general Kerr-Newman solution. For spinning, dyonic black holes, the three-derivative operators might give $\mathcal{O}\left(1/Q^2\right)$ corrections to the extremality bounds. We leave the analysis of this case to future work.

### 5.2.2 Four-Derivative Operators

The three-derivative operators have no contribution on spherically symmetric backgrounds. Thus, the leading shift to the extremality bound comes from four-derivative operators. We consider the action

$$\begin{aligned} S_4 = \int \mathrm{d}^4x\sqrt{-g}\Big(\frac{R}{4} - \frac{1}{4}F^i_{\mu\nu}F^{i\,\mu\nu} + \alpha_{ijkl}\,F^i_{\mu\nu}F^{j\,\mu\nu}F^k_{\rho\sigma}F^{l\,\rho\sigma} + \beta_{ijkl}\,F^i_{\mu\nu}\tilde{F}^{j\,\mu\nu}F^k_{\rho\sigma}\tilde{F}^{l\,\rho\sigma} \\ + \gamma_{ij}\,F^i_{\mu\nu}F^j_{\sigma\rho}W^{\mu\nu\sigma\rho} + \chi_{ijkl}\,\tilde{F}^i_{\mu\nu}F^{j\,\mu\nu}F^k_{\rho\sigma}F^{l\,\rho\sigma} + \omega_{ij}\,F^i_{\mu\nu}\tilde{F}^j_{\sigma\rho}W^{\mu\nu\sigma\rho}\Big). \end{aligned} \tag{5.2.5}$$

Here the Latin indices run from $1$ to the number of gauge fields $N$. This is the most general possible set of four-derivative operators for Einstein-Maxwell theory in 4 dimensions. For a thorough discussion on how these operators comprise a complete basis, see Appendix G. We will see that the parity-odd operators can contribute if we allow for magnetic charges. Our calculation is identical to the one performed in [46] if we set $N \to 1$ and turn on only electric charges. We have chosen units with $M_{\mathrm{Pl}} = 1$ for convenience, though they may be restored via dimensional analysis.

#### 5.2.2.1 Background

First consider the uncorrected theory, which is gravity with $N$ $U(1)$ gauge fields. This theory admits solutions that are black holes with up to $N$ electric and magnetic charges. These solutions take the form:

$$\begin{aligned} ds^2 &= g_{tt}\,dt^2 + g_{rr}\,dr^2 + r^2 d\Omega^2, \qquad F^{i\,tr} = \frac{q^i}{r^2}, \qquad F^{i\,\theta\phi} = \frac{p^i}{r^4\,\sin\theta}\,, \\ -\,g_{tt} &= g^{rr} = 1 - \frac{2M}{r} + \frac{Q^2}{r^2}\,. \end{aligned} \tag{5.2.6}$$

Here $Q^2 = q^i q^i + p^i p^i$. These backgrounds are spherically symmetric, so we will impose this as a requirement on the shifted background[1]. In the case of spherical symmetry, one may rearrange the Einstein equation and integrate to find [46]

$$g^{rr} = 1 - \frac{2M}{r} - \frac{2}{r}\int_r^\infty dr r^2 T_t^{\ t} \, . \tag{5.2.7}$$

For the uncorrected theory, the stress tensor is

$$T_{\mu\nu} = F^i_{\mu\alpha} F^{i\alpha}_\nu - \frac{1}{4} F^i_{\alpha\beta} F^{i\alpha\beta} g_{\mu\nu} \, . \tag{5.2.8}$$

In this case, it is easy to see that the effect of the stress tensor is to add the $\frac{q^2+p^2}{r^2}$ term to $g^{rr}$.

#### 5.2.2.2 Corrections to the Background

Now consider the effect of the four-derivative terms. To compute their effect on the geometry, we must compute their contributions to the stress tensor. We will expand the stress tensor as a power series in the Wilson coefficients as

$$T = T^{(0)} + T^{(1)}_{Max} + T^{(1)}_{Lag} + \ldots \tag{5.2.9}$$

Here we have written two terms that are proportional to the first power of the Wilson coefficients $(\alpha_{ijkl}, \beta_{ijkl}, \ldots)$, because there are two different sources of first-order corrections.

The first change $T^{(1)}_{Max}$ comes from the effect of these operators on solutions to the Maxwell equations, which changes the values of $F^i_{\mu\alpha} F^{i\alpha}_\nu - \frac{1}{4} F^i_{\alpha\beta} F^{i\alpha\beta} g_{\mu\nu}$. Thus, $T^{(1)}_{Max}$ essentially comes from evaluating the zeroth-order stress tensor on the first-order solution of the $F^i$ equations of motion.

The second change $T^{(1)}_{Lag}$ derives from varying the higher-derivative operators with respect to the metric. Thus, this term is essentially the first-order stress tensor, and we will evaluate it on the zeroth-order solutions to the Einstein and Maxwell equations. The remainder of this section will be devoted to computing each of these contributions.

[1] Spherical symmetry ensures that $1/g_{rr} = g^{rr}$, even for the corrected solutions. However, $g_{tt}$ and $1/g_{rr}$ will generally receive different corrections, which is why we do not denote these functions with one symbol such as $f(r)$.

#### 5.2.2.3 Maxwell Corrections

The first source of corrections to the stress tensor derives from including the corrections to the value of $F$. The corrected gauge field equation of motion is

$$\nabla_\mu F^{i\mu\nu} = \nabla_\mu \Big( 8\,\alpha_{ijkl} F^{j\mu\nu} F^k_{\alpha\beta} F^{l\alpha\beta} + 8\,\beta_{ijkl} \tilde{F}^{j\mu\nu} F^k_{\alpha\beta} \tilde{F}^{l\alpha\beta} + 4\,\gamma_{ij} F^j_{\alpha\beta} W^{\mu\nu\alpha\beta} \\ + 4\,\Big( \chi_{ijkl} \tilde{F}^{j\mu\nu} F^k_{\alpha\beta} F^{l\alpha\beta} + \chi_{klij} F^{j\mu\nu} \tilde{F}^k_{\alpha\beta} F^{l\alpha\beta} \Big) + 4\,\omega_{ij} \tilde{F}^j_{\alpha\beta} W^{\mu\nu\alpha\beta} \Big). \tag{5.2.10}$$

We denote the right-hand side of this equation by $\nabla_\mu G^{\mu\nu}$. The first-order solution to the Maxwell equation leads to corrections that equal (see Appendix H)

$$(T^{(1)}_{Max})_t{}^t \;=\; -\left[\sqrt{-g} G^{itr}\right]^{(1)} \left[\sqrt{-g} F^{itr}\right]^{(0)} / (g_{\theta\theta} g_{\phi\phi})\,. \tag{5.2.11}$$

By plugging in the zeroth-order values of the fields into this expression, we compute the corrections to the stress tensor through the Maxwell equation:

$$(T^{(1)}_{Max})_t{}^t \;=\; \frac{8}{r^8} \Big( 2\alpha_{ijkl}\, q^i q^j (q^k q^l - p^k p^l) + 4\,\beta_{ijkl}\, q^i p^j q^k p^l + 2\gamma_{ij}\, q^i q^j \,(Q^2 - Mr) \\ + \chi_{ijkl}\, \left( q^i p^j (q^k q^l - p^k p^l) + 2 q^i q^j q^k p^l \right) +\, 2\omega_{ij}\, q^i p^j \,(Q^2 - Mr) \Big)\,. \tag{5.2.12}$$

The details of this derivation may be found in Appendix H, but we should comment on a few interesting points. First, note the only $G^{itr}$ arises in the result. This is due to the Bianchi identity, which does not allow $G^{i\theta\phi}$ to contribute. The Bianchi identity requires that $\partial_r F_{\theta\phi} = 0$, so in fact $F^i_{\theta\phi}$ can get no corrections at any order.

A subtlety arises from the fact that the metric appears in the expression for the stress tensor. Therefore, it might appear that the first-order corrections to $T_t{}^t$ involve contributions from the first-order value of $F$ and the first-order value of $g$. This would be problematic because the first-order value of $g$ is what we use the stress tensor to compute in the first place. In fact, this is not an issue; only the zeroth-order metric shows up in (5.2.11). This decoupling relies on cancellation between various factors of metric components, as well as spherical symmetry. Without this, the perturbative procedure we use to compute the shift to the metric would not work. We do not expect this decoupling between corrections to the stress tensor and corrections to the metric to happen for general backgrounds. It would be interesting to study the general circumstances under which it occurs.

#### 5.2.2.4 Lagrangian Corrections

The second source of corrections is comparatively straightforward and comes from considering the higher-derivative terms in the Lagrangian as "matter" and varying them with respect to the metric. The variations of each term are given in appendix I. The result is

$$\begin{aligned}(T^{(1)}_{Lag})_t{}^t \quad = \quad & \frac{1}{r^8}\Big(4\,\alpha_{ijkl}\,(p^ip^jp^kp^l + 2q^iq^jp^kp^l - 3q^iq^jq^kq^l) - 4\,\beta_{ijkl}\,q^ip^jq^kp^l \\ & - \frac{4}{3}\,\gamma_{ij}\,\big(q^iq^j(6Q^2 - 2Mr - 3r^2) + p^ip^j(6Q^2 - 10Mr - 3r^2)\big) \\ & - 16\,\chi_{ijkl}\,q^ip^jq^kq^l - \frac{8}{3}\,\omega_{ij}q^ip^j(4Mr - 3r^2)\Big)\,.\end{aligned} \tag{5.2.13}$$

In both cases, we have simplified the expressions by using the symmetries of the tensor appearing in the higher-derivative terms (e.g. $\alpha_{ijkl} = \alpha_{jikl} = \alpha_{klij}$).

### 5.2.3 Leading Shift to Extremality Bound

By adding together both sources of corrections and computing the integral in (5.2.7), we compute the shift to the radial function $g^{rr}$ defined as,

$$g^{rr} = 1 - \frac{2M}{r} + \frac{q^2 + p^2}{r^2} + \Delta g^{rr}. \tag{5.2.14}$$

Then the shift is given by

$$\begin{aligned}\Delta g^{rr} = -\frac{4}{15r^6}\Big(&6\,\alpha_{ijkl}\,(q^iq^j - p^ip^j)(q^kq^l - p^kp^l) \;+\; 24\beta_{ijkl}q^ip^jq^kp^l \\ &+ \gamma_{ij}\;\big(q^iq^j - p^ip^j\big)\left(12Q^2 - 25Mr + 10r^2\right) \\ &+ 12\,\chi_{ijkl}\,q^ip^j\;\left(q^kq^l - p^kp^l\right) + 2\,\omega_{ij}\,q^ip^j\left(12Q^2 - 25Mr + 10r^2\right)\Big).\end{aligned} \tag{5.2.15}$$

To find the shift to extremality that results from this, we examine when the new radial function $g^{rr}(r, M, Q)$ has zeros [2]. This equation is sixth order in $r$, but we are only interested in the first-order shift to the solution. We Taylor-expand near the extremal solution where $r = M$ and $Q = M$, and keep only terms that are first-order in Wilson coefficients:

$$\begin{aligned}g^{rr}(r, M, Q) \;&=\; g^{rr}(M, M, M) + (Q - M)\,\partial_Q g^{rr}|_{(M,M,M)} + (r - M)\,\partial_r g^{rr}|_{(M,M,M)} \\ &=\; \Delta g^{rr}(M, M, M) + (Q - M)\,\partial_Q g^{rr}|_{(M,M,M)}.\end{aligned} \tag{5.2.16}$$

We have kept $M$ fixed. In going from the first to the second line, we have used that the uncorrected

[2]Equivalently we could examine the zeros of $g_{tt}$. This must give identical results since the consistency of the metric signature requires that $g_{tt}$ and $g^{rr}$ have the same set of zeros.

metric vanishes at $(M, M, M)$ so $g^{rr}(M, M, M) = \Delta g^{rr}(M, M, M)$. We also used that the uncorrected metric also has vanishing $r-$derivative at $(M, M, M)$, so the last term on the first line may be removed because it is second-order in Wilson coefficients. The requirement that $g^{rr}$ leads to the condition:

$$g^{rr}(r, M, Q) = 0 \implies Q - M = -\frac{\Delta g^{rr}(M, M, M)}{\partial_Q g^{rr}(M, M, M)}. \tag{5.2.17}$$

Now we evaluate this expression and divide by $m$ to find the result for the extremality bound $|\vec{z}|^2 = Q^2/M^2$

$$\begin{aligned}|\vec{z}| \leq 1 + \frac{2}{5(Q^2)^3}\Big(&2\,\alpha_{ijkl}\,(q^iq^j - p^ip^j)(q^kq^l - p^kp^l) \,+\, 8\beta_{ijkl}q^ip^jq^kp^l - \gamma_{ij}\,\left(q^iq^j - p^ip^j\right)Q^2 \\ &+ 4\,\chi_{ijkl}\,q^ip^j\,\left(q^kq^l - p^kp^l\right) - 2\,\omega_{ij}\,q^ip^jQ^2\Big) + \mathcal{O}\left(\frac{1}{(Q^2)^2}\right).\end{aligned} \tag{5.2.18}$$

This is the main technical result of this chapter. In the next section, we comment on the constraints that black hole decay might place on these coefficients, and we analyze this expression for the case of black holes with two electric charges, and the case of black holes with a single electric and single magnetic charge.

## 5.3 Black Hole Decay and the Weak Gravity Conjecture

As described by [166] and reviewed in Section 5.1, a state with charge-to-mass vector $\vec{z}$ and total charge $Q^2 \equiv \sum_i((q^i)^2 + (p^i)^2)$ is kinematically allowed to decay to a general multiparticle state only if $\vec{z}$ lies in the convex hull of the light charged states. In the case of asymptotically large extremal black holes decaying to finite charge black holes, the spectrum of light states corresponds to the region compatible with the extremality bound. This bound describes a surface in $z$-space of the form

$$|\vec{z}| = 1 + T(\vec{z}, Q^2), \tag{5.3.1}$$

where $T \to 0$ as $Q^2 \to \infty$. The convex hull condition [166] has a natural generalization to the sector of extremal black hole states:

**Black Hole Convex Hull Condition:** *It is kinematically possible for asymptotically large extremal black holes to decay into smaller finite $Q^2$ black holes only if the convex hull of the extremality surface encloses the unit ball $|\vec{z}| \leq 1$.*

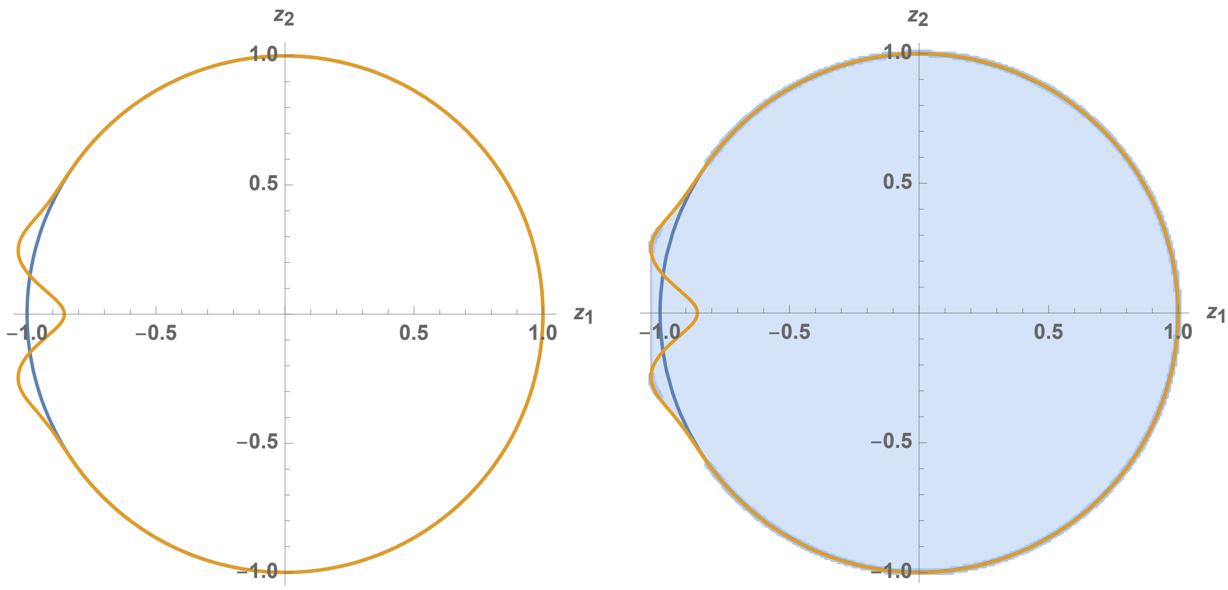



Figure 5.1: (Left): an extremality curve that naively violates the WGC as it does not enclose the unit circle. (Right): the convex completion of the extremality curve *does* enclose the unit circle, hence the WGC is satisfied. For this to be possible the extremality surface must be somewhere locally non-convex, which is shown in Appendix J to be impossible in the perturbative regime.

This means that to determine if the decay of a large black hole is kinematically allowed, we must first determine the convex hull of a complicated surface, a task that may only be tractable numerically. As illustrated in figure 5.1, it is possible for the convex hull of the extremality surface to enclose the unit ball even if the surface itself does not. Furthermore, the extremality surface may be non-convex even if the magnitude of the corrections is arbitrarily small.

The condition simplifies somewhat in the $Q^2 \gg 1$ regime, where the corrections to the unit circle derive from the four-derivative terms and are small as a result. In Appendix J we prove that if $T(\vec{z}, Q^2)$ is a quartic form, as it is in the explicit result (5.2.18), then the smallness of the deviation *does* imply convexity. In this regime, the convex hull condition is simplified in the sense that the extremality surface always bounds a convex region. At a given $Q^2 \gg 1$, and $\vec{z}$, the black hole extremality bound describes a surface in $z$-space of the form

$$|\vec{z}| = 1 + \frac{1}{(Q^2)^3} T_{ijkl} z^i z^j z^k z^l + \mathcal{O}\left(\frac{1}{(Q^2)^2}\right). \tag{5.3.2}$$

The condition for the multi-charge weak gravity conjecture to be satisfied in the perturbative regime degenerates to the more tractable condition:

**(Perturbative) Black Hole Weak Gravity Conjecture:** *It is kinematically possible for asymptotically large extremal black holes to decay into smaller finite $Q^2$ extremal black holes if the quartic extremality form*

$$T(q^i, p^i) = T_{ijkl} z^i z^j z^k z^l , \tag{5.3.3}$$

*is everywhere non-negative. Using the parametrization of the effective action (6.2.1), this bound takes the form*

$$\begin{aligned} T(q^i, p^i) \;&=\; 2\alpha_{ijkl}(q^i q^j - p^i p^j)(q^k q^l - p^k p^l) + 8\beta_{ijkl} q^i p^j q^k p^l - \gamma_{ij}\, Q^2 \left(q^i q^j - p^i p^j\right) \\ &\quad + 4\,\chi_{ijkl}\, q^i p^j \,\left(q^k q^l - p^k p^l\right) - 2\,\omega_{ij}\, Q^2 q^i p^j \geq 0\,, \end{aligned} \tag{5.3.4}$$

*which follows directly from (5.2.18).*

### 5.3.1 Examples

According to the previous section, we can determine whether black holes are stable by checking if the extremality form is anywhere negative. In this section we demonstrate this with a few basic examples.

#### 5.3.1.1 Black Hole With Two Electric Charges

A black hole that is electrically charged under two $U(1)$ groups provides one simple example. In this case, the extremality bound simplifies to

$$(2\alpha_{ijkl} - \gamma_{ij}\delta_{kl}) q^i q^j q^k q^l > 0. \tag{5.3.5}$$

As the $q$ factors project to the completely symmetric part of this tensor, it is convenient to define $T_{ijkl} = 2\alpha_{\{ijkl\}} - \gamma_{\{ij}\delta_{kl\}}$, where we have symmetrized the indices with weight one. Expanding the constraint in components leads to

$$T_{1111}\, q_1^4 + T_{1112}\, q_1^3\, q_2 + T_{1122}\, q_1^2\, q_2^2 + T_{1222}\, q_1\, q_2^3 + T_{2222}\, q_2^4 > 0. \tag{5.3.6}$$

This polynomial must be positive for all possible combinations of $q_1$ and $q_2$. We use the fact that the polynomial in (5.3.6) is homogenous, and divide by $(q_2)^4$. Redefining $q_1/q_2 = x$ simplifies the left-hand-side of the inequality to a polynomial of one variable:

$$T_{1111}\, x^4 + T_{1112}\, x^3 + T_{1122}\, x^2 + T_{1222}\, x + T_{2222} > 0. \tag{5.3.7}$$

This polynomial is quartic so one may solve this by studying the explicit expressions for the roots and demanding that they are not real. However the positivity conditions for fourth order polynomials are much simpler and lead to a set of relations among the components of $T_{ijkl}$ (see, for instance, [168]). This allows the problem to be solved entirely in the case of two charges; for $N > 2$ one must analyze multivariate polynomials.

For an example of a theory that may be in the Swampland, consider the following four-derivative terms:

$$\mathcal{L}_4 = \alpha_{1111}\, F^1_{\mu\nu} F^{1\,\mu\nu} F^1_{\rho\sigma} F^{1\,\rho\sigma} + \alpha_{1122}\, F^1_{\mu\nu} F^{1\,\mu\nu} F^2_{\rho\sigma} F^{2\,\rho\sigma} + \alpha_{2222}\, F^2_{\mu\nu} F^{2\,\mu\nu} F^2_{\rho\sigma} F^{2\,\rho\sigma}, \tag{5.3.8}$$

where $\alpha_{1111} = 2$, $\alpha_{1122} = -8$, and $\alpha_{2222} = 3$. Then the extremality shift becomes

$$2\, q_1^4 - 8\, q_1^2\, q_2^2 + 3\, q_2^4 > 0. \tag{5.3.9}$$

The inequality is satisfied when $q_1 = 0$ or $q_2 = 0$, but at $q_1 = q_2$, the extremality shift is negative. Therefore, a black hole with $q_1 = q_2$ in this theory would not be able to decay to smaller black holes. This model requires the existence of self-repulsive states in the spectrum in either the particle or stringy regimes to evade the Swampland.

#### 5.3.1.2 Dyonic Black Hole

Another simple case occurs when there is only a single gauge field but the black hole has both electric and magnetic charge. Then the extremality bound is obtained by removing all indices from (5.2.18):

$$2\alpha\,(q^2 - p^2)^2 + 8\beta\, q^2 p^2 - \gamma\,(q^2 - p^2)(q^2 + p^2) + 4\chi\, qp(q^2 - p^2) - 2\omega\, qp(q^2 + p^2) \ > \ 0. \tag{5.3.10}$$

We recover the results of [46] when the magnetic charge is set to zero. A single electric charge shifts the extremality as

$$|z_q| = 1 + \frac{2}{5|Q|^2}(2\alpha - \gamma). \tag{5.3.11}$$

However, a single magnetic charge has the opposite sign for $\gamma$:

$$|z_p| = 1 + \frac{2}{5|Q|^2}(2\alpha + \gamma). \tag{5.3.12}$$

Requiring that both types of black holes be able to decay places a stronger constraint on $\alpha$ and $\gamma$:

$$2\alpha > |\gamma|. \tag{5.3.13}$$

If we assume that both $p$ and $q$ are non-zero, we can again divide by $p^4$ as we did in the previous section, and again find a polynomial of a single variable:

$$(2\alpha-\gamma)\,y^4+(4\chi-2\omega)\,y^3+(-4\alpha+8\beta)\,y^2+(-4\chi-2\omega)\,y+(2\alpha+\gamma)\;>\;0. \tag{5.3.14}$$

The generalized bound (5.3.14) coincides exactly with the (regularized forward-limit) scattering positivity bounds derived in [48] for arbitrary linear combinations of external states. It is interesting that the requirement that dyonic black holes are unstable gives a new physical motivation for these generalized scattering bounds.

For the case of a single gauge field, a very physical example comes to mind: the Euler-Heisenberg Lagrangian [169], in which integrating out electron loops induces a four-point interaction among the gauge fields.[3] This model has four derivative terms given by

$$\mathcal{L}_4=\alpha(F_{\mu\nu}F^{\mu\nu})^2+\beta(F_{\mu\nu}\tilde{F}^{\mu\nu})^2, \tag{5.3.15}$$

with $\alpha=4$, $\beta=7$ (up to overall constants that do not effect the problem). The inequality that must be satisfied is the following:

$$4y^4+40y^2+8>0. \tag{5.3.16}$$

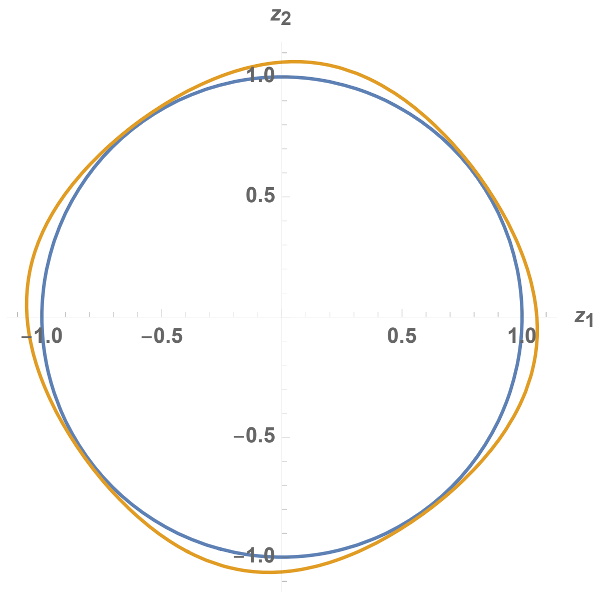
Clearly this holds for all values of $y$. Thus, we have found that the Euler-Heisenberg theory is not in the Swampland. This does not require that we know anything about the spectrum, or that the higher-derivative operators came from integrating out a particle at all. Only the four-derivative couplings are needed to learn that this theory allows nearly extremal black holes to decay.

The condition (5.3.10) exhibits an interesting simplification when $\alpha=\beta$ and the remaining coefficients are set to zero. In this case, the condition on the quartic form then reads

$$\alpha(q^2+p^2)^2>0. \tag{5.3.17}$$

In this special case the extremality surface becomes invariant under orthogonal rotations in charge-space. In fact, it is simple to verify that this is the only choice of coefficients with this feature. The enhanced symmetry is a consequence of the electromagnetic duality invariance of the equations of motion for this choice of coefficients. In the effective action, the necessary condition for

[3] The electron should also contribute to the $WFF$-type operators as well, but this contribution is suppressed by a factor of $1/z$. The electron is extraordinarily superextremal ($z=2\times10^{21}$) so we can safely ignore these terms for our example.



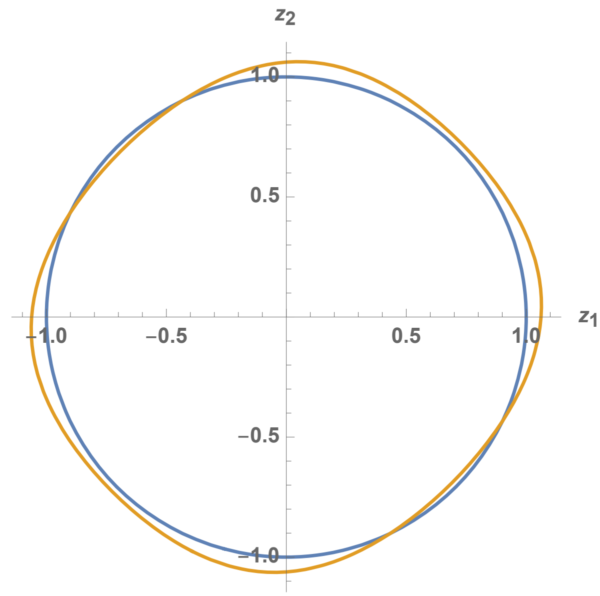



Figure 5.2: (Left): the corrections to the extremality curve are everywhere positive, hence the WGC is satisfied. (Right): the corrections to the extremality curve are *not* everywhere positive; large extremal black holes cannot always decay to intermediate mass black holes, whether or not the WGC is satisfied cannot be decided in the low-energy EFT.

duality invariance is the *Noether-Gaillard-Zumino condition* [170]

$$F_{\mu\nu}\tilde{F}^{\mu\nu} + G_{\mu\nu}\tilde{G}^{\mu\nu} = 0, \quad \text{where} \quad \tilde{G}_{\mu\nu} \equiv 2\frac{\delta S}{\delta F^{\mu\nu}}. \tag{5.3.18}$$

One can verify that this is satisfied if we $\alpha = \beta,\ \gamma = \chi = \omega = 0$ as above, at least to fourth order in derivatives. To make this equation hold to sixth order would require the addition of sixth-derivative operators to the Lagrangian, and so on. For a general analysis of electric-magnetic duality invariant theories, see [158]. In the following section we show that the generalization of the electromagnetic duality group from $U(1)$ in the single charge case, to $U(N)$ in the $N$-charge case plays an essential role in renormalization group running of the four-derivative Wilson coefficients.

### 5.3.2 Unitarity and Causality

Infrared consistency conditions on the low energy effective theory have been used to bound the coefficients of higher-derivative operators. Such constraints were first considered in the context of the weak gravity conjecture in [171], and were extended to the case of multiple gauge fields in [167]. Further arguments based on unitarity and causality were given in [47]. Here we review these arguments and present a few generalizations.

#### 5.3.2.1 Integrating Out Massive Particles

One source of higher derivative corrections derives from integrating out states in the particle regime. By this we mean states that are well described by ordinary QFT on a fixed spacetime background. Such states necessarily have masses smaller than some cutoff scale $\Lambda_{QFT}$, which is the string scale or whatever scale new physics invalidates the QFT description. We have already seen a simple example of this in the Euler-Heisenberg Lagrangian above.

At tree-level, only neutral particles contribute to the four-point interactions. Consider, for example, a dilaton that couples to the field strengths. The Lagrangian for the scalar theory is

$$\mathcal{L} = \frac{R}{4} - \frac{1}{2}(\partial\phi)^2 - \frac{m_\phi^2}{2}\phi^2 - \frac{1}{4}F^i_{\mu\nu}F^{i\,\mu\nu} + \mu_{ij}\phi F^i_{\mu\nu}F^{j\,\mu\nu}. \tag{5.3.19}$$

We integrate out the scalar to find the effective four-derivative coupling by matching to the low-energy EFT at the scale $\Lambda_{\mathrm{UV}} \lesssim m_\phi$

$$\mathcal{L}_4 \supset \frac{M_{\mathrm{Pl}}^4}{m_\phi^2}(\mu_{ij}\mu_{kl} + \mu_{ik}\mu_{jl} + \mu_{il}\mu_{jk})F^i_{\mu\nu}F^{j\,\mu\nu}F^k_{\rho\sigma}F^{l\,\rho\sigma}. \tag{5.3.20}$$

Therefore, in this simple setup, the coefficient $\alpha_{ijkl}$ takes the form

$$\alpha_{ijkl} = \frac{1}{m_\phi^2}(\mu_{ij}\mu_{kl} + \mu_{ik}\mu_{jl} + \mu_{il}\mu_{jk}). \tag{5.3.21}$$

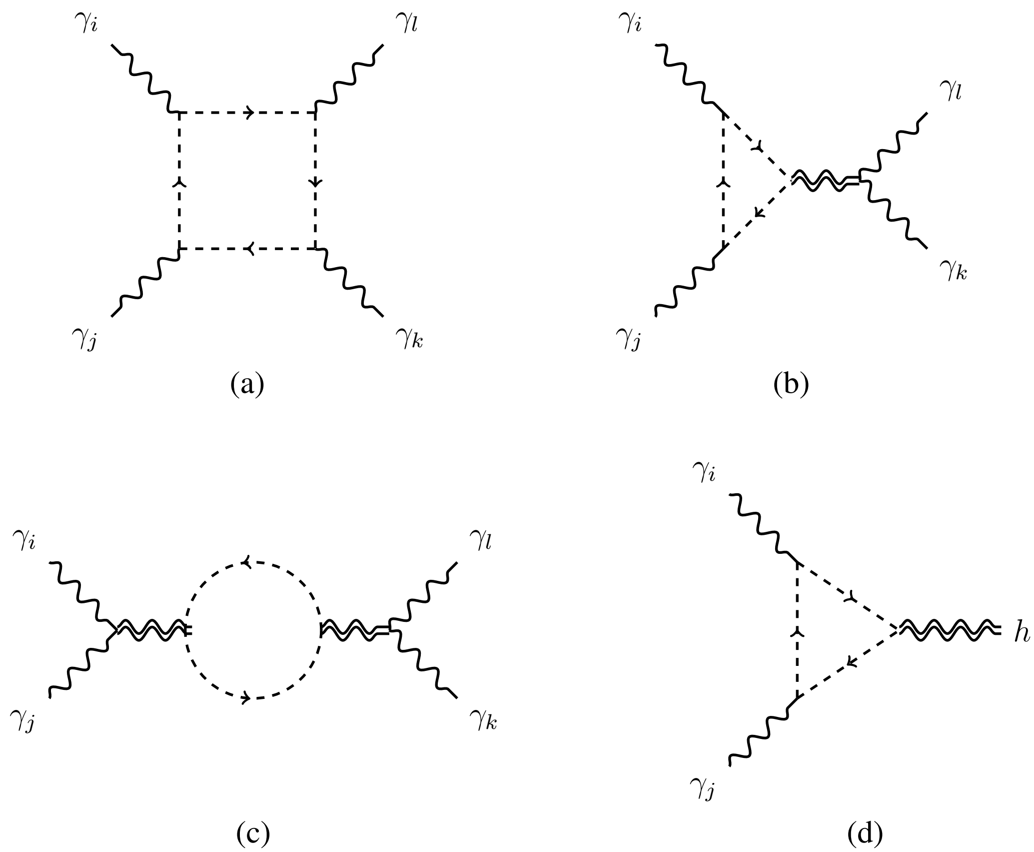
For a single gauge field $\alpha = \frac{3\mu^2}{m_\phi^2}$. Unitarity requires that $\mu$ is real, which implies that $\alpha$ is positive [47]. It is easy to see that this is still the case when there are more gauge fields. The extremality form for this theory is

$$\alpha_{ijkl}q^iq^jq^kq^l = \frac{3}{m_\phi^2}(\mu_{ij}q^iq^j)^2, \tag{5.3.22}$$

which must be positive.[4] The same reasoning shows that integrating out an axion, which couples to $F^i\tilde{F}^j$, generates $\beta_{ijlk}$, and that its contribution to the extremality form is also positive.

Light charged particles cannot contribute at tree-level so their leading contributions are at loop-level. The diagrams that contribute in this case are:

[4]Note that unlike the case of single gauge field, unitarity does not bound all the coefficients separately. For instance, in the two charge case, $\mu_{11} = 1$, $\mu_{22} = -1$, and $\mu_{12} = 0$ would lead to $\alpha_{1122} = -1/m_\phi^2$.




These contribute at the same order except they have relative factors of $z_\phi$, the particle's charge-to-mass ratio, coming from counting couplings and propagators. Diagram (a) goes like $z_\phi^4$, (b) like $z_\phi^2$, (c) like $z_\phi^0$; diagram (d) contributes at order $z_\phi^2$. The field-strength four-point interaction is generated by the first three diagrams. In the limit where $z_\phi \gg 1$, diagram (a) dominates all the others (as we noted above in the Euler-Heisenberg example) and the extremality form becomes

$$T_{ijlk}q^iq^jq^kq^l = \alpha_{ijkl}q^iq^jq^kq^l = (z_\phi^i q^i)^4, \tag{5.3.23}$$

Again, we find a manifestly positive contribution. For $z_\phi$ near or less than one, both $\alpha_{ijkl}$ and $\gamma_{ij}$ are generated by diagrams that are order $z_\phi^0$. In that case this scaling argument does not apply, and the order one constants need to be included in the analysis. These arguments are schematic and largely review what was already considered in [167].

One might wonder whether this analysis is relevant to the parity-odd operators. Interestingly, [172] has shown how to generalize the Euler-Heisenberg Lagrangian by integrating out a monopole or dyonic charge. The effective Lagrangian was derived in that chapter (and earlier in [173]) to be

$$\mathcal{L}_4 = \big(4(\hat{q}^2-\hat{p}^2)^2 + 28\hat{q}^2\hat{p}^2\big)(F^2)^2 + \big(7(\hat{q}^2-\hat{p}^2)^2 + 16\hat{q}^2\hat{p}^2\big)(F\tilde{F})^2 - 12\hat{q}\hat{p}(\hat{q}^2-\hat{p}^2)F^2(F\tilde{F}). \tag{5.3.24}$$

where the $\hat{q}$ and $\hat{p}$ refer to the electric and magnetic charges of the dyon that is integrated out (not the charges of the black hole). This procedure generates the parity-violating four-photon coupling as well as the two parity-even ones. This is not surprising given that magnetic charges violate parity in their interactions with the gauge field. What is more interesting is that this term is *not* a square, unlike every other term appearing in the effective Lagrangian. The sign of the generated term depends on the sign of the product of the electric and magnetic charges of the particle. In terms of the polynomial derived in (5.3.14), the condition that must be met to satisfy the WGC is:

$$\begin{aligned}\left(\hat{q}^4+5\hat{q}^2\hat{p}^2+\hat{p}^4\right)x^4+\,3\left(\hat{q}^3\hat{p}-\hat{q}\hat{p}^3\right)x^3+\,\left(5\hat{q}^4-8\hat{q}^2\hat{p}^2+5\hat{p}^4\right)x^2\\ +\,3\left(\hat{q}^3\hat{p}-\hat{q}\hat{p}^3\right)x\,+\,\left(\hat{q}^4+5\hat{q}^2\hat{p}^2+\hat{p}^4\right)\,>\,0.\end{aligned} \tag{5.3.25}$$

This polynomial is always positive, so the Lagrangian given in (5.3.24) does not allow for stable black holes and satisfies the WGC.

#### 5.3.2.2 Causality Constraints

Another set of arguments for bounds on the EFT coefficients rely on causality. These were first considered in [171] and generalized to multiple gauge fields in [47]. Two methods were used, and they were shown to give the same result. The first is to consider the propagation of photons on a photon gas background. Requiring that photons travel do not travel superluminally constrains the four-photon interaction. The second method uses analyticity and unitarity to relate the EFT coefficients to an integral over the imaginary part of the amplitude, which is manifestly positive. The bounds obtained this way for multiple gauge fields are

$$\sum_{ij}\left(\alpha_{\{ij\}\{kl\}}+\beta_{\{ij\}\{kl\}}\right)u^iv^ju^kv^l\geq 0. \tag{5.3.26}$$

This inequality must hold for any vectors $\vec{u}$ and $\vec{v}$. This bound is independent from the bounds that we have derived in (5.3.4), so it is not enough to imply the WGC on its own.

So far these arguments have only bounded the four-photon interactions. Another causality-based argument was made in [47] that bounds the photon-photon-graviton interaction parameterized by $\gamma$. They argued that the addition of this four-derivative term introduces causality violation at a scale $E\sim M_{\rm Pl}/\gamma^{1/2}$ (a fact noticed in [174]). Therefore new physics must arise at scale $\Lambda_{QFT}\lesssim M_{\rm Pl}/\gamma^{1/2}$, which means $\gamma\lesssim(M_{\rm Pl}/\Lambda_{QFT})^2$. This argument suggests that perhaps the $WFF$ four-derivative terms are generically bounded by causality to be much smaller than a number of possible contributions to the $F^4$ terms. It would be interesting to extend the analysis of [174] to the more general set of operators used here, but this is beyond the scope of our chapter.

## 5.4 Renormalization of Four-Derivative Operators

The Wilson coefficients that appear in the extremality shift (5.2.18) are determined by UV degrees-of-freedom integrated out of the low-energy effective field theory. In Section 5.3 we gave explicit examples of contributions to the Wilson coefficients from integrating out massive particle states, both at tree- and loop-level. To consistently calculate the correction to the extremality bound for a black hole with total charge $Q^2$, we must first calculate the renormalization group evolution from the *matching scale* $\mu^2 \sim \Lambda_{\text{UV}}^2$ to the *horizon scale* $\mu^2 \sim M_{\text{Pl}}^2/Q^2$. For black holes with $Q^2 \gg 1$ these scales can be arbitrarily separated and the effects of the logarithmic running of the Wilson coefficients can be dramatic.

In the single $U(1)$ case it was recently argued [50] that as we RG flow towards the deep IR, $Q^2 \to \infty$, the logarithmic running of a particular combination of Wilson coefficients dominates the extremality shift, independent of the values of the coefficients at the matching scale. Explicitly, the extremality bound takes the form

$$\frac{Q^2}{M^2} \leq 1 + \frac{4}{5Q^2}\left(\frac{c}{16\pi^2}\log\left(\frac{\Lambda_{\text{UV}}^2 Q^2}{M_{\text{Pl}}^2}\right) + 2\alpha_{\text{UV}} - \gamma_{\text{UV}}\right). \tag{5.4.1}$$

If $c > 0$ then at some finite value of the charge $Q^2$ extremal black holes must be self-repulsive. This was shown to be the case in [50] for various explicit theories, including the single $U(1)$ model (1.2.3). Since the renormalization group coefficient $c$ depends only on the massless degrees of freedom, this analysis depends only on the universality class of the model. For those classes in which this conclusion holds, the WGC is always satisfied independently of the details of the UV completion, and in that sense is no longer a useful Swampland criterion.

In this section we show how this argument generalizes to an arbitrary number of $U(1)$ gauge fields. Since there are many more four-derivative operators, we emphasize the importance of a non-renormalization theorem that arises as a consequence of the accidental $U(N)$ electromagnetic duality symmetry of the two-derivative approximation. In the following subsection we give an on-shell proof of this theorem, and then use it to extend the argument above.

### 5.4.1 Non-Renormalization and Electromagnetic Duality

Consider a low-energy effective action of the form

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{4}R - \frac{1}{4}F_{\mu\nu}^i F^{i\mu\nu} + \sum_i c_i \mathcal{O}_i\right], \tag{5.4.2}$$

where the operators $\mathcal{O}_i$ have at least three derivatives. The Wilson coefficient $c_i$ is *renormalized* if $\mathcal{O}_i$ corresponds to a counterterm to an ultraviolet divergence. In terms of on-shell scattering amplitudes, an operator $\mathcal{O}_i$ generates an on-shell local matrix element, and the coefficient $c_i$ is renormalized if there is a corresponding one-loop scattering amplitude with an ultraviolet divergence. Here "corresponding" means that the external states of the matrix element of $\mathcal{O}_i$ must agree with the external states of the loop amplitude. Conversely, an operator $\mathcal{O}_i$ is *not* renormalized if there are no corresponding UV divergent one-loop scattering amplitudes with the correct external states.

We begin by making the observation that the leading, two-derivative, part of the action (5.4.2) has an accidental $O(N)$ flavor symmetry. This leads to a rather trivial non-renormalization theorem:

*In Einstein-Maxwell with $N$ $U(1)$ gauge fields, a four-derivative operator $\mathcal{O}_i$ is renormalized at one-loop only if it generates an on-shell local matrix element that is an invariant tensor of the flavor symmetry group $O(N)$.*

This statement is trivial because there are no Feynman diagrams at one-loop that are not $O(N)$-invariant. Since we are not assuming that $O(N)$ is a symmetry of the UV completion, such symmetry violating higher-derivative operators may appear in the effective action, but they cannot act as counterterms to ultraviolet divergences, and hence their associated Wilson coefficients do not have a logarithmic running. Trivial non-renormalization theorems of this kind follow for all symmetries of the effective action.

The non-trivial non-renormalization theorem we prove below concerns *electromagnetic duality symmetries*, which are only symmetries of the equations of motion, not the action [170]. Consequently, they are not manifest off-shell, meaning diagram-by-diagram in the standard covariant Feynman diagram expansion, and the above reasoning is no longer valid. Nonetheless we will prove that the above non-renormalization theorem is valid verbatim, at least at one-loop, where the flavor symmetry group $O(N)$ is enhanced to the maximal compact electromagnetic duality group $U(N)$.

It is convenient to discuss UV divergences in the context of dimensional regularization where the loop integration is performed in $d = 4 - 2\epsilon$ dimensions and ultraviolet divergences at one-loop appear as $1/\epsilon$ poles. In this context we can classify the sources of UV divergences in on-shell scattering amplitudes:

1. **Cut-Constructible Divergences**: By standard integral reduction algorithms, one-loop am-

plitudes admit a universal decomposition into a sum over a set of *master integrals*:

$$\mathcal{A}_n^{\text{1-loop}} = \sum_i a_i I_i^{(\text{box})} + \sum_j b_j I_j^{(\text{triangle})} + \sum_k c_k I_k^{(\text{bubble})} + \mathcal{R}, \tag{5.4.3}$$

where the master integrals are *scalar* integrals with the indicated topology. Here $a_i, b_j, c_k$ and $\mathcal{R}$ are rational functions of the external kinematic data. The first three contributions are often referred to as the *cut-constructible* part of the amplitude; they contain all of the branch cut discontinuities required by perturbative unitarity at one-loop. These contributions can be completely determined from on-shell unitarity cuts into physical tree amplitudes [13, 138]. This determines the one-loop amplitude up to a rational ambiguity indicated by $\mathcal{R}$. Since the rational part is both UV and IR finite, the divergent structure (both UV and IR) of the one-loop amplitude is completely determined by the tree-level scattering amplitudes. From the definition it is clear that only the master bubble integral $I^{\text{bubble}}$ is UV divergent,

$$\left[I^{\text{bubble}}\left(K^2\right)\right]_{\text{UV}} \equiv \left[\int \frac{\mathrm{d}^{4-2\epsilon} l}{(2\pi)^{4-2\epsilon}} \frac{1}{l^2(l-K)^2}\right]_{\text{UV}} = \frac{i}{16\pi^2\epsilon}, \tag{5.4.4}$$

and therefore what we call the *cut-constructible divergence* is proportional to the sum of the bubble coefficients $c_k$. These coefficients are completely determined by the two-particle unitarity cuts of the one-loop amplitude. It has been shown that the two-particle unitarity cuts of the master bubble integrals are purely rational functions, while the two-particle cuts of the triangle and box integrals give logarithms [13, 138]. By explicitly calculating the two-particle cuts of $\mathcal{A}_n^{\text{1-loop}}$ one can read off the rational part as the associated bubble coefficient. Using the relation between unitarity cuts of one-loop amplitudes and on-shell phase space integrals of tree-amplitudes gives a well-known general formula for the cut-constructible UV divergence

$$[\mathcal{A}_n^{\text{1-loop}}]_{\text{UV}} = \sum_{\text{cuts}} \left[\int \mathrm{d}\mu_{\text{LIPS}} \sum_{\text{states}} \mathcal{A}_L^{\text{tree}} \mathcal{A}_R^{\text{tree}}\right]_{\text{Rational}}, \tag{5.4.5}$$

where the sums on the right-hand-side are taken over all cuts *and* all on-shell states exchanged in each cut. The details of the integration in this formula are not essential to the argument we make below.

2. **UV/IR Mixed Divergences**: In dimensional regularization IR divergences are also regularized as $1/\epsilon$ poles. Even though their physical origin is very different there can be non-trivial cancellations between UV and IR divergences in the on-shell scattering amplitude. Such *mixed* UV divergences are just as important as the cut-constructible ones, and must be included to calculate the correct beta functions [175, 176]. Unfortunately, due to this cancellation they cannot be immediately extracted from the cut-constructible part of the one-loop

amplitude (5.4.3). The strategy is to first independently determine the expected one-loop IR divergence, and then compare against the IR divergences in the cut-constructible part of the amplitude. Any discrepancy must be due to UV/IR cancellations, and so can be used to infer the mixed UV divergences. The true IR divergent structure is determined by the KLN theorem [177]. This states that in an inclusive cross-section, virtual IR divergences from loop integration must cancel against divergences in the initial/final phase space integrals that arise from soft/collinear real emission. Such real emission singularities are fixed by tree-level soft/collinear limits, so we find that again the mixed divergences are completely reconstructible from tree-level, physical data.

We begin with an on-shell description of $U(N)$ duality invariance at tree-level. The three-particle amplitudes are completely fixed [5]:

$$\begin{aligned} \mathcal{A}_3^{\text{tree}}\left(1_h^+, 2_h^+, 3_h^-\right) &= \frac{[12]^6}{[23]^2[31]^2}, \quad \mathcal{A}_3^{\text{tree}}\left(1_h^-, 2_h^-, 3_h^+\right) = \frac{\langle 12\rangle^6}{\langle 23\rangle^2\langle 31\rangle^2}, \\ \mathcal{A}_3^{\text{tree}}\left(1_h^+, 2_{\gamma,i}^+, 3_\gamma^{-,j}\right) &= \delta_i{}^j\frac{[12]^4}{[23]^2}, \quad \mathcal{A}_3^{\text{tree}}\left(1_h^-, 2_{\gamma,i}^+, 3_\gamma^{-,j}\right) = \delta_i{}^j\frac{\langle 13\rangle^4}{\langle 23\rangle^2}, \end{aligned} \tag{5.4.6}$$

where $i, j = 1, ..., N$ are flavor indices. The fact that the on-shell three-particle amplitudes are diagonal in flavor space with unit coupling to the graviton is an on-shell expression of the Einstein equivalence principle. $U(N)$ duality invariance is encoded in the on-shell Ward identity:

$$U_i{}^k U^{*j}{}_l \mathcal{A}_3^{\text{tree}}\left(1_h^+, 2_{\gamma,k}^+, 3_\gamma^{-,l}\right) = \mathcal{A}_3^{\text{tree}}\left(1_h^+, 2_{\gamma,i}^+, 3_\gamma^{-,j}\right), \tag{5.4.7}$$

where $U \in U(N)$. In the explicit expressions above this is seen to hold as a consequence of the fact that $\delta_i{}^j$ is a $U(N)$-invariant tensor. The 4-point amplitudes are simple to calculate using on-shell recursion

$$\begin{aligned} \mathcal{A}_4^{\text{tree}}\left(1_h^+, 2_h^+, 3_h^-, 4_h^-\right) &= \frac{[12]^4\langle 34\rangle^4}{s_{12}s_{13}s_{14}}, \quad \mathcal{A}_4^{\text{tree}}\left(1_h^+, 2_{\gamma,i}^+, 3_h^-, 4_\gamma^{-,j}\right) = \delta_i{}^j\frac{[12]^4\langle 34\rangle^2\langle 23\rangle^2}{s_{12}s_{13}s_{14}}, \\ \mathcal{A}_4^{\text{tree}}\left(1_{\gamma,i}^+, 2_{\gamma,j}^+, 3_\gamma^{-,k}, 4_\gamma^{-,l}\right) &= \delta_i{}^k\delta_j{}^l\frac{[12]^2\langle 34\rangle^2}{s_{13}} + \delta_i{}^l\delta_j{}^k\frac{[12]^2\langle 34\rangle^2}{s_{14}}. \end{aligned} \tag{5.4.8}$$

Again, each of these is a $U(N)$-invariant tensor. As we discussed above, in the standard Lorentz covariant Feynman diagrammatic approach, only the $O(N)$ subgroup of global flavor rotations is manifest. The enhancement to the full $U(N)$ duality invariance in the on-shell amplitudes appears miraculous. A simple way to see that this enhancement continues to all multiplicities is to calculate the tree-amplitudes using on-shell recursion. Here the amplitude is given as a sum

[5] The spinor-helicity conventions used in these expressions are given in [71].

over factorization channels of the form

$$\mathcal{A}_n^{\text{tree}} \sim \sum_{\text{channels}} \sum_{\text{states}} \mathcal{A}_L^{\text{tree}} \mathcal{A}_R^{\text{tree}}. \tag{5.4.9}$$

The precise details of the formula are not important to the argument. It is straightforward to prove $U(N)$ invariance by induction. Assume that all tree amplitudes $\mathcal{A}_m^{\text{tree}}$, with $m < n$ are duality invariant; using the recursive representation (5.4.9) we show that $\mathcal{A}_n^{\text{tree}}$ is duality invariant channel-by-channel. If the exchanged on-shell state in a given channel is a graviton, then $\mathcal{A}_L^{\text{tree}}\mathcal{A}_R^{\text{tree}}$ is a product of invariant tensors, and hence invariant. If the exchanged state is a photon then the sum over helicity and the flavor index takes the form

$$\mathcal{A}_L^{\text{tree}}\left(..., -p_{\gamma,i}^{+}\right) \mathcal{A}_R^{\text{tree}}\left(p_{\gamma}^{-,i}, ...\right) + \mathcal{A}_L^{\text{tree}}\left(..., -p_{\gamma}^{-,i}\right) \mathcal{A}_R^{\text{tree}}\left(p_{\gamma,i}^{+}, ...\right). \tag{5.4.10}$$

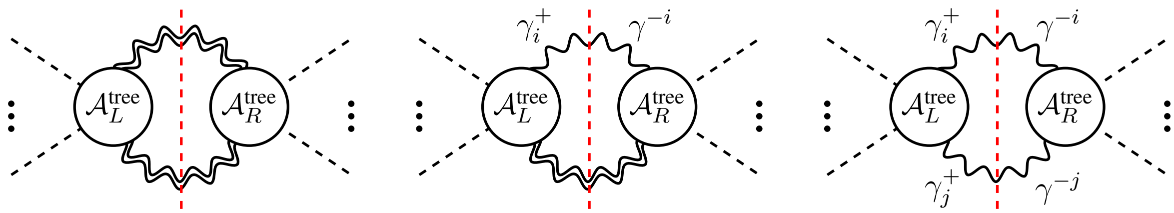
Since this is the contraction of two tensors by the invariant $\delta_i{}^j$, it follows that this sum is likewise an invariant. Together with the explicitly verified duality invariance of the three-point amplitudes, the all-multiplicity Ward identity follows by induction. Here the key property we used was the existence of a valid on-shell recursion for the tree-level S-matrix (5.4.9); a general discussion the necessary conditions for this to exist can be found in [79].

We are now ready to prove the following non-renormalization theorem:

**Non-Renormalization of Duality Violating Operators**: *In Einstein-Maxwell with $N$ $U(1)$ gauge fields, a four-derivative operator $\mathcal{O}_i$ is renormalized at one-loop only if it generates an on-shell local matrix element that is an invariant tensor of the maximal compact electromagnetic duality group $U(N)$.*

This result was first noted long-ago following a detailed calculation of the UV divergence [178, 179], and recently generalized (including massless scalars) to the full non-compact duality group $Sp(2N)$ in [180]. The new result in this section is a simple argument that demonstrates the duality invariance of the divergence *without* the need for a detailed calculation.

We will prove that the total UV divergence is given by a sum over $U(N)$ invariant tensors. Beginning with the cut-constructible part, the logic here is very similar to the inductive proof of tree-level invariance. We will show that the divergence is a $U(N)$ invariant tensor cut-by-cut. In the representation (5.4.5) we consider the contribution of a single two-particle cut; this can be either graviton-graviton, graviton-photon or photon-photon:




Since the tree-amplitudes are invariant, and as in the expression (5.4.10) the exchanged photon flavor indices are contracted with invariant tensors, each case separately generates an invariant tensor. Summing over all states and cuts we conclude that the cut-constructible divergence is duality invariant.

As for the possible mixed divergence, here we begin with the full IR divergence at one-loop. This is given by the universal formula [181]

$$\left[\mathcal{A}_n^{\text{1-loop}}\right]_{\text{IR}} = \frac{ir_\Gamma}{(4\pi)^{2-\epsilon}}\frac{1}{\epsilon^2}\mathcal{A}_n^{\text{tree}}\sum_{i\neq j}^{n}(s_{ij})^{1-\epsilon}, \tag{5.4.11}$$

where the tree-amplitude on the right-hand-side and the loop amplitude on the left-hand-side have the same external states and $r_\Gamma = \Gamma^2(1-\epsilon)\Gamma(1+\epsilon)/\Gamma(2-\epsilon)$. As discussed above, in general there may be non-trivial UV/IR cancellations in the cut-constructible part of the one-loop amplitude. These can be disentangled using knowledge of the full IR divergence. In this case, things are somewhat simpler, and expanding the final factor in (5.4.11) gives

$$\sum_{i\neq j}^{n}(s_{ij})^{1-\epsilon} = \sum_{i\neq j}^{n} s_{ij} + \epsilon\sum_{i\neq j}^{n} s_{ij}\log\left(s_{ij}\right) + \mathcal{O}\left(\epsilon^2\right). \tag{5.4.12}$$

The first term in this sum is zero by momentum conservation. Expanding (5.4.11) the full IR divergence has the form

$$\left[\mathcal{A}_n^{\text{1-loop}}\right]_{\text{IR}} = \frac{i}{16\pi^2\epsilon}\mathcal{A}_n^{\text{tree}}\sum_{i\neq j}^{n} s_{ij}\log\left(s_{ij}\right) + \mathcal{O}\left(\epsilon^0\right). \tag{5.4.13}$$

We see that the coefficient of the IR divergence is a *transcendental* function. We know, however, that the coefficients of UV divergences are always *rational* functions, since they must be removable by adding local counterterms. It follows that there can never be any UV/IR mixing at one-loop in perturbative quantum gravity and hence that the complete UV divergence is given by the cut-constructible part of the amplitude. This completes the proof of the non-renormalization theorem.

It is important to note that this theorem is valid independent of any anomalies in the duality symmetries. Indeed, in the absence of additional massless degrees of freedom, we expect a non-vanishing ABJ anomaly in the duality currents $j_D^\mu$ [182, 183]. Explicitly for the $N = 1$ case:

$$\langle \nabla_\mu j_D^\mu \rangle = \frac{1}{24\pi^2} R_{\mu\nu\rho\sigma} \tilde{R}^{\mu\nu\rho\sigma}. \tag{5.4.14}$$

This is a mixed-gravitational anomaly. The question of how this manifests in on-shell scattering amplitudes in the context of $\mathcal{N} = 4$ supergravity has been a subject of recent interest [164, 184]. Such an anomalous violation of the $U(N)$-invariance at one-loop can appear only in the rational part of the amplitude since the cut-constructible part is completely fixed by unitarity cuts into tree-level amplitudes. The anomaly is therefore irrelevant to the effects of duality invariance on non-renormalization at one-loop. At two-loops however, anomalous rational one-loop amplitudes will have a noticeable effect on ultraviolet divergences and may lead to the renormalization of duality violating six-derivative operators. This question deserves further study.

### 5.4.2 RG Flow and the Multi-Charge Weak Gravity Conjecture

With the non-renormalization theorem proven in the previous section, we now show how the argument given in [50] generalizes to the multi-charge case. By simple dimensional analysis we know that the counter-terms to one-loop divergences in Einstein-Maxwell are four-derivative operators. In Appendix G we give a complete classification of local matrix elements corresponding to four-derivative operators, so together with the non-renormalization theorem proven in the previous section we know that most general *local* UV divergence is given by

$$\left[\mathcal{A}_4^{\text{1-loop}} \left(1^+_{\gamma,i}, 2^+_{\gamma,j}, 3^{-,k}_{\gamma}, 4^{-,l}_{\gamma}\right)\right]_{\text{UV}} = \frac{c}{16\pi^2\epsilon} \left(\delta_i{}^k \delta_j{}^l + \delta_i{}^l \delta_j{}^k\right) [12]^2 \langle 34\rangle^2. \tag{5.4.15}$$

At one-loop, the divergence fixes the dependence of the scattering amplitude on the renormalization group scale $\mu^2$. After adding a counterterm with coefficient $\alpha(\mu)$ to remove the UV divergence, the *physical* scattering amplitude should be independent of $\mu^2$

$$\mathcal{A}_4^{\text{1-loop}} \left(1^+_{\gamma,i}, 2^+_{\gamma,j}, 3^{-,k}_{\gamma}, 4^{-,l}_{\gamma}\right) = \left[\alpha(\mu^2) + \frac{c}{8\pi^2} \log(\mu^2)\right] \left(\delta_i{}^k \delta_j{}^l + \delta_i{}^l \delta_j{}^k\right) [12]^2 \langle 34\rangle^2 + \mathcal{O}\left(\epsilon^0\right), \tag{5.4.16}$$

which gives the logarithmic running of the Wilson coefficient

$$\alpha(\mu^2) = -\frac{c}{8\pi^2} \log\left(\frac{\mu^2}{\Lambda_{\text{UV}}^2}\right), \tag{5.4.17}$$

where $\Lambda_{\text{UV}}$ is some UV matching scale, assumed to be arbitrarily larger than the horizon scale. The ultraviolet divergence in Einstein-Maxwell coupled to $N$ $U(1)$ gauge fields was first calculated long-ago [178, 179], and then recalculated using unitarity methods [181, 185]

$$\left[\mathcal{A}_4^{\text{1-loop}}\left(1^+_{\gamma,i}, 2^+_{\gamma,j}, 3^{-,k}_{\gamma}, 4^{-,l}_{\gamma}\right)\right]_{\text{UV}} = \frac{1}{16\pi^2\epsilon}\left(\frac{137}{120}+\frac{N-1}{20}\right)\left(\delta_i{}^k\delta_j{}^l+\delta_i{}^l\delta_j{}^k\right)[12]^2\langle 34\rangle^2. \tag{5.4.18}$$

This gives the RG coefficient in (5.4.15) as

$$c = \frac{137}{120}+\frac{N-1}{20}. \tag{5.4.19}$$

From this matrix element we can reverse engineer the corresponding four-derivative operator

$$S \supset \alpha(\mu^2)\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right)\int \mathrm{d}^4x\sqrt{-g}\left[\left(F^i_{\mu\nu}F^{j\,\mu\nu}F^k_{\rho\sigma}F^{l\,\rho\sigma}+F^i_{\mu\nu}\tilde{F}^{j\,\mu\nu}F^k_{\rho\sigma}\tilde{F}^{l\,\rho\sigma}\right)\right]. \tag{5.4.20}$$

Note that we have lost manifest duality invariance when passing from on-shell scattering amplitudes to the effective action and so have made the replacement $\delta_i{}^j \to \delta_{ij}$. As an important cross-check, the effect of such an operator on the perturbed metric at leading order in $\alpha$ is given by (5.2.15) to be

$$\Delta g^{rr} = -\frac{24\alpha(\mu^2)}{15r^6}\sum_{i=1}^{N}\left(q_i^2+p_i^2\right), \tag{5.4.21}$$

which manifests the expected electromagnetic duality symmetry, further enhanced to $O(2N)$.

When evaluating the extremality form, $\mu$ should be taken to be the horizon scale $\mu^2 \sim M_{\text{Pl}}^4/M^2 \sim M_{\text{Pl}}^2/Q^2$. Since $c>0$, as $Q^2\to\infty$ the logarithmic term becomes *large* and *positive*. With the logarithmic running included the extremality form at the horizon scale is given by

$$\begin{aligned} T(q^i,p^i) = {} & \frac{1}{8\pi^2}\left(\frac{137}{120}+\frac{N-1}{20}\right)(Q^2)^2\log\left(\frac{\Lambda_{\text{UV}}^2Q^2}{M_{\text{Pl}}^2}\right)+\alpha^{\text{UV}}_{ijkl}\,(q^iq^j-p^ip^j)(q^kq^l-p^kp^l) \\ & + 8\beta^{\text{UV}}_{ijkl}q^ip^jq^kp^l-\gamma^{\text{UV}}_{ij}\,\left(q^iq^j-p^ip^j\right)Q^2+4\,\chi^{\text{UV}}_{ijkl}\,q^ip^j\,\left(q^kq^l-p^kp^l\right) \\ & -2\,\omega^{\text{UV}}_{ij}\,q^ip^jQ^2, \end{aligned} \tag{5.4.22}$$

where $Q^2=\sum_i(q_i^2+p_i^2)$. In this expression $\alpha^{\text{UV}}$, $\beta^{\text{UV}}$, $\gamma^{\text{UV}}$, $\chi^{\text{UV}}$, and $\omega^{\text{UV}}$ refers to the values of the Wilson coefficients at the matching scale $\Lambda_{\text{UV}}$. Importantly, the logarithmic term is $O(2N)$ invariant and therefore gives an isotropic contribution to the extremality form. This contribution scales like $Q^4\log Q$, while the rest of the terms scale like $Q^4$. Therefore it dominates over all other contributions. We conclude that for sufficiently large $Q^2$, the extremality form is positive, independent of the values of the Wilson coefficients at the matching scale $\Lambda_{\text{UV}}$, and consequently the multi-charge WGC is always satisfied in the black hole regime.

Here the full $U(N)$ duality invariance of the UV divergence (enhanced to $O(2N)$ in the quartic form) was essential to the argument. It would not have been enough that some Wilson coefficients had a positive logarithmic running, to prove the multi-charge WGC we require positivity in all directions, which as we have shown follows from a generalized non-renormalization theorem as a consequence of tree-level $U(N)$ duality symmetry of Einstein-Maxwell.

It is interesting to note that we can *almost* reach this same conclusion without knowing the explicit form of the UV divergence (5.4.18). In [47] the causality bound (5.3.26) was applied to the Wilson coefficients at the UV matching scale $\Lambda_{\mathrm{UV}}$ and consequently to constrain the properties of the states integrated out. But this bound must remain valid even deeper in the IR where, as we have seen, the logarithmic running dominates. If the RG coefficient $c$ had been negative, then the bound (5.3.26) is eventually violated, indicating the presence of superluminal propagation at very low energies. Since we expect that Einstein-Maxwell is not inconsistent in the deep IR, it must be the case that $c \geq 0$ even without doing a detailed one-loop calculation. This argument has nothing to say about the possibility that $c = 0$. Only an explicit calculation is sufficient to demonstrate the existence of a non-vanishing one-loop divergence.

# CHAPTER 6

# Higher-Derivative Corrections to Entropy and the Weak Gravity Conjecture in Anti-de Sitter Space

## 6.1 Weak Gravity Conjecture and Black Hole Entropy

Recently, an intriguing argument was made by Remmen, Cheung and Liu relating the black hole WGC in flat space and a conjectured positivity of higher-derivative correcions to the Wald entropy [49] of thermodynamically stable black holes. The authors first show that, near extremality, the shift to the extremality bound at fixed charge and temperature is proportional to the shift in entropy at fixed charge and mass. They then present an argument that the higher-derivative corrections should increase the entropy, thereby proving the black hole WGC. The argument for the entropy shift positivity is not expected to be fully general; it applies to higher-derivative corrections that arise from integrating out massive particles at tree-level. Nonetheless, it is curious that the entropy shift is proportional to the extremality shift. This fact was given a simple thermodynamic proof in [186], where no assumptions were made about the particulars of the background.

So far, however, these ideas have not fully made their way to Anti-de Sitter space. From the WGC point of view, it is easy to see why: the relationship between mass and charge of an extremal black holes in AdS is *already non-linear at the two-derivative level*[1]. Therefore it is not at all clear what is gained by studying the higher-derivative corrections to the extremal mass-to-charge ratio[2]. Furthermore, massive particles emitted from a black hole cannot escape to infinity in AdS as they can in flat space, so if the WGC allows for the instability of black holes in AdS, it must be through a completely different mechanism.

Regardless, the entropy-extremality relationship is expected to hold in AdS as it does in flat space (and indeed, an example in $AdS_4$ was given in [186]). Therefore, this chapter addresses two main issues in Anti-de Sitter space: first, we check the purported relationship between the entropy

[1]By “extremal,” we mean that the temperature is zero. This is not the same as the BPS limit in AdS.

[2]Other aspects of the WGC have been discussed in AdS. See e.g. [187–190].

shift and the extremality shift, and indeed we find that it holds for the AdS-Reissner-Nordström backgrounds. Second, we examine the conjecture that the entropy shift is positive when the leading-order solution is a minimum of the action. By computing the entropy shift explicitly, we see that its positivity for stable black holes implies that the coefficient of Riemann-squared is *universally positive*. This has interesting consequences for a potential bound on $\eta/s$, as we comment on in the section 6.6.

## 6.2 Corrections to the Geometry

We consider Einstein-Maxwell theory in the presence of a negative cosmological constant in a $(d+1)$-dimensional AdS spacetime of size $l$. The first non-trivial terms in the derivative expansion of the effective action arise at the four-derivative level, and by appropriate field redefinitions we may choose a complete basis of dimension-independent operators:

$$\begin{aligned} I = -\frac{1}{16\pi}\int d^{d+1}x\sqrt{-g}\Bigg[&\frac{d(d-1)}{l^2} + R - \frac{1}{4}F^2 \\ &+ l^2\epsilon\Big(c_1 R_{abcd}R^{abcd} + c_2 R_{abcd}F^{ab}F^{cd} + c_3(F^2)^2 + c_4 F^4\Big)\Bigg]. \end{aligned} \tag{6.2.1}$$

Note that additional CP-odd terms can arise in specific dimensions, but will not contribute to the static, stationary spherically symmetric black holes that we are considering here. This basis parallels that of [191], which used the same set of dimensionless Wilson coefficients, but focused on the $(4+1)$-dimensional case. Depending on the origin of the AdS length scale $l$, one may expect these coefficients to be parametrically small, of the form $c_i \sim (\Lambda l)^{-2}$, where $\Lambda$ denotes the scale at which the EFT breaks down. In particular, this will be the case in order for the action (6.2.1) to be under perturbative control. We have also introduced the small bookkeeping parameter $\epsilon$, which will allow us to keep track of which terms are first order in the $c_i$ coefficients.

### 6.2.1 The Zeroth Order Solution

At the two-derivative level, this action admits a family of AdS-Reissner-Nordström black holes parametrized by uncorrected mass $m$ and charge $q$,

$$\begin{gathered} ds^2 = -f(r)dt^2 + g(r)^{-1}dr^2 + r^2 d\Omega^2_{d-1,k}\,, \qquad f(r) = g(r) = k - \frac{m}{r^{d-2}} + \frac{q^2}{4r^{2d-4}} + \frac{r^2}{l^2}, \\ A = \left(-\frac{1}{c}\frac{q}{r^{d-2}} + \Phi\right)dt, \qquad c = \sqrt{\frac{2(d-2)}{(d-1)}}, \qquad \Phi = \frac{1}{c}\frac{q}{r_h^{d-2}}\,. \end{gathered} \tag{6.2.2}$$

Here $r_h$ is the outer horizon radius, and the parameter $k = 0, \pm 1$ specifies the horizon geometry, with $k = 1$ corresponding to the unit sphere. The constant $\Phi$ is chosen so that the $A_t$ component of the gauge field vanishes on the horizon, and represents the potential difference between the asymptotic boundary and the horizon.

Typically, we will consider lower case letters $(m, q, ...)$ to be parameters in the theory, while upper case letters $(M, Q, S, T, ...)$ will denote physical quantities that may or may not receive corrections. We will add a subscript zero (e.g. $M_0$) to denote the uncorrected contribution to quantities that do receive order $c_i$ corrections. The shifts, which are equal to the corrected quantities minus the uncorrected ones, will be denoted by the $\epsilon$ derivative. However, we will sometimes use $\Delta$ when it is convenient, with subscripts indicating quantities held fixed, for example, we have

$$(\Delta M)_T \equiv \lim_{\epsilon\to 0}\left(M(T,\epsilon) - M_0(T)\right) \equiv \lim_{\epsilon\to 0}\left(\frac{\partial M}{\partial \epsilon}\right)_T. \tag{6.2.3}$$

Finally, in Sections (6.4) and (6.5) we will use dimensionless quantities $(\nu, \xi)$ for convenience. These are defined by $\nu = (r_h)_0/l$ and $Q = (1-\xi)Q_{\text{ext}}$.

### 6.2.2 The First Order Solution

We now turn to the first order solution in terms of the Wilson coefficients $c_i$. We follow the procedure outlined in Ref. [46], but work in an AdS$_{d+1}$ background. While general $(d+1)$-dimensional results may be worked out analytically, we took a shortcut of working with explicit dimensions four through eight and then fitting the coefficients to extract results for arbitrary dimension. Since the four-derivative terms are built from tensors with eight indices and hence four metric contractions, the resulting expressions will scale at most as $d^4$. The coefficients are hence fully determined by working in five different dimensions.

Following [46], we start with the *effective* stress tensor, where corrections come from two sources. The first is from substituting in the corrected Maxwell field to the zeroth order electromagnetic stress tensor, and the second is from the explicit four-derivative corrections to the stress tensor evaluated on the zeroth order solution. The result of computing both of these contributions to the time-time component of the stress tensor is

$$
\begin{aligned}
T_t^{\ t} =& -\frac{(d-1)(d-2)\,q^2}{4\,r^{2d-2}} + \frac{d(d-1)}{l^2} \\
&+ c_1\Bigg(\frac{(d-2)(8d^3-24d^2+15d+3)\,q^4l^2}{8r^{4d-4}} - \frac{(d-1)(d-2)(4d^2-9d+3)\,mq^2l^2}{r^{3d-2}} \\
&\qquad + k\frac{4d(d-1)(d-2)^2\,l^2q^2}{r^{2d}} - \frac{d(d-1)(d-2)(d-3)\,l^2m^2}{r^{2d}} \\
&\qquad + \frac{(d-2)(2d-3)(2d^2-5d+1)\,q^2}{r^{2d-2}} + \frac{2d(d-3)}{l^2}\Bigg) \\
&+ c_2\Bigg(\frac{(d-1)^3(d-2)\,q^4l^2}{r^{4d-4}} - \frac{(d-1)^2(3d^2-8d+4)\,q^2ml^2}{r^{3d-2}} + k\frac{2d(d-1)^2(d-2)\,q^2l^2}{r^{2d}} \\
&\qquad + \frac{2(d-1)^3(d-2)\,q^2}{r^{2d-2}}\Bigg) + (2c_3+c_4)\left(\frac{(d-1)^2(d-2)^2q^4l^2}{2\,r^{4d-4}}\right).
\end{aligned}
\tag{6.2.4}
$$

The shift to the geometry may be obtained from the corrections to the stress tensor [46],

$$
\Delta g = \frac{1}{(d-1)r^{d-2}}\int dr\,r^{d-1}\Delta T_t^{\ t}\,, \tag{6.2.5}
$$

and after integrating the $\mathcal{O}(c_i)$ terms in (6.2.4), we find

$$
\begin{aligned}
\Delta g(r) =& c_1\Bigg(-\frac{(d-2)(8d^3-24d^2+15d+3)\,q^4l^2}{8(d-1)(3d-4)r^{4d-6}} + \frac{(d-2)(4d^2-9d+3)\,mq^2l^2}{2(d-1)r^{3d-4}} \\
&\qquad - k\frac{4(d-2)^2\,l^2q^2}{r^{2d-2}} + \frac{(d-2)(d-3)\,l^2m^2}{r^{2d-2}} - \frac{(2d-3)(2d^2-5d+1)\,q^2}{(d-1)r^{2d-4}} \\
&\qquad + \frac{2(d-3)r^2}{(d-1)l^2}\Bigg) \\
&+ c_2\Bigg(-\frac{(d-1)^2(d-2)\,q^4l^2}{(3d-4)r^{4d-6}} + \frac{(3d^2-8d+4)\,q^2ml^2}{2r^{3d-4}} - k\frac{2(d-1)(d-2)\,q^2l^2}{r^{2d-2}} \\
&\qquad - \frac{2(d-1)^2\,q^2}{r^{2d-4}}\Bigg) \\
&+ (2c_3+c_4)\left(-\frac{(d-1)(d-2)^2q^4l^2}{(6d-8)r^{4d-6}}\right).
\end{aligned}
\tag{6.2.6}
$$

The time component of the metric can then be obtained using the relation [46]

$$f(r) = (1 + \gamma(r))g(r), \tag{6.2.7}$$

where $\gamma(r)$ is defined by[3]

$$\gamma(r) = -\frac{1}{(d-2)} \int drr \left( T_t{}^t - T_r{}^r \right). \tag{6.2.8}$$

For our particular case we find:

$$\gamma(r) = \left( c_1 \frac{(d-2)(2d^2 - 5d + 1)}{(d-1)} + c_2 d(d-2) \right) \frac{q^2 l^2}{r^{2d-2}}. \tag{6.2.9}$$

Finally, we have

$$\begin{aligned} F_{tr} = \sqrt{\frac{(d-2)(d-1)}{2}} \Bigg[ & (1 - 8c_2) \frac{q}{r^{d-1}} + 4c_2(d-1)(d-2) \frac{qml^2}{r^{2d-1}} \\ & + \left( \frac{c_1}{2} \frac{(2d^2 - 5d + 1)}{(d-1)} - \frac{c_2}{2}(7d - 12) - 4\left(2c_3 + c_4\right)(d-1) \right) (d-2) \frac{q^3 l^2}{r^{3d-3}} \Bigg], \end{aligned} \tag{6.2.10}$$

which we note is independent of the geometry parameter $k$, as was the case in [192].

### 6.2.3 Asymptotic Conditions and Conserved Quantities

The first order solution can be summarized as

$$ds^2 = -\left(1 + \gamma(r)\right) g(r) dt^2 + g(r)^{-1} dr^2 + r^2 d\Omega_{d-1,k}^2, \tag{6.2.11}$$

where

$$g(r) = k - \frac{m}{r^{d-2}} + \frac{q^2}{4r^{2d-4}} + \frac{r^2}{l^2} + \Delta g. \tag{6.2.12}$$

The corrected metric functions, $\Delta g$ and $\gamma(r)$, are given in (6.2.6) and (6.2.8), respectively. In addition, the full electric field is given in (6.2.10). For a given zeroth order AdS radius $l$, this solution is specified by two parameters, $m$ and $q$, which correspond to the mass and charge of the uncorrected black hole. At the same time, the corrected solution includes a number of integration constants, two of which we have implicitly set to zero in the integral expressions for $\Delta g$ and $\gamma(r)$. The constant related to $\Delta g$ can be absorbed by a shift in $m$, and a third constant from the corrected Maxwell equation can be absorbed by a shift in $q$. The constant related to $\gamma(r)$ can be

[3] We note that the definition of $\gamma$ implies that it is positive provided that the null energy condition holds.

absorbed at the linearized level by a rescaling of the time coordinate, and hence can be thought of as a redshift factor.

In order to make the correspondence between the parameters of the solution, $m$ and $q$, and the physical mass and charge of the black hole more precise, consider the part of $\Delta g$ that is leading in $r$. We can see that there is a term that goes like $c_1 \frac{r^2}{l^2}$ that dominates over all other terms in the correction. Therefore, for large values of $r$, the solution takes the form

$$\begin{aligned} f(r) \approx g(r) &= k - \frac{m}{r^{d-2}} + \left(1 + c_1 \frac{2(d-3)}{d-1}\right) \frac{r^2}{l^2} + \cdots , \\ F_{tr} &= \sqrt{\frac{(d-2)(d-1)}{2}} (1 - 8c_2) \frac{q}{r^{d-1}} + \cdots . \end{aligned} \tag{6.2.13}$$

Our first observation is that the AdS radius gets modified because the Riemann-squared term is non-vanishing on the original uncorrected background. This suggests that we define an effective AdS radius

$$l^2 = \lambda^2 l_{\text{eff}}^2, \qquad \lambda^2 = \left(1 + c_1 \frac{2(d-3)}{(d-1)}\right). \tag{6.2.14}$$

This shift by $\lambda$ is unavoidable when turning on the $c_1$ Wilson coefficient. However, in principle we still have a choice of whether we hold $l$ or $l_{\text{eff}}$ fixed when turning on the four-derivative corrections.

In what follows, we always choose to keep $l$ fixed. Then, since the effective AdS radius is shifted, the asymptotic form of the metric is necessarily modified as well. From a holographic point of view, this leads to a modification of the boundary metric

$$ds^2 \sim r^2 \left(\frac{dt^2}{l^2} + d\Omega_{d-1,k}^2\right) \quad \longrightarrow \quad ds^2 \sim r^2 \left(\frac{dt^2}{l_{\text{eff}}^2} + d\Omega_{d-1,k}^2\right). \tag{6.2.15}$$

This is generally undesirable, as we would like to compare thermodynamic quantities in a framework where we hold the boundary metric fixed while turning on the Wilson coefficients. One way to avoid this shift in the boundary metric is to introduce a 'redshift' factor

$$t = \bar{t}/\lambda, \tag{6.2.16}$$

to compensate for the shift in $l_{\text{eff}}$. In terms of the time $\bar{t}$, the solution now takes the form

$$\begin{aligned} ds^2 &= -\bar{f}(r)\, d\bar{t}^2 + g(r)^{-1} dr^2 + r^2 d\Omega_{d-1,k}^2, \\ F_{\bar{t}r} &= \lambda^{-1} F_{tr} = \sqrt{\frac{(d-2)(d-1)}{2}} (1 - 8c_2) \frac{q/\lambda}{r^{d-1}} + \cdots , \end{aligned} \tag{6.2.17}$$

where

$$\bar{f}(r) = \lambda^{-2}(1+\gamma(r))g(r) = k/\lambda^2 - \frac{m/\lambda^2}{r^{d-2}} + \frac{r^2}{l^2} + \cdots,$$
$$g(r) = k - \frac{m}{r^{d-2}} + \frac{r^2}{l_{\text{eff}}^2} + \cdots. \tag{6.2.18}$$

We now turn to the charge and mass of the solution measured with respect to the redshifted $\bar{t}$ time. For the charge $Q$, we take the conserved Noether charge

$$Q = \frac{1}{16\pi}\int_{\Sigma_{d-1}} *\mathcal{F}, \tag{6.2.19}$$

where $\mathcal{F}$ is the effective electric field

$$\mathcal{F}_{\mu\nu} = F_{\mu\nu} + l^2\left(-4c_2 R_{\mu\nu\rho\sigma}F^{\rho\sigma} - 8c_3 F_{\mu\nu}(F^2) - 8c_4 F_{\nu\rho}F^{\rho\sigma}F_{\sigma\mu}\right). \tag{6.2.20}$$

The result is

$$Q = \frac{1+8c_2}{16\pi}\omega_{d-1}\lambda r^{d-1}F_{\bar{t}r}\Big|_{r\to\infty} = \sqrt{\frac{(d-2)(d-1)}{2}}\frac{\omega_{d-1}}{16\pi}q, \tag{6.2.21}$$

where $\omega_{d-1}$ is the volume of the unit $S^{d-1}$. The $1/16\pi$ factor arises from the prefactor in the action (6.2.1) where we have set Newton's constant $G = 1$.

Unlike in the asymptotically Minkowski case, some care needs to be taken in obtaining the mass of the black hole. With an eye towards holography, we choose to define the mass from the boundary stress tensor [193]. The standard approach to holographic renormalization involves the addition of appropriate local boundary counterterms so as to render the action finite. This was performed in [192] for $R^2$-corrected bulk actions, and since only the $c_1 R_{abcd}R^{abcd}$ term in (6.2.1) leads to an additional divergence, we can directly use the result of [192]. The result is

$$M = \frac{\omega_{d-1}}{16\pi}(1+4c_1(d-3))\frac{(d-1)m}{\lambda}, \tag{6.2.22}$$

where we have taken into account the scaling of the mass by the redshift factor $\lambda$. Substituting in $\lambda$ from (6.2.14) then gives

$$M = \frac{\omega_{d-1}}{16\pi}(d-1)(1+\rho)m, \tag{6.2.23}$$

where

$$\rho = c_1\frac{(d-3)(4d-5)}{d-1}. \tag{6.2.24}$$

Note that we are taking the mass here to *exclude* the Casimir energy that is normally part of the boundary stress tensor. This will be important when comparing with the thermodynamic quantities extracted from the regulated on-shell action in Section (6.4). Working in the setup of holographic renormalization ensures that the mass $M$ and charge $Q$ defined in (6.2.23) and (6.2.21), respectively, yield a consistent framework for black hole thermodynamics.

## 6.3 Mass, Charge, and Entropy from the geometry shift

Given the first-order solution, we now consider shifts to the mass, $\Delta M$, and entropy, $\Delta S$, of the black hole induced by the four-derivative corrections. In these computations it is important to keep in mind what is being held fixed as we turn on the Wilson coefficients $c_i$. The main parameters we consider here are the mass $M$ and charge $Q$, which are related to the two parameters, $m$ and $q$, of the solution by (6.2.23) and (6.2.21), respectively. In addition we consider the thermodynamic quantities $T$ (temperature) and $S$ (entropy), although they are not all independent. Note that we always consider the AdS radius $l$ to be fixed, although interesting results have been obtained by mapping it to thermodynamic pressure.

Singly-charged, non-rotating black holes may be described by any two of mass $M$, charge $Q$ and the horizon radius $r_h$. Of course, any number of other parameters may be used as well, such as the temperature $T$ or an extremality parameter, such as was used in [49]. If we further impose the extremality condition $T = 0$ on the solution, then only a single parameter is needed. Clearly this is only true for non-rotating black holes with a single gauge field, as more general solutions may have additional charges or angular momenta.

Here we mainly focus on the effect of higher-derivative corrections on extremal or near extremal black holes. In particular, we consider the extremality shift $\Delta(M/Q)$ and the entropy shift $\Delta S$. However, it is important to keep in mind what is being held fixed when we turn on the higher-derivative corrections, as the results will depend on this choice. For example, we will see below that the shift to $M/Q$ depends on whether the mass, charge or horizon radius is held fixed when comparing the corrected with uncorrected quantities.

### 6.3.1 Mass, Charge, and Extremality

Recall that, in our first-order solution, the geometry is essentially given by the radial function

$$g^{rr} = g(r) = k - \frac{m}{r^{d-2}} + \frac{q^2}{4r^{2d-4}} + \frac{r^2}{l^2} + \Delta g\,, \tag{6.3.1}$$

where $\Delta g$ denotes the contributions of the higher-derivative corrections to the geometry, and $\epsilon$ is a small parameter we use to keep track of where $\mathcal{O}(c_i)$ corrections come in. Using the fact that both $g(r_h)$ and $g'(r_h)$ vanish at extremality, we may express the extremal mass and charge as a function of the horizon radius,

$$
\begin{aligned}
M_{\text{ext}} &= 2V(d-1)r_h^{d-2}\left(\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right)(1+\epsilon\rho)+\epsilon\,\Delta g+\frac{r_h}{2(d-2)}\epsilon\,\Delta g'\right)\,,\\
Q_{\text{ext}}^2 &= 2V^2(d-1)(d-2)r_h^{2(d-2)}\left(k+\frac{d}{d-2}\frac{r_h^2}{l^2}+\epsilon\,\Delta g+\frac{r_h}{d-2}\epsilon\,\Delta g'\right)\,,
\end{aligned}
\tag{6.3.2}
$$

where $M$ and $Q$ are the asymptotic quantities defined in (6.2.23) and (6.2.21), and we have defined $V=\omega_{d-1}/16\pi$. Though we have expressed $M$ and $Q$ as functions of $r_h$, these expressions are valid regardless of which of the three quantities is being held fixed. For example, if we work at fixed charge, then $Q$ gets no $\mathcal{O}(\epsilon)$ corrections, in which case $M$ and $r_h$ will both receive corrections.

#### 6.3.1.1 Extremality at Leading Order

Before discussing the extremality and entropy shifts, we consider the leading order relations between $M_0$, $Q_0$ and $(r_h)_0$ for extremal black holes. We will suppress the 0 subscripts in this subsection, but we mean the uncorrected quantities. Setting $\epsilon=0$ in (6.3.2) immediately gives the relations

$$
\begin{aligned}
M_{\text{ext}} &= 2V(d-1)r_h^{d-2}\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right)\,,\\
Q_{\text{ext}}^2 &= 2V^2(d-1)(d-2)r_h^{2(d-2)}\left(k+\frac{d}{d-2}\frac{r_h^2}{l^2}\right)\,.
\end{aligned}
\tag{6.3.3}
$$

In principle, we can eliminate $r_h$ from these equations to obtain the relation between mass and charge for extremal AdS black holes. However, for general dimension $d$, there is no simple expression that directly encodes this relation. Nevertheless, we can consider the limit of small and large black holes.

For small black holes ($r_h \ll l$), we take $k=1$ (ie a spherical horizon) and find

$$
M_{\text{ext}} \sim Q_{\text{ext}} \sim r_h^{d-2}\,,
\tag{6.3.4}
$$

so one recovers the simple $M\sim Q$ scaling that appears in flat space. (Note that asymptotically Minkowski black holes necessarily have spherical horizons.) For large black holes ($r_h \gg l$), on the other hand, the scaling is very different from that of flat space,

$$
M_{\text{ext}} \sim r_h^d\,,\qquad Q_{\text{ext}} \sim r_h^{d-1}\qquad \Rightarrow\qquad M_{\text{ext}} \sim (Q_{\text{ext}})^{\frac{d}{d-1}}\,.
\tag{6.3.5}
$$

In fact, this is precisely the scaling behavior expected based on the relationship between minimal scaling dimension and charge for boundary operators with large global charges [194].

#### 6.3.1.2 Mass Shift at Fixed Charge

Now we consider the effect of four-derivative corrections. If we hold the charge fixed, then the shift to extremality is entirely due to the change in the mass. This may computed from the expression (6.3.2) for the mass by taking a derivative with respect to $\epsilon$, which parametrizes the higher-derivative corrections, leading to

$$\begin{aligned}\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} &= V(d-1)r_h^{d-2}\Bigg(2\Delta g + \frac{1}{d-2}r_h\Delta g' \\ &+ 2\rho\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right) + \frac{2}{(d-2)r_h}\left((d-2)^2k + d(d-1)\frac{r_h^2}{l^2}\right)\left(\frac{\partial r_h}{\partial \epsilon}\right)\Bigg),\end{aligned} \tag{6.3.6}$$

where we have taken into account the fact that when the charge is fixed, we must allow the horizon radius $r_h$ to vary with $\epsilon$. To compute the shift $\partial r_h/\partial\epsilon$, we use the fact that we are holding $Q$ fixed. Then we use the expression for $Q_{\text{ext}}$ in (6.3.2) and demand that $(\partial Q/\partial\epsilon)_{T=0} = 0$ to obtain an equation for $\partial r_h/\partial\epsilon$. This procedure leads to the rather simple result

$$\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} = V(d-1)r_h^{d-2}\left(\Delta g + 2\rho\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right)\right). \tag{6.3.7}$$

Note that the dependence on $\Delta g'$ has vanished. From the geometric point of view, this non-trivial cancellation is crucial for the extremality-entropy relation to hold.

#### 6.3.1.3 Charge Shift at Fixed Mass

If we instead hold the mass fixed, the entire shift in the extremality is due to the shift in charge. Following the same procedure as in the fixed charge case, but this time demanding $\partial M_{\text{ext}}/\partial\epsilon = 0$, we find the relation:

$$\left(\frac{\partial Q^2}{\partial \epsilon}\right)_{M,T=0} = -2V^2(d-1)(d-2)r_h^{2d-4}\left(\Delta g + 2\rho\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right)\right). \tag{6.3.8}$$

Here we also find a cancellation of all $\Delta g'$ terms. Moreover, this shift is proportional to the mass shift at fixed charge

$$\left(\frac{\partial Q^2}{\partial \epsilon}\right)_{M,T=0} = -2V(d-2)r_h^{d-2}\left(\frac{dM}{d\epsilon}\right)_{Q,T=0}. \tag{6.3.9}$$

This relationship more clear when we write this as the shift of $Q$ rather than $Q^2$. Using $\Delta Q^2 = 2Q\Delta Q$, we find

$$Q\left(\frac{\partial Q}{\partial \epsilon}\right)_{M,T=0} = -V(d-2)r_h^{d-2}\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} . \tag{6.3.10}$$

Finally, we use $\Phi = Q/(d-2)Vr^{d-2}$ to write:

$$\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} = -\Phi\left(\frac{\partial Q}{\partial \epsilon}\right)_{M,T=0} . \tag{6.3.11}$$

So we see that the mass shift is related to the charge shift times the potential. In Appendix L, we derive this statement for a general thermodynamic system and show that it holds for any extensive charge and its conjugate.

One physical consequence of this fact is that the entropy-extremality relationship (with a different proportionality factor) will hold regardless of whether the mass or charge is held fixed. As far as we know, this has not been noticed before in the literature.

#### 6.3.1.4 Summary of Extremality Shifts

The shifts to extremality may be obtained from these mass and charge shifts. For completeness, we also present calculation at fixed horizon radius, as this extremality shift has previously been considered in the literature as well [191, 192],

$$\begin{aligned}
\left(\frac{M}{Q}\right)_{Q,T=0} &= \left(\frac{M}{Q}\right)_0\left(1+\rho+\Delta g\,\frac{1}{2\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right)}\right), \\
\left(\frac{M}{Q}\right)_{M,T=0} &= \left(\frac{M}{Q}\right)_0\left(1+\rho\,\frac{k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}}{k+\frac{d}{d-2}\frac{r_h^2}{l^2}}+\Delta g\,\frac{1}{2\left(k+\frac{d}{d-2}\frac{r_h^2}{l^2}\right)}\right), \\
\left(\frac{M}{Q}\right)_{r_h,T=0} &= \left(\frac{M}{Q}\right)_0\left(1+\rho+\frac{\Delta g\left(k+\frac{d+1}{d-2}\frac{r_h^2}{l^2}\right)+r_h\Delta g'\frac{1}{(d-2)^2}\frac{r_h^2}{l^2}}{2\left(k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}\right)\left(k+\frac{d}{d-2}\frac{r_h^2}{l^2}\right)}\right),
\end{aligned} \tag{6.3.12}$$

where the corrections are encoded in $\rho$ and $\Delta g$ given in (6.2.24) and (6.2.6), respectively (and $\Delta g'$ as well for the fixed $r_h$ case). For these final results, we have set $\epsilon = 1$. However, the expressions are only valid to first order in the Wilson coefficients $c_i$. Here the uncorrected charge to mass ratio may be obtained from (6.3.3), and takes the form

$$\left(\frac{M}{Q}\right)_0 \quad = \quad \sqrt{\frac{2(d-1)}{d-2}}\frac{k+\frac{d-1}{d-2}\frac{r_h^2}{l^2}}{\sqrt{k+\frac{d}{d-2}\frac{r_h^2}{l^2}}}\,. \tag{6.3.13}$$

Note that, in (6.3.12), the horizon radius $r_h$ may be taken to be the uncorrected radius, and can be obtained from either $M$ or $Q$ using the leading order expressions (6.3.3). In (6.3.13), the leading order expression for $r_h$ should be used. Finally, note that $\Delta g$ depends on the parameters $m$ and $q$ as well as the radius $r$. The $m$ and $q$ parameters are directly obtained from $M$ and $Q$ using (6.2.23) and (6.2.21), and again the leading order horizon radius can be used in $\Delta g$.

### 6.3.2 Wald Entropy

We now compare the shift in mass at fixed charge and temperature to the shift in entropy at fixed mass and charge. The entropy for black holes in higher-derivative theories is given by the Wald entropy [195]:

$$S = -2\pi \int_\Sigma \frac{\delta \mathcal{L}}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\nu}\epsilon_{\rho\sigma}\,. \tag{6.3.14}$$

For spherically symmetric backgrounds, the integral over the horizon $\Sigma$ gives a factor of the area $A$. The two-derivative contribution to the entropy is simply $S^{(2)} = A/4$, while the four-derivative terms yield

$$S^{(4)} = -2\pi A \frac{\delta\Delta\mathcal{L}}{\delta R_{\mu\nu\rho\sigma}}\epsilon_{\mu\nu}\epsilon_{\mu\nu}\bigg|_{\partial^4} = -\frac{A}{4}l^2(4c_1 R_{trtr} + 2c_2 F_{tr}F_{tr})\,. \tag{6.3.15}$$

The total entropy is the sum of these terms,

$$S = \frac{A}{4}\left(1-\epsilon\left(4c_1 l^2 R_{trtr} + 2c_2 l^2 F_{tr}F_{tr}\right)\right)\bigg|_{r_h}\,, \tag{6.3.16}$$

where we once again introduced $\epsilon$ to parametrize the expansion. Here the horizon area is given by $A = \omega_{d-1} r_h^{d-1}$, where $r_h$ is the corrected horizon radius. On the other hand, the $R_{trtr}$ and $F_{tr}F_{tr}$ terms need only be computed on the zeroth-order background,

$$\begin{aligned} R_{trtr} &= \frac{1}{l^2} + \frac{(2d-3)(Q/V)^2}{2(d-1)r^{2d-2}} - \frac{(d-2)M/V}{2r^d}\,, \\ F_{tr}F_{tr} &= \frac{(Q/V)^2}{r^{2d-2}}\,. \end{aligned} \tag{6.3.17}$$

It does not matter whether we use the corrected or uncorrected quantities here because they already show up in a term that is order $\epsilon$. Note also that, while the expression for the Wald entropy (6.3.16) is given in terms of $M$, $Q$ and $r_h$ of the fully corrected solution, only two of these quantities are independent.

We now examine the entropy shift for a given solution at fixed mass $M$ and charge $Q$. For the moment, we work at arbitrary $M$ and $Q$, and not necessarily at extremality. The general expression for the entropy shift is then

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{Q,M} = \frac{A}{4}\left((d-1)\left(\frac{1}{r_h}\frac{\partial r_h}{\partial \epsilon}\right)_{Q,M} - \left(4c_1 l^2 R_{trtr} + 2c_2 l^2 F_{tr}F_{tr}\right)\right), \tag{6.3.18}$$

where the first term was obtained by

$$\frac{1}{A}\frac{\partial A}{\partial \epsilon} = (d-1)\frac{1}{r_h}\frac{\partial r_h}{\partial \epsilon}. \tag{6.3.19}$$

Here it is important to note that the horizon radius $r_h$ receives a correction when working at fixed $M$ and $Q$. If, on the other hand, we were to keep the horizon radius fixed (as is done in [191]), we would find only the second (interaction) term in (6.3.18), and the entropy shift would be independent of $c_3$ and $c_4$.

To compute $\partial r_h/\partial\epsilon$, we start with the horizon condition $g(r_h) = 0$ where $g(r)$ is given by (6.3.1) with $m$ and $q$ rewritten in terms of $M$ and $Q$. Taking a derivative and solving for $\partial r_h/\partial\epsilon$ then gives

$$\frac{1}{r_h}\frac{\partial r_h}{\partial \epsilon} = -\frac{\rho M + V(d-1)r_h^{d-2}\Delta g}{(d-2)(M-(M_{\text{ext}})_0)}. \tag{6.3.20}$$

where $(M_{\text{ext}})_0$ is the leading order extremal mass given in (6.3.3). As we can see, this expression diverges if the leading order solution is extremal. This is in fact not a surprise, as leading order extremality implies a double root at the horizon. The higher order corrections will lift this double root and hence cannot be parametrized as a linear shift in $\epsilon$.

In order to avoid the divergence, we can instead consider a leading order solution taken slightly away from extremality. As long as we are sufficiently close to extremality, the first term in (6.3.18) will dominate the entropy shift. Noting further that, at extremality, the numerator of (6.3.20) becomes proportional to the mass shift (6.3.7) at fixed charge, we can rewrite (6.3.18) as

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{Q,M} = -\frac{A}{4}\left(\frac{d-1}{(d-2)(M-(M_{\text{ext}})_0)}\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} + \frac{d-1}{d-2}\rho + 4c_1 l^2 R_{trtr} + 2c_2 l^2 F_{tr}F_{tr}\right). \tag{6.3.21}$$

The deviation away from extremality can be written in terms of the leading order temperature,

$$4\pi T_0 = |g'((r_h)_0)|_{\epsilon=0} = \frac{(d-2)(M-(M_{\text{ext}})_0)}{V(d-1)(r_h)_0^{d-1}}\,. \tag{6.3.22}$$

The total shift to the entropy is then given by

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{Q,M} = -\frac{1}{T_0}\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} - \frac{A}{4}\left(\frac{d-1}{d-2}\rho + 4c_1 l^2 R_{trtr} + 2c_2 l^2 F_{tr}F_{tr}\right)\,. \tag{6.3.23}$$

Finally, as $T_0 \to 0$ we reproduce the relation [49, 186]

$$\left(\frac{\partial M}{\partial \epsilon}\right)_{Q,T=0} = -T_0\left(\frac{\partial S}{\partial \epsilon}\right)_{Q,M}\,. \tag{6.3.24}$$

Note that this relation was obtained using only the general feature that the corrected geometry may be written in terms of a shift $\Delta g$ to the radial function $g(r)$. In particular, we never had to use the explicit form of $\Delta g$ given in (6.2.6).

### 6.3.3 Explicit Results for the Entropy Shifts

In order to compare with the next section, we include some explicit results for the mass shifts. In Section (6.6), we will see what constraints may be placed on the EFT coefficients by imposing that entropy shift is positive. We'll use the mass shift here, to remove the factor of $T_0$. The entropy shift is positive when the mass shift at constant charge is negative. It is easy to see that the shifts here are positive when all the coefficients are positive.

For AdS$_4$, we find:

$$T_0\Delta S = \frac{1}{5r_h l^2}\Big(4c_1(l^2+3r_h^2)^2 + 2c_2(l^2+3r_h^2)(l^2+18r_h^2) + 8(2c_3+c_4)(l^2+3r_h^2)^2\Big)\,. \tag{6.3.25}$$

For AdS$_5$, we get:

$$\begin{aligned} T_0\Delta S = \frac{\pi}{16l^2}\Big(&c_1(31l^4+128l^2r_h^2+138r_h^4) \\ &+ c_2 24(l^2+2r_h^2)(l^2+6r_h^2) + (2c_3+c_4)72(l^2+2r_h^2)^2\Big)\,. \end{aligned} \tag{6.3.26}$$

AdS$_6$:

$$T_0 \Delta S = \frac{2\pi}{99 l^2}\Big( c_1 r_h (369 l^4 + 1263 l^2 r_h^2 + 1124 r_h^4) \\ + c_2 4 r_h (3l^2 + 5 r_h^2)(27 l^2 + 100 r_h^2) + (2c_3 + c4) 96 r_h (3l^2 + 5r_h^2)^2 \Big) \,. \tag{6.3.27}$$

AdS$_7$:

$$T_0 \Delta S = \frac{\pi^2}{224 l^2}\Big( c_1 \, (1384 l^4 r_h^2 + 4236 l^2 r_h^4 + 3345 r_h^6) \\ + c_2 \, 40 (2l^2 + 3r_h^2)(16 l^2 + 45 r_h^2) + (2c_3 + c4) \, 800 (2l^2 + 3r_h^2)^2 \Big) \,. \tag{6.3.28}$$

## 6.4 Thermodynamics from the On-Shell Euclidean Action

The ultimate goal of this chapter is to determine the leading higher-derivative corrections to relations between certain global properties of black hole solutions. These relations are of a thermodynamic nature, and arise by taking various derivatives of the free-energy corresponding to the appropriate ensemble. As is well-known [196], the classical free-energy of a black hole can be calculated using the saddle-point approximation of the Euclidean path integral with appropriate boundary conditions. In the *Gibbs* or *grand canonical ensemble*, the appropriate quantity is the Gibbs free-energy, which may be calculated from the on-shell Euclidean action

$$\beta G(T, \Phi) = I_E[g^E_{\mu\nu}(T, \Phi) \, , A^E_\mu(T, \Phi)], \tag{6.4.1}$$

where $\beta = T^{-1}$, and $g^E_{\mu\nu}(T, \Phi)$ and $A^E_\mu(T, \Phi)$ are Euclideanized solutions to the classical equations of motion with temperature $T$ and potential $\Phi$. Similarly in the *canonical ensemble* the corresponding quantity is the Helmholtz free-energy, given by

$$\beta F(T, Q) = I_E[g^E_{\mu\nu}(T, Q) \, , A^E_\mu(T, Q)], \tag{6.4.2}$$

where $g^E_{\mu\nu}(T, Q)$ and $A^E_\mu(T, Q)$ are Euclideanized solutions with temperature $T$ and electric charge $Q$. In both expressions, $I_E$ is the *renormalized* Euclidean on-shell action.

The Euclidean action with cosmological constant is IR divergent when evaluated on a solution. However, it may be given a satisfactory finite definition by first regularizing the integral with a radial cutoff $R$. To render the variation principle well-defined on a spacetime with boundary we must add an appropriate set of Gibbons-Hawking-York (GHY) [197, 198] (in the case of the canonical ensemble, also Hawking-Ross [199]) terms *in addition to* a set of boundary counterterms. The complete on-shell action then consists of three contributions

$$I_E = I_{\text{bulk}} + I_{\text{GHY}} + I_{\text{CT}}. \tag{6.4.3}$$

If the counterterms are chosen correctly, they will cancel the divergence of the bulk and Gibbons-Hawking-York terms, rendering the results finite as $R \to \infty$. In AdS there is a systematic approach to generating such counterterms via the method of *holographic renormalization* [193,200, 201]; since the logic of this approach is well-described in detail elsewhere (see e.g. [202]) we will not review it further, but simply make use of known results. Explicit expressions for the needed GHY and counterterms (including the four-derivative corrections used in this chapter) valid in AdS$_d$, $d = 4, 5, 6$ can be found in [192, 203].

Once the free-energy is calculated, the remaining thermodynamic quantities can be determined straightforwardly by using the definitions of the free-energies and the first-law of black hole thermodynamics

$$F = E - TS, \quad G = E - TS - \Phi Q, \quad dE = TdS + \Phi dQ. \tag{6.4.4}$$

The expressions calculated using these Euclidean methods should agree with the Lorentzian or geometric calculations in the previous section. Note, however, that there is a bit of a subtlety with the notion of black hole mass here, as the thermodynamic relations are for the energy $E$ of the system. In holographic renormalization, there is always an ambiguity in the addition of finite counterterms that shift the value of the on-shell action. The standard approach is to fix the ambiguity by demanding that even-dimensional global AdS has zero vacuum energy while odd-dimensional global AdS has non-zero vacuum energy that is interpreted as a Casimir energy in the dual field theory. In this case the thermodynamic energy is the sum of the black hole mass and the Casimir energy

$$E = M + E_c, \tag{6.4.5}$$

and the mass $M$ of the black hole is only obtained after subtracting out the Casimir energy contribution, as we did in Section (6.2).

The purpose of introducing this alternative approach is not just to give a cross-check on the results of the previous section, but also to verify a recent general claim by Reall and Santos [204]. In this chapter, the $\mathcal{O}(\epsilon)$ corrections we are considering can be calculated by first evaluating the free-energy or on-shell action at the same order. Naively, this would require evaluating three contributions

$$\begin{aligned} I_E[g^E_{\mu\nu}, A^E_\mu] = I^{(2)}_E[g^{(2)E}_{\mu\nu}, A^{(2)E}_\mu] + \epsilon \left( \frac{\partial}{\partial \epsilon} I^{(2)}_E[g^{(2)E}_{\mu\nu} + \epsilon g^{(4)E}_{\mu\nu}, A^{(2)E}_\mu + \epsilon A^{(4)E}_\mu] \right)\bigg|_{\epsilon=0} \\ + \epsilon I^{(4)}_E[g^{(2)E}_{\mu\nu}, A^{(2)E}_\mu] + \mathcal{O}\left(\epsilon^2\right), \end{aligned} \tag{6.4.6}$$

where $(2)$ and $(4)$ denote two and four derivative terms in the action and their corresponding perturbative contributions to the solution. The central claim in [204] is that the first term at $\mathcal{O}(\epsilon)$ is actually *zero*, and that therefore we do not need to explicitly calculate the $\mathcal{O}(\epsilon)$ corrections to the equations of motion. For black hole solutions of the type considered in this chapter, we can evaluate the leading corrections without much difficulty, but for more general situations with less symmetry this may not be possible. In such a case the Euclidean method is more powerful, as has recently been demonstrated with calculation of corrections involving angular momentum [205] or dilaton couplings [206].

Although the result of [204] was demonstrated in the grand canonical ensemble, it is straightforward to see that it implies an identical claim about the leading corrections in the canonical ensemble. While the quantities of interest can be extracted from either, the explicit expressions encountered in the latter are usually far simpler and therefore more convenient. Recall that we can change ensemble by a Legendre transform of the free-energy

$$F(T,Q)=G(T,\Phi(Q))+\Phi(Q)Q, \quad Q=-\left(\frac{\partial G}{\partial \Phi}\right)_T, \tag{6.4.7}$$

where the right-hand-side is defined in terms of the implicit inverse function $\Phi(Q)$. At fixed $T$ and $Q$, the potential $\Phi$ receives corrections from the higher-derivative interactions, and so, expanding the right-hand-side to $\mathcal{O}(\epsilon)$, we have

$$\begin{aligned} F(T,Q)=G^{(2)}(T,\Phi^{(2)}(Q))&+\epsilon\left(\frac{\partial}{\partial \epsilon}G^{(2)}(T,\Phi^{(2)}(Q)+\epsilon\Phi^{(4)}(Q))\right)\Bigg|_{\epsilon=0} \\ &+\epsilon G^{(4)}(T,\Phi^{(2)}(Q))+\Phi^{(2)}(Q)Q+\epsilon\Phi^{(4)}(Q)Q+\mathcal{O}\left(\epsilon^2\right). \end{aligned} \tag{6.4.8}$$

Recognizing that

$$\left(\frac{\partial}{\partial \epsilon}G^{(2)}(T,\Phi^{(2)}(Q)+\epsilon\Phi^{(4)}(Q))\right)\Bigg|_{\epsilon=0}=\Phi^{(4)}(Q)\left(\frac{\partial G^{(2)}}{\partial \Phi}\right)_T\Bigg|_{\Phi=\Phi^{(2)}(Q)}=-\Phi^{(4)}(Q)Q, \tag{6.4.9}$$

we see that the leading correction to the Helmholtz free energy is simply given by

$$F(T,Q)=F^{(2)}(T,Q)+\epsilon G^{(4)}(T,\Phi^{(2)}(Q))+\mathcal{O}\left(\epsilon^2\right). \tag{6.4.10}$$

In terms of the on-shell Euclidean action, using the result of Reall and Santos, this is then equivalent to

$$F^{(4)}(T,Q)=\frac{1}{\beta}I_E^{(4)}\left(g_{\mu\nu}^{(2)E}(T,Q),A_\mu^{(2)E}(T,Q)\right), \tag{6.4.11}$$

where here $I_E^{(4)}$ denotes the contribution of the four-derivative terms to the renormalized on-shell action. Note that this includes potential four-derivative Gibbons-Hawking-York terms, but as this argument makes clear, will not include any additional Hawking-Ross terms. This expression is the analogue of the Reall-Santos result, but in the canonical ensemble. It says that the leading correction to the Helmholtz free-energy is given by evaluating the four-derivative part of the renormalized on-shell action on a solution to the two-derivative equations of motion with temperature $T$ and charge $Q$.

Below we will give a brief review of the well-known thermodynamic relations at two-derivative order, and then using the above result we will calculate the leading corrections and verify explicitly that they agree with the results of the previous section.

### 6.4.1 Two-Derivative Thermodynamics

As described above, the regularized on-shell action has a bulk as well as various boundary contributions. At two-derivative order and in $d$-dimensions these have the explicit form

$$\begin{aligned} I^{(2)}_{\text{bulk}} &= -\frac{1}{16\pi}\int d^{d+1}x\sqrt{g}\Big(\frac{d(d-1)}{l^2}+R-\frac{1}{4}F^2\Big), \\ I^{(2)}_{\text{GHY}} &= -\frac{1}{8\pi}\int d^dx\sqrt{h}K, \\ I^{(2)}_{\text{CT}} &= \frac{1}{8\pi}\int d^dx\sqrt{h}\left(\frac{d-1}{l}+\frac{l}{2(d-2)}\mathcal{R}\right), \end{aligned} \tag{6.4.12}$$

where $h_{ab}$ and $\mathcal{R}_{ab}$ are the metric and Ricci tensor of the induced geometry on the boundary at $r=R$. Note that in $I^{(2)}_{\text{CT}}$ we have included the minimal set of counterterms necessary to cancel the IR divergence in $d=3$ and $d=4$. For $d>4$, additional counterterms beginning at quadratic order in the boundary Riemann tensor are necessary to cancel further divergences.

The regularized bulk action has a well-defined variational principle provided that $\delta A_a=0$ at $r=R$. This amounts to holding $\Phi$ fixed, and thus it corresponds to boundary conditions compatible with the grand canonical ensemble. For many applications, we will want to hold the charge fixed. From a thermodynamic point of view, we want to use the extensive quantity $Q$ instead of the intensive $\Phi$, so we must compute the Helmholtz free energy instead of the Gibbs free energy. Holding $Q$ fixed requires different boundary conditions, and in particular the further addition of a Hawking-Ross boundary term [199]

$$I^{(2)}_{\text{HR}} = \frac{1}{16\pi}\int d^dx\sqrt{h}n_\mu F^{\mu b}A_b\,, \tag{6.4.13}$$

where $n_\mu$ is the normal vector on the boundary and $A_a$ is the pull-back of the gauge potential. To summarize, the total two-derivative on-shell action

$$I^{(2)}_E = I^{(2)}_{\text{bulk}}+I^{(2)}_{\text{GHY}}+I^{(2)}_{\text{HR}}+I^{(2)}_{\text{CT}}, \tag{6.4.14}$$

evaluated on the Euclideanized solution to the two-derivative equations of motion

$$
\begin{aligned}
ds_E^2 =& f(r)d\tau^2 + g(r)^{-1}dr^2 + r^2 d\Omega_{d-1}^2\,, \qquad f(r) = g(r) = 1 - \frac{m}{r^{d-2}} + \frac{q^2}{4r^{2d-4}} + \frac{r^2}{l^2}\,,\\
A_E =& i\left(-\frac{1}{c}\frac{q}{r^{d-2}} + \Phi\right)d\tau, \qquad c = \sqrt{\frac{2(d-2)}{(d-1)}}, \qquad \Phi = \frac{1}{c}\frac{q}{l^{d-2}\nu^{d-2}}\,,
\end{aligned}
\tag{6.4.15}
$$

is equal to $\beta F^{(2)}(T,Q)$, where $F^{(2)}$ is the two-derivative contribution to the Helmholtz free-energy. In the above we have introduced the dimensionless variable $\nu \equiv (r_h)_0/l$, where $(r_h)_0$ is the location of the outer-horizon of the two-derivative solution with temperature $T$ and charge $Q$. Note also that here, and for the remainder of this section, we will consider only spherical $k = 1$ black holes. Since $\nu$ satisfies $f(\nu) = 0$, we can solve for the parameter $m$ as

$$
m = \nu^{d-2} + \frac{q^2}{4\nu^{d-2}} + \frac{\nu^d}{l^2}. \tag{6.4.16}
$$

In the Euclidean approach to calculating the leading corrections to the thermodynamics, it will prove natural to continue to use $\nu$ and $q$ to parametrize the space of black hole solutions, even when the four-derivative corrections are included. This means that it is also natural to write all thermodynamic quantities in these variables, which requires the use of standard thermodynamic derivative identities to rewrite derivatives. Recall that the parameter $q$ and the physical charge $Q$ are not the same, but are related by an overall constant given in (6.2.21). Therefore holding $Q$ fixed is the same as holding $q$ fixed. Explicitly, the two-derivative free-energy calculated in this way in AdS$_4$ is given by

$$
F_{d=3}^{(2)}(q,\nu) = -\frac{l\nu^3}{4} + \frac{l\nu}{4} + \frac{3q^2}{16l\nu}, \tag{6.4.17}
$$

and in AdS$_5$ by

$$
F_{d=4}^{(2)}(q,\nu) = -\frac{1}{8}\pi l^2\nu^4 + \frac{1}{8}\pi l^2\nu^2 + \frac{5\pi q^2}{32l^2\nu^2} + \frac{3\pi l^2}{32}. \tag{6.4.18}
$$

Once the free-energy is calculated, the entropy and energy are given by

$$
S = -\left(\frac{\partial F}{\partial T}\right)_Q, \qquad E = F + TS. \tag{6.4.19}
$$

In terms of our natural variables, we can reexpress the entropy as

$$
S(q,\nu) = \left(\frac{\partial F}{\partial \nu}\right)_q \left[\left(\frac{\partial T}{\partial \nu}\right)_q\right]^{-1}, \tag{6.4.20}
$$

where the temperature is given by

$$T(q,\nu)=\frac{(d-2)q^2l^{1-d}\nu^{1-d}}{4\pi}+\frac{(d-1)\nu^2+d-2}{4\pi l}. \tag{6.4.21}$$

Note that this expression is exact, meaning it does not receive corrections when we include the four-derivative interactions. It is therefore useful to introduce the function

$$q_{\text{ext}}^2(\nu)=-\frac{2\left(d\nu^2+d-\nu^2-2\right)(l\nu)^{d-2}}{(d-2)}, \tag{6.4.22}$$

such that taking the limit $q^2\to q_{\text{ext}}^2(\nu)$ is equivalent to taking the extremal limit $T\to 0$.

If we extract the energy $E=F+TS$ from the expressions (6.4.17) and (6.4.18), we find that it agrees with the mass, (6.2.23), for $\text{AdS}_4$ but not $\text{AdS}_5$. This is not surprising as the thermodynamic energy $E$ and mass $M$ of the black hole in $\text{AdS}_5$ differ by a Casimir energy contribution that is independent of $q$ and $\nu$. We can, of course remove the Casimir energy by the addition of *finite boundary counterterms*, or equivalently by a change in holographic renormalization scheme. The expression (6.4.18) is calculated in a *minimal subtraction* scheme, in which the possible finite counterterms are zero and the Casimir energy is present.

Physically, it is useful work in a scheme in which the energy $E$ coincides with the mass $M$ of the black hole, without a Casimir contribution. In such a *zero Casimir* scheme, the energy of pure $\text{AdS}_5$ is defined to be zero. Calculating the free-energy from the on-shell action of pure $\text{AdS}_5$ with generically parametrized four-derivative counterterms we find that this scheme requires the following modification from the minimal subtraction counterterms

$$I_{\text{CT}}^{(2)}\longrightarrow I_{\text{CT}}^{(2)}+\frac{1}{8\pi}\int d^4x\sqrt{h}\left(-\frac{l^3}{96}\right)\mathcal{R}^2. \tag{6.4.23}$$

The free energy calculated with this modified on-shell action agrees exactly with the expectation using (6.2.23). Note that the entropy, since it is given by a derivative of the free-energy, is independent of the choice of scheme. The zero Casimir scheme is a physically motivated choice, but certainly not unique.

### 6.4.2 Four-Derivative Corrections to Thermodynamics

To evaluate the four-derivative corrections we make use of the result (6.4.11). As in the two-derivative contribution, the on-shell action is properly defined by a regularization and renormalization procedure. For the operators in (6.2.1) with Wilson coefficients $c_2$, $c_3$ and $c_4$ the required $I_{\text{bulk}}^{(4)}$ contribution is actually finite, while for the term in (6.2.1) proportional to $c_1$, we must again regularize and renormalize by adding infinite boundary counterterms. The required explicit expressions, as well as the complete set of four-derivative GHY terms, can be found in [192, 203]. The calculation is otherwise identical to the two-derivative contribution described above, and in $\mathrm{AdS}_4$ we find

$$\begin{aligned}F^{(4)}_{d=3}(q,\nu) =& c_1\left(-\frac{(20l^4\nu^4-5l^2\nu^2q^2+q^4)}{20l^5\nu^5}-\frac{3\nu}{l}\right)+\frac{c_2q^2\left(l^2\left(20l^2\nu^2-7q^2\right)-60l^4\nu^4\right)}{80l^7\nu^5}\\&-\frac{c_3q^4}{5l^5\nu^5}-\frac{c_4q^4}{10l^5\nu^5}.\end{aligned}\tag{6.4.24}$$

The complete free-energy, up to $\mathcal{O}(\epsilon^2)$ contributions, is then given by

$$F_{d=3}(q,\nu)=F^{(2)}_{d=3}(q,\nu)+\epsilon F^{(4)}_{d=3}(q,\nu)+\mathcal{O}(\epsilon^2).\tag{6.4.25}$$

From this explicit expression we can then calculate the entropy

$$\begin{aligned}S_{d=3}=\pi l^2\nu^2&-\frac{4\pi c_1\epsilon\left(4l^4\nu^4\left(1-3\nu^2\right)-3l^2\nu^2q^2+q^4\right)}{4l^2\left(3\nu^2-1\right)\nu^4+3\nu^2q^2}-\frac{\pi c_2q^2\epsilon\left(12l^2\nu^2\left(\nu^2-1\right)+7q^2\right)}{4l^2\left(3\nu^2-1\right)\nu^4+3\nu^2q^2}\\&-\frac{16\pi c_3q^4\epsilon}{4l^2\left(3\nu^2-1\right)\nu^4+3\nu^2q^2}-\frac{8\pi c_4q^4\epsilon}{4l^2\left(3\nu^2-1\right)\nu^4+3\nu^2q^2}+\mathcal{O}(\epsilon^2),\end{aligned}\tag{6.4.26}$$

and mass (which coincides with the thermal energy)

$$\begin{aligned}M_{d=3}=&\frac{1}{2}l\left(\nu^3+\nu\right)+\frac{q^2}{8l\nu}+\frac{c_1q^4\epsilon\left(q^2-4l^2\nu^2\left(9\nu^2+2\right)\right)}{40l^5\left(3\nu^2-1\right)\nu^7+30l^3\nu^5q^2}\\&+\frac{c_2q^2\epsilon\left(80l^4\nu^4\left(-9\nu^4+6\nu^2+1\right)-8l^2\nu^2\left(39\nu^2+7\right)q^2+7q^4\right)}{40l^3\nu^5\left(4l^2\nu^2\left(3\nu^2-1\right)+3q^2\right)}\\&+\frac{2c_3q^4\epsilon\left(q^2-4l^2\nu^2\left(9\nu^2+2\right)\right)}{5l^3\nu^5\left(4l^2\nu^2\left(3\nu^2-1\right)+3q^2\right)}+\frac{c_4q^4\epsilon\left(q^2-4l^2\nu^2\left(9\nu^2+2\right)\right)}{5l^3\nu^5\left(4l^2\nu^2\left(3\nu^2-1\right)+3q^2\right)}+\mathcal{O}(\epsilon^2).\end{aligned}\tag{6.4.27}$$

Taking the extremal limit we find the following expression for the mass shift

$$\begin{aligned}(\Delta M_{d=3})_{Q,T=0}=&-\frac{4c_1l\left(3\nu^2+1\right)^2}{5\nu}-\frac{2c_2l(3\nu^2+1)(18\nu^2+1)}{5\nu}\\&-\frac{16c_3l\left(3\nu^2+1\right)^2}{5\nu}-\frac{8c_4l\left(3\nu^2+1\right)^2}{5\nu},\end{aligned}\tag{6.4.28}$$

which agrees exactly with (6.3.25). Strictly, the two expressions are parameterized in terms of different variables ($\nu$ the uncorrected horizon vs. $r_h$ the corrected horizon), but these differ by $\mathcal{O}(\epsilon)$, and so when we take $\epsilon\to 0$ the two functions are the same.

Similarly we can calculate the shift in the microcanonical entropy, which will be important in the subsequent section for analyzing conjectured bounds on the Wilson coefficients. The actual expression is given in (6.5.10), and can be calculated straightforwardly using standard thermodynamic derivative identities

$$(\Delta S)_{Q,E} = \lim_{\epsilon\to 0}\left[\left(\frac{\partial S}{\partial \epsilon}\right)_{q,\nu} - \left(\frac{\partial E}{\partial \epsilon}\right)_{q,\nu}\frac{\left(\frac{\partial S}{\partial \nu}\right)_q}{\left(\frac{\partial E}{\partial \nu}\right)_q}\right]. \tag{6.4.29}$$

The calculation for AdS$_5$ is similar, but in this case we have to be cautious about the Casimir energy. We calculate the free-energy in the physically motivated zero Casimir scheme. To do so, we again fix the finite counterterms by evaluating the four-derivative on-shell action on pure AdS$_5$. Requiring the Casimir energy to vanish requires the following modification from the minimal subtraction counterterm action

$$I_{\text{CT}}^{(4)} \longrightarrow I_{\text{CT}}^{(4)} + \frac{1}{8\pi}\int d^4x\sqrt{h}\left(-\frac{5c_1 l^3}{48}\right)\mathcal{R}^2. \tag{6.4.30}$$

Using this we calculate the four-derivative contribution to the renormalized free-energy

$$\begin{aligned} F_{d=4}^{(4)} =& \frac{1}{256}\pi c_1\left(-\frac{43q^4}{l^8\nu^8} + \frac{24\left(5\nu^2+8\right)q^2}{l^4\nu^4} - 32\left(13\nu^4+41\nu^2+18\right)\right) \\ &+ \frac{3\pi c_2\left(8l^4\nu^4q^2-3q^4\right)}{32l^8\nu^8} - \frac{9\pi c_3 q^4}{16l^8\nu^8} - \frac{9\pi c_4 q^4}{32 l^8\nu^8}. \end{aligned} \tag{6.4.31}$$

We also obtain the entropy

$$\begin{aligned} S_{d=4} =& \frac{1}{2}\pi^2 l^3\nu^3 + \frac{\pi^2 c_1\epsilon\left(8l^8\left(26\nu^2+41\right)\nu^{10} + 6l^4\left(5\nu^2+16\right)\nu^4q^2 - 43q^4\right)}{4l^3\nu^3\left(4l^4\left(2\nu^2-1\right)\nu^4+5q^2\right)} \\ &+ \frac{6\pi^2 c_2\epsilon\left(4l^4\nu^4q^2-3q^4\right)}{l^7\left(8\nu^9-4\nu^7\right)+5l^3\nu^3q^2} - \frac{36\pi^2 c_3 q^4\epsilon}{l^7\left(8\nu^9-4\nu^7\right)+5l^3\nu^3q^2} \\ &- \frac{18\pi^2 c_4 q^4\epsilon}{l^7\left(8\nu^9-4\nu^7\right)+5l^3\nu^3q^2} + \mathcal{O}\left(\epsilon^2\right), \end{aligned} \tag{6.4.32}$$

and mass

$$\begin{aligned} M_{d=4} =& \frac{3\pi\left(4l^4\left(\nu^2+1\right)\nu^4+q^2\right)}{32l^2\nu^2} \\ &+ c_1\left[\frac{\pi\epsilon\left(384l^{12}\left(\nu^2+1\right)\left(26\nu^4+23\nu^2+6\right)\nu^{12} - 32l^8\left(27\nu^4+32\nu^2+18\right)\nu^8q^2\right)}{256l^8\nu^8\left(4l^4\left(2\nu^2-1\right)\nu^4+5q^2\right)}\right. \\ &\qquad \left. + \frac{\pi\epsilon\left(-4l^4\left(684\nu^2+253\right)\nu^4q^4+129q^6\right)}{256l^8\nu^8\left(4l^4\left(2\nu^2-1\right)\nu^4+5q^2\right)}\right] \\ &+ \frac{3\pi c_2 q^2\epsilon\left(32l^8\left(10\nu^2+3\right)\nu^8 - 4l^4\left(54\nu^2+19\right)\nu^4q^2+9q^4\right)}{32l^8\nu^8\left(4l^4\left(2\nu^2-1\right)\nu^4+5q^2\right)} \\ &+ \frac{9\pi c_3 q^4\epsilon\left(3q^2-4l^4\nu^4\left(18\nu^2+7\right)\right)}{16l^8\nu^8\left(4l^4\left(2\nu^2-1\right)\nu^4+5q^2\right)} \\ &+ \frac{9\pi c_4 q^4\epsilon\left(3q^2-4l^4\nu^4\left(18\nu^2+7\right)\right)}{32l^8\nu^8\left(4l^4\left(2\nu^2-1\right)\nu^4+5q^2\right)} + \mathcal{O}\left(\epsilon^2\right). \end{aligned} \tag{6.4.33}$$

The extremal mass shift is given by

$$(\Delta M_{d=4})_{Q,T=0} = -\frac{1}{16}\pi c_1\left(138\nu^4+128\nu^2+31\right)-\frac{3}{2}\pi c_2\left(2\nu^2+1\right)\left(6\nu^2+1\right) \\ -9\pi c_3\left(2\nu^2+1\right)^2-\frac{9}{2}\pi c_4\left(2\nu^2+1\right)^2\,, \tag{6.4.34}$$

which agrees exactly with the result (6.3.26). Likewise we can calculate the correction to the microcanonical entropy using (6.4.29), the explicit expressions are given in Appendix K.

## 6.5 Constraints From Positivity of the Entropy Shift

Having derived the general entropy shift at fixed mass, we may now determine what constraints on the EFT coefficients are implied by the assumption that it is positive. Recall that the argument of [49] for the positivity of the entropy shift assumes the existence of a number of quantum fields $\phi$ with mass $m_\phi$, heavy enough so that they can be safely integrated out. In particular, such fields are assumed to couple to the graviton and photon in such a way that, after being integrated out, they generate *at tree-level* the higher-dimension operators we are considering (with the corresponding operator coefficients scaling as $c_i \sim 1/m_\phi$). This assumption is essential to the proof; it may be that the entropy shift is universally positive (see [205] for a number of examples), but proving such a statement for non-tree-level completions would require a different argument from the one laid out here.

We revisit the logic of [49] in the context of flat space, before discussing how it may be extended to AdS asymptotics, and denote the Euclidean on-shell action of the theory that includes the heavy scalars $\phi$ by $I_{\rm UV}[g, A, \phi]$. First, note that when the scalars are set to zero and are non-dynamical, the action reduces to that of the pure Einstein-Maxwell theory,

$$I_{\rm UV}[g, A, 0] = I^{(2)}[g, A]\,. \tag{6.5.1}$$

This is a statement relating the value of the functionals $I_{\rm UV}$ and $I^{(2)}$ (the two-derivative action) when we pick particular configurations for the fields. These fields may or may not be solutions to the equations of motion. Next, consider the corrected action, $I_C = I^{(2)} + I^{(4)}$, and note that it obeys

$$I_C[g+\Delta g, A+\Delta A] \simeq I_{\rm UV}[g, A, \phi]\,. \tag{6.5.2}$$

Here we have in mind that the fields are valid solutions of the respective theories, *i.e.* $[g, A, \phi]$ is a solution of the UV theory and $[g+\Delta g, A+\Delta A]$ is a solution to the four-derivative corrected theory. The UV theory and that with an infinite series of higher-derivative corrections should have exactly the same partition function; therefore, this expression is an equality up to quantum corrections and corrections that are $\mathcal{O}(\epsilon^2)$. Finally, let us choose $[g, A, \phi]$ to be solutions of the UV theory with charge $Q$ and temperature $T$, and $[g_0, A_0]$ to be field configurations in the pure Einstein-Maxwell theory *with the same charge and temperature* as those of the UV theory. One then finds the following inequality,

$$I_C[g + \Delta g, A + \Delta A]_{T,Q} \simeq I_{\text{UV}}[g, A, \phi]_{T,Q} < I_{\text{UV}}[g_0, A_0, 0]_{T,Q} = I^{(2)}[g_0, A_0]_{T,Q} \,. \tag{6.5.3}$$

Since $[g, A, \phi]$ is a solution of the UV theory, it extremizes the action. To ensure the inequality that appears in (6.5.3), one must further require that this solution is a *minimum* of the action. The inequality then follows because $[g_0, A_0, 0]$ is *not* a solution to the equations of motion, for the same charge and temperature. Finally, as long as one works in the same ensemble, the boundary terms will be the same for both actions and thus don't affect the argument.

In general, different theories will have different relationships between mass, charge, and temperature. We are interested in the entropy shift at fixed mass and charge. Therefore we must compare the two action functionals at different temperatures. For simplicity, we use $T_4/T_2$ for the temperature that corresponds to mass $M$ and charge $Q$ for the theory with/without higher-derivative corrections, respectively. Then we have:

$$\begin{aligned}
F_C(Q, T_4) &< F_2(Q, T_4), \\
F_C(Q, T_4) &< F_2(Q, T_2) + (T_4 - T_2)\partial_T F_2(Q, T_2), \\
F_C(Q, T_4) &< F_2(Q, T_2) - (T_4 - T_2)S_2, \\
M - S_4 T_4 &< M - S_2 T_2 - (T_4 - T_2)S_2, \\
-S_4 T_4 &< -T_4 S_2, \\
\Delta S &> 0,
\end{aligned} \tag{6.5.4}$$

at fixed $M$ and $Q$ (and in the zero Casimir energy scheme).

Now that we have outlined the argument in flat space, we can ask whether it can be immediately extended to AdS. One subtle point in the derivation outlined above is that the free-energy is only finite after the subtraction of the free-energy of a reference background. In the flat space context, the contributions of such terms to the two actions are identical because the asymptotic charges are the same. Thus, this issue does not affect the validity of the argument.

In AdS, the story is a little different– the free-energy is computed using holographic renormalization. Different counterterms are required to render the two-derivative action $I^{(2)}$ and the corrected action $I_C$ finite. Moreover, $I_{UV}$ may also require a different set of counterterms involving contributions from the scalar, and unlike the bulk contribution, there is no reason to expect that their on-shell values are less than their off-shell values. This is a potential hole in the positivity argument in AdS. Apart from this issue, the rest of the argument can be immediately applied to AdS.

### 6.5.1 Thermodynamic Stability

As we've seen, the above proof requires that the uncorrected backgrounds are minima of the action. Thermodynamically, this amounts to the condition that the black holes are stable under thermal and electrical fluctuations. This translates to the following requirements on the free-energies,

$$\left(\frac{\partial^2 F}{\partial T^2}\right)_Q \leq 0, \qquad \left(\frac{\partial^2 G}{\partial T^2}\right)_\Phi \leq 0, \qquad \epsilon_T = \left(\frac{\partial^2 F}{\partial Q^2}\right)_T \geq 0\,. \tag{6.5.5}$$

These conditions may be rewritten in terms of the specific heat and permittivity of the black hole, which can be used to determine, respectively, the thermal stability and electrical stability of the black hole [207,208]. We'll ignore the specific heat at constant $\Phi$ now, as we are interested in the stability in the canonical ensemble, and consider

$$C_Q = T\left(\frac{\partial S}{\partial T}\right)_Q \geq 0, \qquad \epsilon_T = \left(\frac{\partial Q}{\partial \Phi}\right)_T \geq 0\,. \tag{6.5.6}$$

Positivity of the specific heat is equivalent to the statement that larger black holes should heat up and radiate more, while smaller ones should become colder and radiate less. When the quantity $\epsilon_T$ is negative the black hole is unstable to electrical fluctuations, meaning that when more charge is placed into it, its chemical potential *decreases*. We expect that it should instead increase, to make it more difficult to move a charge from outside to inside the black hole – thus making it harder to move away from equilibrium [208]. We may compute these quantities using the results of the previous section. For AdS$_4$, we find

$$C_Q = \frac{2\pi l^2\nu^2(1+3\nu^2)(2-\xi)\xi}{2-6\xi+3\xi^2+3\nu^2(4-6\xi+3\xi^2)}, \qquad \epsilon_T = \frac{(\xi-2)\xi+3\nu^2(2-2\xi+\xi^2)}{\nu l\left(2-6\xi+3\xi^2+3\nu^2(4-6\xi+3\xi^2)\right)}\,, \tag{6.5.7}$$

where we recall that $\nu = r_h/l$ and $Q = (1-\xi)Q_{\text{ext}}$. These results have been obtained previously *e.g.* in [209]. We find that both of these coefficients are positive when either

$$\nu < \nu^* = \frac{1}{\sqrt{3}}\,, \qquad \xi < \xi^* = 1 - \sqrt{\frac{1-3\nu^2}{1+3\nu^2}}\,, \tag{6.5.8}$$

holds, or when

$$\nu > \nu^* = \frac{1}{\sqrt{3}}\,, \qquad 0 < \xi < 1\,, \tag{6.5.9}$$

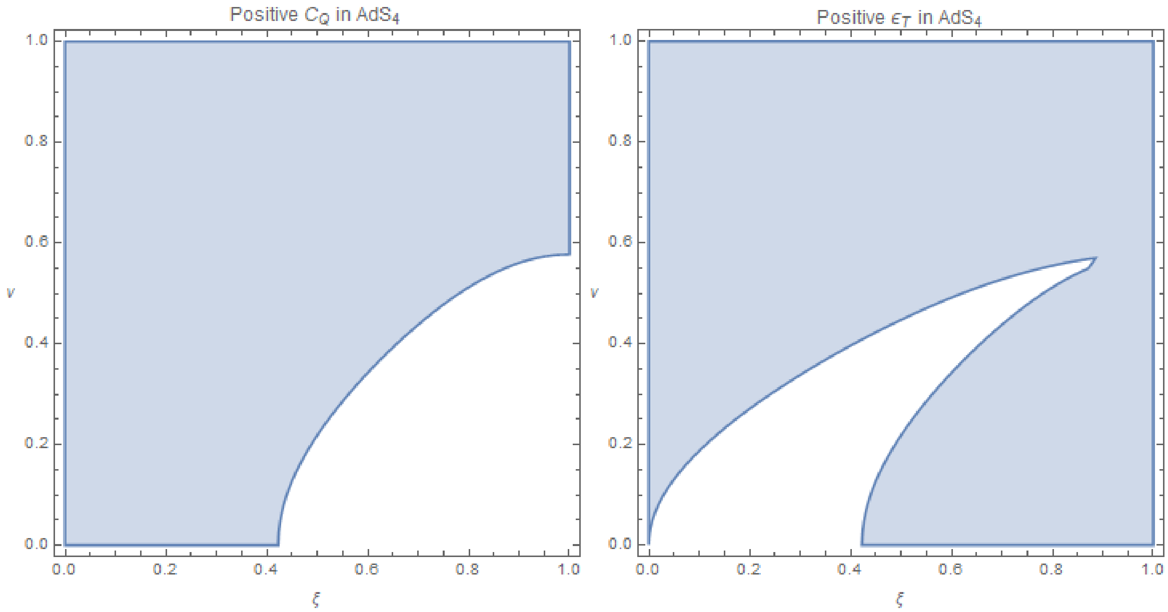



Figure 6.1: Blue represents the regions of parameter space where each quantity is positive.

is satisfied.

Thus, for small black holes stability requires that the extremality parameter be less than some function of the radius, $\xi < \xi^*$. In particular, extremal black holes, for which $\xi \to 0$, are stable while neutral black holes, which correspond to $\xi \to 1$, are not. The implication of (6.5.9) is that above a certain radius ($r_h > l/\sqrt{3}$) all black holes are thermodynamically stable. This behavior is visible from Fig. 6.1, where we have plotted the allowed parameter space based on the $C_Q$ and $\epsilon_T$ conditions separately. This raises an interesting point in making contact with the flat space limit: if we require both parameters to be positive, there are no stable black holes at $\nu = 0$. Note that in [49] only $C_Q$ was considered. However, in applications involving AdS/CFT, we believe that both the specific heat and electrical permittivity should be taken into account.

Here we have only considered the leading-order stability. The higher-derivative corrections will shift the point where the specific heat crosses from positive to negative. However, in proving the extremality-entropy relation, we are only interested in the extremal limit, which is not affected by this consideration. In principal we could compute the order $\epsilon$ shifts to the stability conditions to obtain small corrections to the entropy bounds.

### 6.5.2 Constraints on the EFT Coefficients

The entropy shift in AdS$_4$ for a black hole with an *arbitrary size and charge* takes the following form,

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{Q,M} = \frac{l(1+3\nu^2)}{5\nu T}\Big(c_1\left(4-6\xi+19\xi^2-16\xi^3+4\xi^4+12\nu^2(\xi-1)^4\right)$$
$$+c_2(\xi-1)^2\left(2-14\xi+7\xi^2+3\nu^2(12-14\xi+7\xi^2)\right)+8(2c_3+c_4)(1+3\nu^2)(\xi-1)^4\Big)\,, \tag{6.5.10}$$

where the temperature is given by the expression

$$T(r_h,\xi) = -\frac{(1+3\nu^2)(\xi-2)\xi}{4\pi\nu l}\,.$$

We can see from the $\xi$ dependence of the latter that in the $\xi \to 0$ limit the shift to the entropy blows up. If we examine the leading part in $1/\xi$, we find that it is proportional to the mass shifts we have computed above. Thus, in the extremal limit we have

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{\xi\to 0} = \frac{l^2}{5r_hT}\Big(4c_1(1+3\nu^2)^2+2c_2(1+3\nu^2)(1+18\nu^2)+8(2c_3+c4)(1+3\nu^2)^2\Big)\,. \tag{6.5.11}$$

It is also interesting to note that in the chargeless limit $\xi \to 1$ the dependence of (6.5.10) on $c_2, c_3$ and $c_4$ drops out entirely, and we are left with an entropy shift of the simple form

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{\xi\to 1} = \frac{l}{\nu T}c_1\left(1+3\nu^2\right)\,. \tag{6.5.12}$$

Our results above show that the large black holes are stable in the chargeless limit, which implies that the $c_1$ coefficient must be positive.

In Fig. 6.2, we have graphed the constraints on the coefficients that arise from demanding that the entropy shift is positive. We have included both the constraints from the extremal entropy shift and from considering the shift of all stable black holes. Considering only extremal black holes may be interesting because it is equivalent to the condition that the extremality shift, $\Delta(M/Q)$, is negative. Thus we may look at the constraints implied by positive entropy shift and by negative extremality shift independently. Note that we have divided by $c_1$, which we have already proven to be positive. We may write out the all the constraints obtained:

$$\begin{aligned} c_1 &\geq 0, \\ c_2 &\geq 0, \\ c_3 &\geq -\frac{1}{8}c_1(2+c_2). \end{aligned} \tag{6.5.13}$$

We have computed the corresponding bounds for $\mathrm{AdS}_5$ through $\mathrm{AdS}_7$. The results may be found in Appendix K. We would, however, like to comment on $\mathrm{AdS}_5$, where the positivity of the coefficient of the Riemann-squared term is of particular interest. The stability analysis yields results that are qualitatively similar to (6.5.8) and (6.5.9), but with the following definitions

$$\xi^* = 1 - \sqrt{\frac{1-2\nu^2}{1+2\nu^2}}, \qquad \nu^* = \frac{1}{\sqrt{2}}\,. \tag{6.5.14}$$

Once again, we see that large black holes are stable for all values of the charge.

When we examine the entropy shift in the neutral limit, we find

$$\frac{\pi l^2}{32T}c_1\left(87+164\nu^2+52\nu^4\right)\,, \tag{6.5.15}$$

whose overall sign is completely determined by that of $c_1$. This means that there are stable black holes where the sign of the entropy shift is the same as the sign of the coefficient of $R^2_{abcd}$. Thus, a positive entropy shift for stable black holes implies that $c_1$ is positive. In fact, a positive value of $c_1$ was the necessary ingredient in [210] for obtaining the violation of the KSS bound[4]. It is also interesting to note that in $d > 3$, this sign constraint was shown to follow from the assumption of a unitary tree-level UV completion [211]. The entropy constraints given in this chapter are then strictly stronger since they also apply in $d = 3$.

In closing, we stress that we are not claiming that the entropy shift should be universally positive; the proof outlined above only applies when the higher-derivative corrections are generated by integrating out massive fields at tree-level (and relies on assuming that the corresponding solutions minimize the effective action). However, it is interesting that the conjecture that the entropy shift is universally positive appears to suggest that violations of the KSS bound are required to occur. Our results extend and make more precise the earlier claim by some of us [192] of a link between the WGC and the violation of the KSS bound. We will come back to this point in Section 6.6.

[4]We have checked the calculation with a different basis, choosing to use Gauss-Bonnet instead of Riemann squared. As expected, we find that the coefficient of the Gauss-Bonnet term is positive.

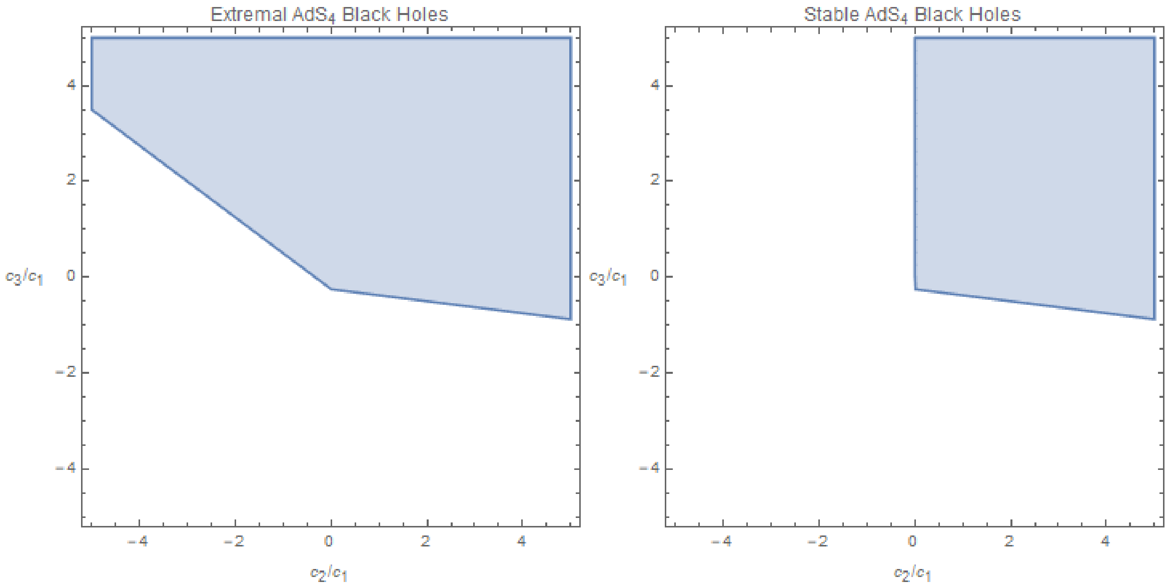



Figure 6.2: Blue regions are allowed after imposing that the entropy shift is positive. (Left): Allowed region after imposing that extremal black holes have positive entropy shift (Right): Allowed region after imposing that all stable black holes have positive entropy shift

### 6.5.3 Flat Space Limit

As we've pointed out above, we can not compare the results we have given above to the flat space limit. This is because if we impose both $C_Q > 0$ and $\epsilon_T > 0$, we find that there are no stable black holes in the flat space limit $\nu \to 0$ (as suggested by figure 6.1). In AdS/CFT, we expect that both conditions are necessary to ensure thermodynamic stability; nonethless, we may remove the condition $\epsilon_T > 0$ in order to compare with the flat space limit. In this case, we find that stability requires

$$\xi^* = 1 - \frac{1}{\sqrt{3}}\sqrt{\frac{1-3\nu^2}{1+3\nu^2}}, \qquad \nu^* = \frac{1}{\sqrt{3}}, \tag{6.5.16}$$

for the AdS$_4$ black holes, and

$$\xi^* = 1 - \frac{1}{\sqrt{2}}\sqrt{\frac{1-2\nu^2}{1+2\nu^2}}, \qquad \nu^* = \frac{1}{\sqrt{2}}, \tag{6.5.17}$$

for the AdS$_5$ black holes. This allows for a more direct comparison between the two cases. In figure 6.3, we contrast the bounds obtained in AdS and flat space. The bounds in AdS are stronger, as they should be given that there is an extra parameter's worth of stable black holes. Note also that $c_1 > 0$ is implied by positivity in AdS, but not in flat space, because in flat space there are no stable neutral black holes.

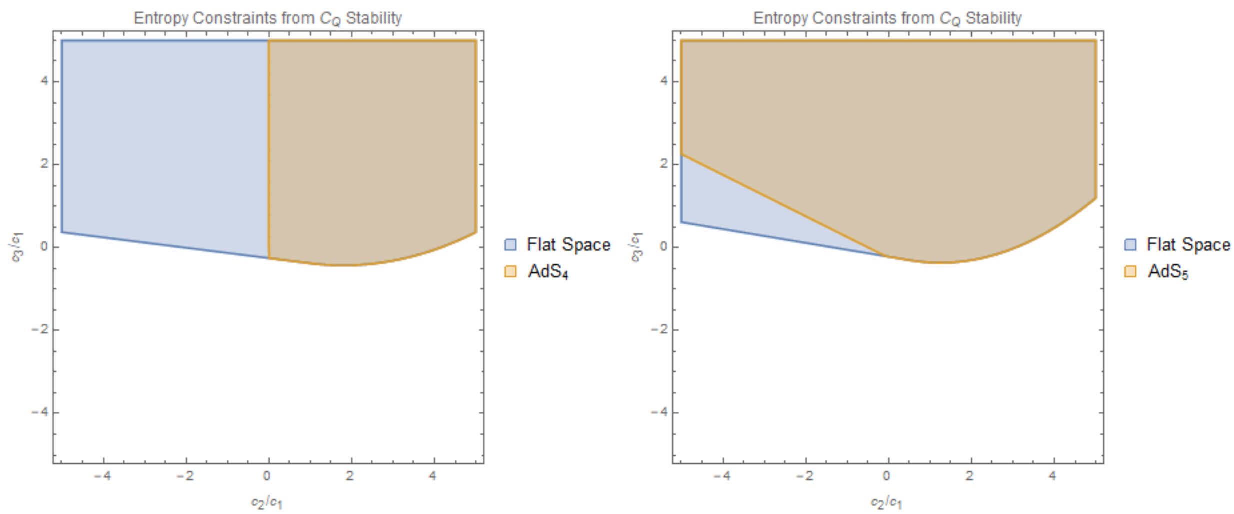



Figure 6.3: The blue regions are allowed in flat space and the orange in AdS– note that the AdS regions are a subset of those from flat space.

## 6.6 Holography and the Shear Viscosity to Entropy Ratio

In this chapter, we have examined the relationship between the higher-derivative corrections to entropy and extremality in Anti-de Sitter space. As we have seen, extremality is considerably more complicated in AdS because the relationship between mass, charge, and horizon radius at extremality is non-linear. Nonetheless, we have verified the relation [49,186] between the entropy shift at fixed charge and mass and the extremality shift at fixed charge and temperature. There is a sharp dependence on which quantities are held fixed in AdS. This is in contrast to flat space, where the linear relationship between mass, charge, and horizon radius removes this issue. We have also provided a more general proof of this relation in Appendix L, and extended the result to show that there is a third proportional quantity, which is the extremality shift at fixed mass and temperature.

When viewed geometrically, these statements seem almost accidental. In Section (6.4), we performed the same calculation from a thermodynamic point of view by computing the free energy from the renormalized on-shell action. From this point of view, issues concerning "which quantity is held fixed" translate to "which ensemble is used." In addition to providing an additional check on the results from Section (6.3), this provides a non-trivial confirmation of the results of [204], which states that the shifted geometry is not needed to compute the thermodynamic quantities.

Assuming that the entropy shift is positive places constraints on the Wilson coefficients. However, a crucial difference appears in AdS when compared to flat space. The stability criterion depends on the horizon radius over the AdS length, and goes to zero at large horizon radius. This means that there are stable neutral black holes that are asymptotically AdS. For neutral black holes, the entropy shift is dominated by $c_1$, which is the coefficient of the Riemann squared term, so the positivity of the entropy shift implies the positivity of this coefficient. In AdS$_5$, this coefficient may be related to the central charges of the dual field theory [200, 212, 213] by

$$c_1 = \frac{1}{8}\frac{c-a}{c}\,. \tag{6.6.1}$$

Thus, the positivity of the entropy shift appears to be violated in theories where $c - a < 0$. In [214], a number of superconformal field theories were examined, and all were found to satisfy $c - a > 0$. It is worth noting there are non-interacting theories where $c - a < 0$; for example, $\frac{a}{c} = \frac{31}{18}$ for a free theory of only vector fields [215]. However, such theories do not have weakly curved gravity duals. If there are any bulk theories where $c_1 < 0$, we are not aware of them. The question of whether holographic theories necessarily correspond to $c-a$ non-negative is interesting for a number of reasons – both from a fundamental point of view and for phenomenological applications.

In particular, recall that the range of the Wilson coefficients and the sign of $c - a$ played an important role in the physics of the shear viscosity to entropy ratio $\eta/s$ and how it deviates from its universal $1/4\pi$ result [216, 217], as discussed extensively in the literature (see [218] for a review of the status of the shear viscosity to entropy bound). Indeed, it is interesting to compare our results to the higher-derivative corrections to $\eta/s$, which (for the AdS$_5$ case of interest to us here) were shown [191] to be given by

$$\frac{\eta}{s} = \frac{1}{4\pi}\left(1 - 8c_1 + 4(c_1 + c_2)\frac{q^2}{r_0^6}\right)\,, \tag{6.6.2}$$

where $r_0$ is a parameter of the solution defined in [191]; the factor $q^2/r_0^6$ goes from 0 (for neutral black holes) to 2 (at extremality). Our bounds on $c_1$ imply that neutral black holes will necessarily have a negative viscosity shift, violating the KSS bound. Models where this is realized are known to exist—the first UV complete counter-example to the KSS bound was given in [46]. For extremal black holes, the dependence on $c_1$ drops out and only the sign of $c_2$ matters, $\eta/s = \frac{1}{4\pi}(1 + 8c_2)$. For AdS$_5$, the $c_2$ coefficient may have both positive and negative values. However, imposing the null energy condition implies an additional constraint on the range of $c_2$, which in AdS$_5$ takes the form

$$\frac{13}{12}\,c_1 + c_2 > 0\,. \tag{6.6.3}$$

This may be seen by first noticing that the definition of the parameter $\gamma$ in equation (6.2.8) implies $\gamma > 0$ as long as the null energy condition holds. Then the bound in (6.6.3) may be derived from the specific form of $\gamma$ given in (6.2.9). This alone is sufficient to bound $c_2$ from below, when $c_1$ is non-negative. Thus, one can see that utilizing such constraints it is at least in principle possible to bound $\eta/s$ *from below*, in specific cases. To what extent this can be done generically is still an open question.

It might be interesting to try to relate the extremality bounds to the transport coefficients of the boundary theory in a more concrete way. As the corrections to $\eta/s$ depend only on $c_1$ and $c_2$ in five dimensions, it is clear that the shift to extremality is not captured by the physics that controls $\eta/s$ alone. One might wonder, however, if some other linear combination of transport coefficients, such as the conductivity or susceptibility[5], might be related to the extremality shift. From a purely CFT point of view, this is certainly not that strange; the philosophy of conformal hydrodynamics is that scaling symmetry ties together ultraviolet quantities $(a, c)$ that characterize the CFT to the transport coefficients, which characterize the IR, long-wavelength behavior of the theory. If we believe that EFT coefficients in the bulk are related to these UV quantities (as is known in the case of $c_1$), then a correspondence between higher-derivatives and hydrodynamics is very natural. The question is to what extent this can be used to efficiently constrain IR quantities. Finally, we should note that extending our analysis to holographic theories that couple gravity to scalars would be useful to make contact with the efforts to generate non-trivial temperature dependence for $\eta/s$ (see *e.g.* the discussion in [220, 221]), which is expected to play a key role in understanding the dynamics of the strongly coupled quark gluon plasma.

Our results also have potential to make contact with the work on CFTs at large global charge [222]. As we've seen above, the extremality curve for AdS-Reissner-Nordström black holes is non-linear even at the two-derivative level. In an analysis of the minimum scaling dimension for highly charged 3D CFTs states of a given charge, it was found [194] that $\Delta \sim q^{3/2}$. This is in striking agreement with the extremality relationship $m \sim q^{3/2}$ that holds for large black holes. The large charge OPE may be powerful because it offers an expansion parameter, $1/q$, which may be used even for CFTs which are strongly coupled. In principle, it should be possible to match our higher-derivative corrections to the extremality bound with corrections to the minimum scaling dimension that are subleading in $1/q$. This might allow one to use the large charge OPE to compute the EFT coefficients of the bulk dual of specific theories where the minimum scaling dimensions are known.

## 6.7 Weak Gravity Conjecture in AdS

One of the motivations for this work is to address to question of to what extent the WGC is constraining in Anti-de Sitter space. It is not obvious that it should be. In flat space, one looks for higher-derivative corrections to shift the extremality bound $m(q)$ to have a slope that is greater than one. In that case, a single nearly extremal black holes is (kinematically) allowed to decay to

[5] These have been considered in [219], which already in 2008 had an interesting comment about a possible relation to the WGC.

two smaller black holes, which can fly apart off to infinity and decay further if they wish.

In AdS, the extremality bound $m(q)$ has a slope that is greater than one at the two-derivative level. Therefore one might expect that large black holes are already able to decay without any new particles or higher-derivative corrections. This picture may be too naive, however; the AdS radius introduces a long range potential that is proportional to $\frac{r^2}{l^2}$. This causes all massive states emitted from the black hole to fall back in, contrary to the situation in flat space.

A different decay path is provide by the dynamical instability [223–226], whereby charged black branes are unstable to formation of a scalar condensate. This occurs only if the theory also has a scalar with charge $q$ and dimension $\Delta$ that satisfies

$$(m_\phi l)^2 \leq \frac{1}{2}(q_\phi g M_{Pl} l)^2 - \frac{3}{2}\,. \tag{6.7.1}$$

Note that, even in the limit of large AdS-radius $l$, this does not approach the bound we have for small black holes, which is $m \leq q$. Numerical work in [224] suggests that the endpoint of the instability is a state where all the charge is carried by the scalar condensate. Similar requirements appear for the superradiant instability of small black holes [227,228]. For a more thorough review, see [187]. In either case, it is curious that in AdS, a condition similar to the flat space WGC allows for black holes to decay through an entirely different mechanism.

Another remarkable hint of the WGC comes from its connection to cosmic censorship. In [229, 230], it is shown that a class of solutions of Einstein-Maxwell theory in AdS$_4$ that appear to violate cosmic censorship [231] are removed if the theory is modified to include a scalar whose charge is great enough to satisfy the weak gravity bound[6].

It may be possible to study these solutions in the presence of higher-derivative corrections. One might ask whether there is a choice of higher-derivative terms such that the singular solutions are removed. It would be interesting to check if this occurs when the higher-derivative terms are those that are obtained by integrating out a scalar of sufficient charge. It would also be interesting to compare constraints obtained by requiring cosmic censorship with constraints due to positivity of the entropy shift.

A more general proof of the WGC in AdS was given in [190]. In that paper, it was shown that, under mild assumptions, entanglement entropy for the boundary dual of an extremal black brane should go like the surface area of the entangling subregion, which is in tension with the volume law scaling predicted by the Ryu-Takayanagi formula. The contradiction is removed when one introduces a WGC-satisfying state. This violates one of the assumptions that imply the area law for the entropy– that is, the assumption that correlations decay exponentially with distance.

This form of the WGC in particularly interesting to us because it makes no reference to whether

[6]The bound they consider is the bound for superradiance of small black holes, which requires $\Delta \leq ql$.

or not the WGC-satisfying state is a particle, or a non-perturbative object like a black hole. Therefore, the contradiction pointed out in that paper may be lifted if the higher-derivative corrections allow for black holes with charge greater than mass. Heavy black holes in AdS have masses far greater than their charge– therefore we expect that the WGC-satisfying states might be provided by small black holes whose higher-derivative corrections shift the extremality bound to allow slightly more charge.

# APPENDIX A

# Manifestly Local Soft Subtracted Recursion

In this appendix, we derive the manifestly local form (3.2.6) of the subtracted recursion relations. For a given factorization channel, consider from the recursion relations (3.2.5) the expression

$$\frac{\hat{\mathcal{A}}_L^{(I)}(z_I^{\pm})\hat{\mathcal{A}}_R^{(I)}(z_I^{\pm})}{F(z_I^{\pm})P_I^2(1-z_I^{\pm}/z_I^{\mp})} = \sum_{z_I=z_I^{\pm}} \text{Res}_{z=z_I} \frac{\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)}{z\,F(z)\,\hat{P}_I^2} = \oint_{\mathcal{C}} dz\, \frac{\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)}{z\,F(z)\,\hat{P}_I^2}\,, \tag{A.0.1}$$

where the contour surrounds only the two poles $z_I^{\pm}$. The second equality is non-trivial and deserves clarification. In the second expression, the subamplitudes $\hat{\mathcal{A}}_L^{(I)}(z)$ and $\hat{\mathcal{A}}_R^{(I)}(z)$ are only defined precisely on the residue values $z = z_I^{\pm}$ for which the internal momentum $\hat{P}_I$ is on-shell; in general one cannot just think of $\hat{\mathcal{A}}_{L,R}^{(I)}(z)$ as functions of $z$. However, in the product $\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)$, one can eliminate the internal momentum $\hat{P}_I$ in favor of the $n$ shifted external momenta by using momentum conservation. Then the resulting expression can be analytically continued in $z$ away from the residue value. This is implicitly what has been done in performing the second step in (A.0.1).

Let us assess the large-$z$ behavior of the integrand in (A.0.1). The L and R subamplitudes have couplings $g_L$ and $g_R$ such that $g_L g_R = g_n$, with $g_n$ the coupling of $\mathcal{A}_n$. Their mass-dimensions are related as $[g_L] + [g_R] = [g_n]$. Hence, using $n_L + n_R = n + 2$ and (3.2.13), we find that the numerator behaves at large $z$ as

$$\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z) \to z^{D_L} z^{D_R} = z^{6-n-[g_n]-\sum_{i=1}^n s_i - 2s_P} = z^{D+2-2s_P}\,, \tag{A.0.2}$$

where $s_P$ denotes the spin of the particle exchanged on the internal line and $D$ is the large $z$ behavior of the $\mathcal{A}_n$ which we know satisfies $D - \sum_{i=1}^n \sigma_i < 0$, by the assumption that the amplitude $\mathcal{A}_n$ is recursively constructible by the criterion (3.2.7). We therefore conclude that the integrand in (A.0.1) behaves as $z^{D-1-\sum_{i=1}^n \sigma_i - 2s_P}$, i.e. it goes to zero as $1/z^2$ or faster. Hence, there is no simple pole at $z \to \infty$.

If we deform the contour, we get the sum over all poles $z \neq z_I^{\pm}$ in $\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)/(z\,F(z)\,\hat{P}_I^2)$. Let us *assume that* $\mathcal{A}_L^{(I)}$ *and* $\mathcal{A}_R^{(I)}$ *are both local*: they have no poles and hence we pick up exactly the simple poles at $z = 0$ and $z = 1/a_i$ for $i = 1, 2, \ldots, n$. We then conclude that the soft recursion relations take the form

$$\mathcal{A}_n = \sum_I \sum_{z'=0,\frac{1}{a_1},\ldots,\frac{1}{a_n}} \sum_{|\psi^{(I)}\rangle} \text{Res}_{z=z'} \frac{\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)}{z\,F(z)\,\hat{P}_I^2}\,, \tag{A.0.3}$$

where $F(z) = \prod_{i=1}^n (1 - a_i z)^{\sigma_i}$. This form of the recursion relation is manifestly rational in the momenta.

Note that only the $z = 0$ residues give pole terms in $\mathcal{A}_n$. Therefore the sum of the $1/a_i$ residues over all channels must be a local polynomial in the momenta. For example, it is valid for the reconstruction of the 6-point scalar amplitude of NLSM, but not for the reconstruction of an 8-point amplitude.

# APPENDIX B

# Explicit Expressions for Amplitudes

In this appendix, we present expressions for the 4- and 6-point amplitudes of the theories discussed in the main text. The 6-point amplitudes were reconstructed with the 4-point ones as input, by means of the subtracted recursion relations and the the supersymmetry Ward identities also discussed in the main text.

## B.1 Supersymmetric $\mathbb{CP}^1$ NLSM

Below, we list the amplitudes for the $\mathbb{CP}^1$ $\mathcal{N}=1$ supersymmetric NLSM. This model is discussed in Section 3.5 as an illustration of our methods.

The 4-point amplitudes are:

$$\mathcal{A}_4(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}}) = \frac{1}{\Lambda^2} s_{13}\,, \tag{B.1.1}$$

$$\mathcal{A}_4(1_Z 2_{\bar{Z}} 3_\psi^+ 4_\psi^-) = -\frac{1}{\Lambda^2}\,[23]\,\langle 24\rangle = \frac{1}{2\Lambda^2}\langle 4|p_1 - p_2|3]\,, \tag{B.1.2}$$

$$\mathcal{A}_4(1_\psi^+ 2_\psi^- 3_\psi^+ 4_\psi^-) = -\frac{1}{\Lambda^2}\,[13]\,\langle 24\rangle\,. \tag{B.1.3}$$

They serve as the input for computing the 6-point amplitudes recursively:

$$\begin{aligned}
&\mathcal{A}_6(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}} 5_Z 6_{\bar{Z}}) \\
&\quad = \frac{1}{\Lambda^4}\left[\left(\frac{s_{13}s_{46}}{p_{123}^2} + (1\leftrightarrow 5) + (3\leftrightarrow 5)\right) + (2\leftrightarrow 4) + (2\leftrightarrow 6) + 3p_{135}^2\right], \qquad \text{(B.1.4)}\\
&\mathcal{A}_6(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}} 5_\psi^+ 6_\psi^-) \\
&\quad = \frac{1}{\Lambda^4}\left[\left(\frac{s_{13}\,[54]\,\langle 46\rangle}{p_{123}^2} + (2\leftrightarrow 4)\right) - \left(\frac{s_{24}\,[51]\,\langle 16\rangle}{p_{156}^2} + (1\leftrightarrow 3)\right)\right.\\
&\qquad\qquad \left. - \left(\left(\frac{[54]\,\langle 4|p_{126}|2]\langle 26\rangle}{p_{126}^2} + (1\leftrightarrow 3)\right) + (2\leftrightarrow 4)\right) + \langle 6|p_{135}|5]\right], \qquad \text{(B.1.5)}\\
&\mathcal{A}_6(1_Z 2_{\bar{Z}} 3_\psi^+ 4_\psi^- 5_\psi^+ 6_\psi^-) \\
&\quad = \frac{1}{\Lambda^4}\left[-\left(\frac{[31]\,\langle 1|p_{123}|5]\langle 46\rangle}{p_{123}^2} - (3\leftrightarrow 5)\right) + \left(\frac{[35]\,\langle 4|p_{126}|2]\langle 26\rangle}{p_{126}^2} - (4\leftrightarrow 6)\right)\right.\\
&\qquad\qquad \left. - \left(\left(\frac{[51]\,\langle 16\rangle\,[32]\,\langle 24\rangle}{p_{156}^2} - (3\leftrightarrow 5)\right) - (4\leftrightarrow 6)\right)\right], \qquad \text{(B.1.6)}\\
&\mathcal{A}_6(1_\psi^+ 2_\psi^- 3_\psi^+ 4_\psi^- 5_\psi^+ 6_\psi^-) \\
&\quad = \frac{1}{\Lambda^4}\left[\left(\frac{[13]\,\langle 2|p_{123}|5]\langle 46\rangle}{p_{123}^2} - (1\leftrightarrow 5) - (3\leftrightarrow 5)\right) - (2\leftrightarrow 4) - (2\leftrightarrow 6)\right]. \qquad \text{(B.1.7)}
\end{aligned}$$

Note that only the pure scalar amplitudes and the 2-fermion amplitudes have local terms. The 6-point amplitudes satisfy the NMHV supersymmetry Ward identities.

## B.2 Supersymmetric Quartic Galileon Theory

Below, we list the amplitudes of an $\mathcal{N} = 1$ supersymmetric quartic Galileon. This model was discussed in detail in [52] and reviewed in Section 3.6. The 4-point amplitudes are

$$\mathcal{A}_4(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}}) = \frac{1}{\Lambda^6} s_{12}s_{13}s_{23}\,, \qquad \text{(B.2.1)}$$

$$\mathcal{A}_4(1_Z 2_{\bar{Z}} 3_\psi^+ 4_\psi^-) = \frac{1}{\Lambda^6} s_{12}s_{23}\,[32]\,\langle 24\rangle = \frac{1}{2\Lambda^6} s_{12}s_{23}\langle 4|p_1 - p_2|3]\,, \qquad \text{(B.2.2)}$$

$$\mathcal{A}_4(1_\psi^+ 2_\psi^- 3_\psi^+ 4_\psi^-) = -\frac{1}{\Lambda^6}\,[13]\,\langle 24\rangle s_{12}s_{23}\,. \qquad \text{(B.2.3)}$$

At 6-point, only the amplitudes with at most two fermions are constructible with soft subtracted recursion relations. The remaining ones are fixed by the supersymmetry Ward identities, we find

$$\begin{aligned}
&\mathcal{A}_6(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}} 5_Z 6_{\bar{Z}}) \\
&\quad = \frac{1}{\Lambda^{12}}\left[\left(\frac{s_{12}s_{13}s_{23}s_{45}s_{46}s_{56}}{p_{123}^2} + (1 \leftrightarrow 5) + (3 \leftrightarrow 5)\right) + (2 \leftrightarrow 4) + (2 \leftrightarrow 6)\right], \qquad \text{(B.2.4)}
\end{aligned}$$

$$\begin{aligned}
&\mathcal{A}_6(1_Z 2_{\bar{Z}} 3_Z 4_{\bar{Z}} 5_\psi^+ 6_\psi^-) \\
&\quad = \frac{1}{\Lambda^{12}}\left[\left(\frac{s_{12}s_{13}s_{23}s_{45}s_{56}\,[54]\,\langle 46\rangle}{p_{123}^2} + (2 \leftrightarrow 4)\right) - \left(\frac{s_{16}s_{23}s_{24}s_{34}s_{56}\,[51]\,\langle 16\rangle}{p_{156}^2} + (1 \leftrightarrow 3)\right)\right. \\
&\qquad\quad \left. + \left(\left(\frac{s_{12}s_{16}s_{34}s_{45}\,[53]\,\langle 3|p_{126}|2]\langle 26\rangle}{p_{126}^2} + (1 \leftrightarrow 3)\right) + (2 \leftrightarrow 4)\right)\right], \qquad \text{(B.2.5)}
\end{aligned}$$

$$\begin{aligned}
&\mathcal{A}_6(1_Z 2_{\bar{Z}} 3_\psi^+ 4_\psi^- 5_\psi^+ 6_\psi^-) \\
&\quad = \frac{1}{\Lambda^{12}}\left[\left(\frac{[31]\,\langle 1|p_{46}|5]\langle 46\rangle}{p_{123}^2} - (3 \leftrightarrow 5)\right) + \left(\frac{[35]\,\langle 4|p_{16}|2]\langle 26\rangle}{p_{126}^2} - (4 \leftrightarrow 6)\right)\right. \\
&\qquad\quad \left. - \left(\left(\frac{[32]\,\langle 24\rangle\,[51]\,\langle 16\rangle}{p_{156}^2} - (3 \leftrightarrow 5)\right) - (4 \leftrightarrow 6)\right)\right], \qquad \text{(B.2.6)}
\end{aligned}$$

$$\begin{aligned}
&\mathcal{A}_6(1_\psi^+ 2_\psi^- 3_\psi^+ 4_\psi^- 5_\psi^+ 6_\psi^-) \\
&\quad = \frac{1}{\Lambda^{12}}\left[\left(\frac{[13]\,\langle 2|p_{13}|5]\langle 46\rangle}{p_{123}^2} - (1 \leftrightarrow 5) - (3 \leftrightarrow 5)\right) - (2 \leftrightarrow 4) - (2 \leftrightarrow 6)\right]. \qquad \text{(B.2.7)}
\end{aligned}$$

None of the amplitudes have local terms.

## B.3 Chiral Perturbation Theory

Below, we list the color-ordered amplitudes of the $\frac{U(N)\times U(N)}{U(N)}$ sigma model, with higher derivative corrections, referred to as chiral perturbation theory in the main text. Different color orderings are related to the ones listed by momentum relabelling. At 4-point we have

$$\mathcal{A}_4[1,2,3,4] = \frac{g_2}{\Lambda^2}t + \frac{g_6}{\Lambda^6}t\left(s^2+t^2+u^2\right) + \frac{g_8}{\Lambda^8}st^2u + \mathcal{O}(\Lambda^{-10}) \qquad \text{(B.3.1)}$$

and at 6-point

$$
\begin{aligned}
&\mathcal{A}_6[1,2,3,4,5,6] \\
&= \frac{g_2^2}{\Lambda^4}\left[\frac{s_{13}s_{46}}{p_{123}^2}+\frac{s_{24}s_{15}}{p_{234}^2}+\frac{s_{35}s_{26}}{p_{345}^2}-s_{24}-s_{26}-s_{46}\right] \\
&+ \frac{g_2 g_6}{\Lambda^8}\left[\frac{s_{13}s_{46}}{p_{123}^2}\left(s_{12}^2+s_{13}^2+s_{23}^2+s_{45}^2+s_{46}^2+s_{56}^2\right)\right. \\
&\quad + \frac{s_{24}s_{15}}{p_{234}^2}\left(s_{23}^2+s_{24}^2+s_{34}^2+s_{56}^2+s_{15}^2+s_{16}^2\right)+\frac{s_{35}s_{26}}{p_{345}^2}\left(s_{34}^2+s_{35}^2+s_{45}^2+s_{16}^2+s_{26}^2+s_{12}^2\right) \\
&\quad - 2\big(s_{26}^3+s_{23}s_{26}^2+s_{25}s_{26}^2+s_{34}s_{26}^2+s_{45}s_{26}^2+s_{23}^2s_{26}+s_{25}^2s_{26}+s_{34}^2s_{26}+s_{35}^2s_{26}+s_{45}^2s_{26} \\
&\qquad + s_{23}s_{34}s_{26}+s_{23}s_{35}s_{26}+s_{25}s_{35}s_{26}+s_{34}s_{36}s_{26}+s_{23}s_{45}s_{26}+s_{34}s_{45}s_{26}+s_{36}s_{45}s_{26} \\
&\qquad + s_{46}^3+s_{24}s_{25}^2+s_{24}s_{35}^2+s_{24}s_{45}^2+s_{23}s_{46}^2+s_{25}s_{46}^2+s_{34}s_{46}^2+s_{35}s_{46}^2+s_{36}s_{46}^2 \\
&\qquad + s_{45}s_{46}^2+s_{24}s_{35}s_{36}+s_{25}^2s_{46}+s_{34}^2s_{46}+s_{35}^2s_{46}+s_{36}^2s_{46}+s_{45}^2s_{46}+s_{23}s_{25}s_{46} \\
&\qquad + s_{25}s_{34}s_{46}+s_{23}s_{45}s_{46}+s_{34}s_{45}s_{46}+s_{35}s_{45}s_{46}+s_{36}s_{45}s_{46}\big) \\
&\quad - 4\big(s_{24}^3+s_{25}s_{24}^2+s_{35}s_{24}^2+s_{45}s_{24}^2+s_{23}^2s_{24}+s_{34}^2s_{24}+s_{36}^2s_{24}+s_{23}s_{25}s_{24}+s_{25}s_{34}s_{24} \\
&\qquad + s_{23}s_{35}s_{24}+s_{25}s_{35}s_{24}+s_{34}s_{35}s_{24}+s_{26}s_{36}s_{24}+s_{23}s_{45}s_{24}+s_{25}s_{45}s_{24}+s_{34}s_{45}s_{24} \\
&\qquad + s_{35}s_{45}s_{24}+s_{36}s_{45}s_{24}+s_{23}s_{25}s_{26}+s_{25}s_{26}s_{34}+s_{25}s_{26}s_{45}+s_{23}^2s_{46}+s_{25}s_{26}s_{46} \\
&\qquad + s_{23}s_{34}s_{46}+s_{23}s_{35}s_{46}+s_{34}s_{35}s_{46}+s_{23}s_{36}s_{46}+s_{25}s_{36}s_{46}+s_{26}s_{36}s_{46}+s_{34}s_{36}s_{46} \\
&\qquad + s_{35}s_{36}s_{46}+s_{25}s_{45}s_{46}+s_{26}s_{45}s_{46}\big) \\
&\quad - 6\big(s_{23}s_{24}^2+s_{34}s_{24}^2+s_{36}s_{24}^2+s_{26}^2s_{24}+s_{46}^2s_{24}+s_{23}s_{26}s_{24}+s_{25}s_{26}s_{24}+s_{23}s_{34}s_{24} \\
&\qquad + s_{26}s_{34}s_{24}+s_{23}s_{36}s_{24}+s_{25}s_{36}s_{24}+s_{26}s_{45}s_{24}+s_{25}s_{46}s_{24}+s_{35}s_{46}s_{24}+s_{45}s_{46}s_{24} \\
&\qquad + s_{26}s_{46}^2+s_{25}s_{34}s_{36}+s_{25}s_{36}s_{45}+s_{26}^2s_{46}+s_{23}s_{26}s_{46}+s_{26}s_{34}s_{46}\big) \\
&\quad \left. - 8s_{24}\big(s_{24}s_{26}+s_{34}s_{36}+s_{23}s_{46}+s_{24}s_{46}+s_{34}s_{46}+s_{36}s_{46}\big)-12s_{24}s_{26}s_{46}\right]+\mathcal{O}(\Lambda^{-10})\,.
\end{aligned}
\tag{B.3.2}
$$

These amplitudes are discussed in further detail in Section 3.6.6.

# APPENDIX C

# Recursion Relations and Ward Identities

We show that if the seed amplitudes of a recursive theory satisfy a set of Ward identities, then all recursively constructible $n$-point amplitudes also satisfy them. For Abelian groups, this follows from two features:

(a) additive charges have Ward identities that simply state that the sum of charges of the states in an amplitude must vanish.

(b) CPT conjugate states sitting on either end of a factorization channel have equal and opposite charges.

Hence recursion will result in amplitudes that respect the Abelian symmetry so long as the seed amplitudes do.

Now consider Ward identities generated by elements of a semi-simple Lie algebra. In the root space decomposition of the algebra, we can choose a triplet of generators: raising operators $\mathcal{T}_+$, lowering operators $\mathcal{T}_-$, and "diagonal" $\mathcal{T}_0$ generators, for each positive root that satisfy the algebra

$$[\mathcal{T}_+, \mathcal{T}_-] = \mathcal{T}_0\,, \quad [\mathcal{T}_+, \mathcal{T}_0] = -2\mathcal{T}_+\,, \quad [\mathcal{T}_-, \mathcal{T}_0] = 2\mathcal{T}_-\,. \tag{C.0.1}$$

In order for representations of this algebra to be physical, CPT must be an algebra automorphism. The CPT charge conjugation generator $\mathcal{C}$ must also flip the sign of the additive $\mathcal{T}_0$-charge. So we determine the action of $\mathcal{C}$ to be

$$\begin{aligned}
\mathcal{C} \cdot \mathcal{T}_0 \cdot X &= -\mathcal{T}_0 \cdot \mathcal{C} \cdot X = -\mathcal{T}_0 \cdot \tilde{X}\,, \\
\mathcal{C} \cdot \mathcal{T}_+ \cdot X &= -\mathcal{T}_- \cdot \mathcal{C} \cdot X = -\mathcal{T}_- \cdot \tilde{X}\,, \\
\mathcal{C} \cdot \mathcal{T}_- \cdot X &= -\mathcal{T}_+ \cdot \mathcal{C} \cdot X = -\mathcal{T}_+ \cdot \tilde{X}\,,
\end{aligned} \tag{C.0.2}$$

where $X$ is a physical state and we have defined the conjugate state $\tilde{X}$ to be the charge conjugate of $X$, i.e. $\tilde{X} = \mathcal{C} \cdot X$.

If the S-matix is recursively constructible (at some order in the derivative expansion) then each $n$-point amplitude is given as a sum over factorization singularities with residues given in terms of a product of amplitudes with fewer external states

$$\mathcal{A}_n(1,\cdots,n)=\sum_I\sum_X\underset{z=z_I^\pm}{\text{Res}}\,\frac{\hat{\mathcal{A}}_L^{(I)}(z)\hat{\mathcal{A}}_R^{(I)}(z)}{z\hat{P}_I(z)^2F(z)}\,, \tag{C.0.3}$$

where $I$ labels all possible factorization channels and $X$ the exchanged internal states. Since $\mathcal{T}_0$ is diagonal, the Ward identity generated by $\mathcal{T}_0$ works just like in the Abelian case – charges can be assigned to the physical states and recursion preserves this charge in any $n$-point amplitude. More complicated are the non-diagonal generators $\mathcal{T}_\pm$. For simplicity, we present the argument explicitly for $SU(2)_R$ Ward identities as they apply to the $\mathcal{N}=2$ NLSM described in Section 3.5.2. For $SU(2)_R$, the action of $\mathcal{T}_+$ on the fermion helicity states is given in (3.5.23). The scalar and vectors are singlets under $SU(2)_R$.

The statement of the $SU(2)_R$ Ward identity is that $\mathcal{T}_+\cdot\mathcal{A}_n(1,...,n)=0$. The inductive assumption is that this holds true for the lower-point amplitudes in the recursive expression for $\mathcal{A}_n(1,...,n)$. We already know from Section 3.5.2 that $SU(2)_R$ is a symmetry of the 3- and 4-point amplitudes, so that provides the basis of induction.

The action of $\mathcal{T}_+$ on the recursive expression for an $n$-point amplitude is

$$\mathcal{T}_+\cdot\mathcal{A}_n(1,...,n)\equiv\sum_{i=1}^n(-1)^{P_i}\mathcal{A}_n(1,...,\mathcal{T}_+\cdot i,...,n) \tag{C.0.4}$$

$$\begin{aligned}=\sum_I\sum_X\underset{z=z_I^\pm}{\text{Res}}\Bigg[&\sum_{i\in I}(-1)^{P_i}\frac{\hat{\mathcal{A}}_L^{(I)}(\ldots,\mathcal{T}_+\cdot i,\ldots,X)\hat{\mathcal{A}}_R^{(I)}(\ldots)}{z\hat{P}_I(z)^2F(z)}\\&+\sum_{i\notin I}(-1)^{P_i}\frac{\hat{\mathcal{A}}_L^{(I)}(\ldots)\hat{\mathcal{A}}_R^{(I)}(\tilde{X},\ldots,\mathcal{T}_+\cdot i,\ldots)}{z\hat{P}_I(z)^2F(z)}\Bigg]\,,\end{aligned} \tag{C.0.5}$$

where $P_i=0$ or $1$ corresponds to the additional signs in the prefactors for the action of $\mathcal{T}_+$ as given in Table 3.5.23. We now prove that this expression vanishes channel by channel. Without loss of generality, we will show that the contribution from the $(1\ldots k)^\pm$ channel vanishes independently, where $+$ means the contribution from the $z^\pm$ residue. The argument follows for all other factorization channels by replacing $(1\ldots k)^\pm$ by $I^\pm$. For the $(1\ldots k)$-channel, the relevant part of (C.0.4) that we want to show vanishes is

$$\begin{aligned}\sum_X\Bigg[&\Bigg(\sum_{i=1}^k(-1)^{P_i}\hat{\mathcal{A}}_L(1,\ldots,\mathcal{T}_+\cdot i,\ldots,k,X)\Bigg)\hat{\mathcal{A}}_R(\tilde{X},k+1,\ldots,n)\\&+\hat{\mathcal{A}}_L(1,\ldots,k,X)\Bigg(\sum_{i=k+1}^n(-1)^{P_i}\hat{\mathcal{A}}_R(\tilde{X},k+1,\ldots,\mathcal{T}_+\cdot i,\ldots,n)\Bigg)\Bigg].\end{aligned} \tag{C.0.6}$$

By the inductive assumption, the lower-point amplitudes respect the $\mathcal{T}_+$ Ward identities

$$\sum_{i=1}^{k}(-1)^{P_i}\hat{\mathcal{A}}_L(1,\ldots,\mathcal{T}_+\cdot i,\ldots,k,X) \;=\; (-1)^{P_X+1}\hat{\mathcal{A}}_L(1,\ldots,k,\mathcal{T}_+\cdot X)\,, \tag{C.0.7}$$

and similarly for $\hat{\mathcal{A}}_R$. Using this relation and splitting the sum over particles $X$ allows us to rewrite (C.0.6) as

$$\begin{aligned}
&-\sum_X(-1)^{P_X}\Big[\hat{\mathcal{A}}_L(1,\ldots,k,\mathcal{T}_+\cdot X)\hat{\mathcal{A}}_R(\tilde{X},k+1,\ldots,n)\Big]\\
&-\sum_{X'}(-1)^{P_{\tilde{X}'}}\Big[\hat{\mathcal{A}}_L(1,\ldots,k,X')\hat{\mathcal{A}}_R(\mathcal{T}_+\cdot\tilde{X}',k+1,\ldots,n)\Big]\,.
\end{aligned} \tag{C.0.8}$$

In the second line we have made a change of dummy summation variable that we now exploit further.

It is non-trivial, but turns out to be true for $SU(2)_R$ as we have explicitly checked, that if we define $X'=\mathcal{T}_+\cdot X$ and sum over $X$ instead of $X'$, the second line of (C.0.8) gives exactly the same result. We can then write (C.0.8) as

$$\begin{aligned}
-\sum_X\Big[&(-1)^{P_X}\hat{\mathcal{A}}_L(1,\ldots,k,\mathcal{T}_+\cdot X)\hat{\mathcal{A}}_R(\tilde{X},k+1,\ldots,n)\\
&+(-1)^{P_{\tilde{X}'}}\hat{\mathcal{A}}_L(1,\ldots,k,\mathcal{T}_+\cdot X)\hat{\mathcal{A}}_R(\mathcal{T}_+\cdot C\cdot\mathcal{T}_+\cdot X,k+1,\ldots,n)\Big]\,.
\end{aligned} \tag{C.0.9}$$

Since $\mathcal{T}_+\cdot C\cdot\mathcal{T}_+\cdot X=\mathcal{T}_+\cdot\mathcal{T}_-\cdot\tilde{X}$, this becomes

$$\begin{aligned}
-\sum_X\Big[&(-1)^{P_X}\hat{\mathcal{A}}_L(1,\ldots,k,\mathcal{T}_+\cdot X)\hat{\mathcal{A}}_R(\tilde{X},k+1,\ldots,n)\\
&+(-1)^{P_{\mathcal{T}_-\cdot\tilde{X}}+Q_{\tilde{X}}+1}\hat{\mathcal{A}}_L(1,\ldots,k,\mathcal{T}_+\cdot X)\hat{\mathcal{A}}_R(\mathcal{T}_+\cdot\mathcal{T}_-\cdot\tilde{X},k+1,\ldots,n)\Big]\,.
\end{aligned} \tag{C.0.10}$$

where $Q_X$ refers to the prefactors for the action of $\mathcal{T}_-$ as given in Table 3.5.23. This vanishes when $\mathcal{T}_+\cdot\mathcal{T}_-\cdot\tilde{X}=\tilde{X}$ and $P_{\mathcal{T}_-\cdot\tilde{X}}+Q_{\tilde{X}}=0$ for any state $X$ such that $\mathcal{T}_+\cdot X\neq 0$. For $SU(2)_R$, we can check explicitly that these conditions are satisfied. The only states for which $\mathcal{T}_+\cdot X\neq 0$ are $X=\psi^{2+}$ and $\psi_1^-$. Their conjugates are $\tilde{X}=\psi_2^-$ and $\psi^{2+}$, respectively, and by (3.5.23) we have

$$\mathcal{T}_+\cdot\mathcal{T}_-\cdot\psi^{1+}=\mathcal{T}_+\cdot\psi^{2+}=\psi^{1+}\qquad\mathcal{T}_+\cdot\mathcal{T}_-\cdot\psi_2^-=\mathcal{T}_+\cdot\psi_1^-=\psi_2^-\,, \tag{C.0.11}$$

$$P_{\mathcal{T}_-\cdot\psi^{1+}}+Q_{\psi^{1+}}=0+0=0\qquad P_{\mathcal{T}_-\cdot\psi_2^-}+Q_{\psi_2^-}=1+1=0\ (\mathrm{mod}\ 2)\,. \tag{C.0.12}$$

If follows that from the inductive step that all amplitudes satisfy the $SU(2)_R$ Ward identities when the seed amplitudes do.

# APPENDIX D

# Structure of Contact Terms

In Section 4.3.1 we argued, by a combination of dimensional analysis, little group scaling and requiring vanishing as $\mu^2 \to 0$, that contact terms could appear in the mDBI$_4$ amplitudes in the NSD sector in the form of some contraction of the form

$$\mathcal{A}_n^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^+, n_{\overline{\phi}}\right)\Big|_{\text{contact}} \sim \mu^2 |2]^2 |3]^2 \ldots |n-1]^2, \tag{D.0.1}$$

where $n$ is even. In this appendix we will give a short proof that there is a unique such contact term for each $n$. We begin by noting that any candidate term has the form of a sum over terms where each term is a sum over cyclic contractions of the spinors. For example for $n = 12$ typical terms might have the form

$$([23][34][45][56][67][72])\,([89][9,10][10,11][11,8])\,, \tag{D.0.2}$$

or

$$([23][34][42])\,([56][67][75])\,([89][9,10][10,11][11,8])\,. \tag{D.0.3}$$

Neither term by itself is a candidate contact term since it does not have the appropriate Bose symmetry. We should take expression (D.0.2) and symmetrize over each pair of spinors, beginning with 3 and 4 gives

$$([23][34][45] + [24][43][35])\,[56][67][72]\,([89][9,10][10,11][11,8])\,, \tag{D.0.4}$$

applying the Schouten identity then gives

$$= -[34]^2\,([25][56][67][72])\,([89][9,10][10,11][11,8])\,. \tag{D.0.5}$$

This has reduced a cyclic contraction of length 6 to a product of cyclic contractions of *strictly shorter* length. By Bose symmetrizing over all pairs of spinors we can reduce any possible contact term to a sum over product of cyclic contractions of length 2. Terms such as (D.0.3) with odd cyclic contractions vanish after Bose symmetrization. The final expression then has the unique form

$$\mathcal{A}_n^{\mathrm{mDBI}_4}\left(1_\phi, 2_\gamma^+, \ldots, (n-1)_\gamma^+, n_{\bar{\phi}}^-\right)\Big|_{\text{contact}} = c_n \mu^2 \left([23]^2[45]^2 \ldots [n-2, n-1]^2 + \ldots\right), \quad \text{(D.0.6)}$$

where $+\ldots$ denotes the sum over all ways of partitioning the set $\{2, \ldots, n\}$ into subsets of length 2. This completes the proof that there is a unique possible contact term at each multiplicity.

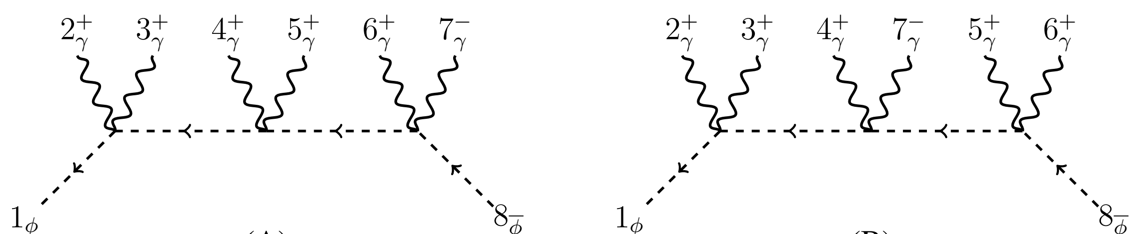
# APPENDIX E

# T-Duality Constraints on 8-point Amplitudes

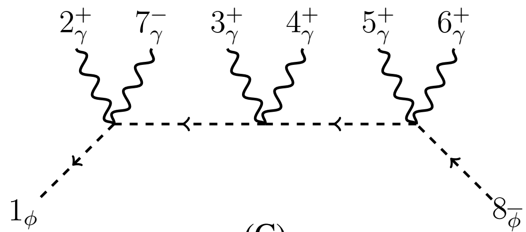
Following our discussion in Section 4.3.2, we now investigate how T-duality constrains the 8-point amplitudes in $\text{mDBI}_4$. Begin with the dimensional reduction followed by the soft limit of particle 7 for the NSD 8-point $\text{mDBI}_4$ Ansatz

$$\begin{aligned}&\mathcal{A}_8^{\text{mDBI}_4}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+, 7_\gamma^+, 8_{\overline{\phi}}\right)\\ &\xrightarrow{3d+\text{soft}} -\frac{1}{8}\left[\frac{2(\mu^2)^3 s_{23}s_{45}(p_6\cdot p_7)}{(s_{123}+\mu^2)(s_{68}+\mu^2)}\right] + c_8\mu^2 s_{23}s_{45}s_{67} + \mathcal{P}\left(2,3,4,5,6,7\right). \qquad \text{(E.0.1)}\end{aligned}$$

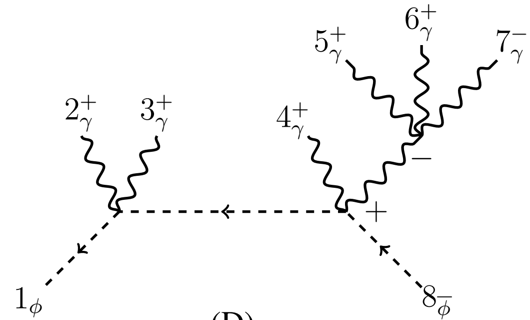
The MHV amplitude has a more complicated structure, there are more factorization graphs which are not related by permutations of external lines. Explicitly



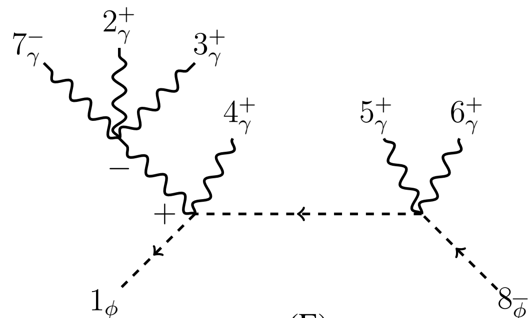

(A) (B)




(C)




(D)




(E)

In this topological decomposition the amplitude has the form

$$\begin{aligned}\mathcal{A}_8^{\text{mDBI}_4}&\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+, 7_\gamma^-, 8_{\overline{\phi}}\right)\\ &= \mathcal{A}_{8(\text{A})}^{\text{mDBI}_4} + \mathcal{A}_{8(\text{B})}^{\text{mDBI}_4} + \mathcal{A}_{8(\text{C})}^{\text{mDBI}_4} + \mathcal{A}_{8(\text{D})}^{\text{mDBI}_4} + \mathcal{A}_{8(\text{E})}^{\text{mDBI}_4},\end{aligned} \tag{E.0.2}$$

where

$$\mathcal{A}_{8(\text{A})}^{\text{mDBI}_4} \xrightarrow{3d+\text{soft}} \frac{-(\mu^2)^2 s_{23}s_{45}\left(2(p_7\cdot p_8)(s_{68}+\mu^2)+2\mu^2(p_7\cdot p_6)\right)}{(s_{123}+\mu^2)(s_{68}+\mu^2)} + \ldots \tag{E.0.3}$$

$$\mathcal{A}_{8(\text{B})}^{\text{mDBI}_4} \xrightarrow{3d+\text{soft}} \frac{-(\mu^2)^2 s_{23}s_{56}\left(4(p_7\cdot p_{123})(p_4\cdot p_{123})-2s_{123}(p_4\cdot p_7)\right)}{(s_{123}+\mu^2)(s_{568}+\mu^2)} + \ldots \tag{E.0.4}$$

$$\mathcal{A}_{8(\text{C})}^{\text{mDBI}_4} \xrightarrow{3d+\text{soft}} \frac{-(\mu^2)^2 s_{34}s_{56}\left(2(p_7\cdot p_1)(s_{12}+\mu^2)+2\mu^2(p_2\cdot p_7)\right)}{(s_{12}+\mu^2)(s_{568}+\mu^2)} + \ldots \tag{E.0.5}$$

$$\mathcal{A}_{8(\text{D})}^{\text{mDBI}_4} \xrightarrow{3d+\text{soft}} -\frac{(\mu^2)^2 s_{23}\left(4(p_7\cdot p_{56})(p_4\cdot p_{56})-2s_{56}(p_4\cdot p_7)\right)}{s_{123}+\mu^2} + \ldots \tag{E.0.6}$$

$$\mathcal{A}_{8(\text{E})}^{\text{mDBI}_4} \xrightarrow{3d+\text{soft}} -\frac{(\mu^2)^2 s_{56}\left(4(p_7\cdot p_{23})(p_4\cdot p_{23})-2s_{23}(p_4\cdot p_7)\right)}{s_{568}+\mu^2} + \ldots \tag{E.0.7}$$

Here $+\ldots$ corresponds to summing over all topologically inequivalent relabelings of the positive helicity photons. Note that we do not include a contact contribution, as discussed in Appendix D.

From the singularity structure it is clear that diagrams A, B and C must cancel against the contribution of the NSD amplitude. For diagrams A and C it is easy to pick out the relevant pieces proportional to $(\mu^2)^3$. For diagram B this is a little less obvious and requires a little algebra first.

The key idea is to recognize that there is something special about $p_4$ since it is the positive helicity particle in the middle of the diagram. We will see that something nice happens if we use momentum conservation and on-shellness to remove $p_4$ from the expression. That is we use

$$p_4 = -p_{123} - p_{568}, \tag{E.0.8}$$

and the on-shell constraint

$$p_4^2 = 0 \Rightarrow p_{123} \cdot p_{568} = -\frac{1}{2}\left(s_{123} + s_{568}\right). \tag{E.0.9}$$

Using this on the numerator of B gives

$$\begin{aligned} &4(p_7 \cdot p_{123})(p_4 \cdot p_{123}) - 2s_{123}(p_4 \cdot p_7) \\ &\quad = -2(p_7 \cdot p_{123})\left(s_{123} - s_{568}\right) + 2s_{123}(p_{123} \cdot p_7 + p_{568} \cdot p_7) \\ &\quad = 2(p_7 \cdot p_{123})(s_{568} + \mu^2) + 2(p_7 \cdot p_{568})(s_{123} + \mu^2) + 2\mu^2(p_4 \cdot p_7). \end{aligned} \tag{E.0.10}$$

We can therefore more usefully rewrite B in the form

$$\begin{aligned} &\mathcal{A}_{8(\mathrm{B})}^{\mathrm{mDBI}}\left(1_\phi, 2_\gamma^+, 3_\gamma^+, 4_\gamma^+, 5_\gamma^+, 6_\gamma^+, 7_\gamma^-, 8_{\overline{\phi}}\right) \\ &\quad \xrightarrow{3d+\text{soft}} \frac{-2(\mu^2)^3 s_{23}s_{56}(p_4 \cdot p_7)}{(s_{123}+\mu^2)(s_{568}+\mu^2)} - \frac{2(\mu^2)^2 s_{23}s_{56}(p_7 \cdot p_{123})}{s_{123}+\mu^2} - \frac{2(\mu^2)^2 s_{23}s_{56}(p_7 \cdot p_{568})}{s_{568}+\mu^2} + \ldots \end{aligned} \tag{E.0.11}$$

We now see explicitly that the non-local contributions from the MHV amplitude cancel completely. What remains is a sum of terms with only a single propagator. This is important since we want the remaining terms to cancel against each other, this couldn't happen unless some of the singularities disappeared upon dimensional reduction and soft limits since the topologically distinct graphs, by definition, have distinct singularity structure.

To finish the calculation we pick a singularity and verify that the sum of all contributions vanishes. Due to charge conjugation symmetry all such calculations are identical so we only need to verify a single case explicitly. We will choose the singularity associated with $s_{123} = -\mu^2$, this receives contributions from diagrams A, B and D. Summing the relevant terms

$$
\begin{aligned}
&-\frac{2(\mu^2)^2 s_{23}s_{45}(p_8\cdot p_7)}{s_{123}+\mu^2}-\frac{2(\mu^2)^2 s_{23}s_{56}(p_7\cdot p_{123})}{s_{123}+\mu^2}\\
&\qquad-\frac{(\mu^2)^2 s_{23}\left(4(p_7\cdot p_{56})(p_4\cdot p_{56})-2s_{56}(p_4\cdot p_7)\right)}{s_{123}+\mu^2}+\mathcal{C}\left(4,5,6\right)\\
&=-\frac{2(\mu^2)^2 s_{23}s_{456}(p_8\cdot p_7)}{s_{123}+\mu^2}-\frac{2(\mu^2)^2 s_{23}s_{456}(p_7\cdot p_{123})}{s_{123}+\mu^2}-\frac{2(\mu^2)^2 s_{23}s_{456}(p_7\cdot p_{456})}{s_{123}+\mu^2}\\
&=0.
\end{aligned}
\tag{E.0.12}
$$

As in the 6-point case we find that all of the factoring terms in the NSD and MHV mDBI$_4$ amplitudes cancel against each other and vanish in the T-dual soft configuration. Since the possible contact term is $\mathcal{O}\left(p_7\right)$, we must choose $c_8 = 0$ for compatibility with T-duality.

# APPENDIX F

# Evaluating Rational Integrals

A *rational integral* in this context is defined as an integral in $d = 4 - 2\epsilon$ dimensions, for which the integrand vanishes in $d = 4$. A powerful and general method for evaluating these integrals was given in [232] where the following *dimension shifting formula* was derived

$$\int \frac{\mathrm{d}^{4-2\epsilon} l}{(2\pi)^{4-2\epsilon}} (l_{-2\epsilon}^2)^p f(l) = (4\pi)^p \frac{\Gamma(-\epsilon + p)}{\Gamma(-\epsilon)} \int \frac{\mathrm{d}^{4+2p-2\epsilon} l}{(2\pi)^{4+2p-2\epsilon}} f(l), \tag{F.0.1}$$

where $f(l)$ is some rational function of the $d$-dimensional loop momentum. This formula allows us to exchange integrals with explicit factors of $l_{-2\epsilon}^2$ for integrals without such factors evaluated in higher dimensions. The integral on the left-hand-side of (F.0.1) is formally defined as a tensor integral

$$\int \frac{\mathrm{d}^{4-2\epsilon} l}{(2\pi)^{4-2\epsilon}} (l_{-2\epsilon}^2)^p f(l) \equiv \left( \prod_{i=1}^{p} g_{\mu_i \nu_i}^{[-2\epsilon]} \right) \int \frac{\mathrm{d}^{4-2\epsilon} l}{(2\pi)^{4-2\epsilon}} \left( \prod_{j=1}^{p} l^{\mu_j} l^{\nu_j} \right) f(l), \tag{F.0.2}$$

where $g_{\mu\nu}^{[-2\epsilon]}$ is the metric tensor projected onto the non-physical $-2\epsilon$-dimensional momentum subspace. The utility of the formula (F.0.1) is that it gives an efficient way to bypass calculating tensor reduction for integrands of arbitrarily high-rank; in this chapter all integrals can be exchanged using this method to either scalar or rank-2 tensor integrals. Even with this simplification, obtaining explicit results to all orders in $\epsilon$ is a very difficult problem, for which only a small fraction of the necessary integrals are known. At $\mathcal{O}(\epsilon^0)$ however, the formula (F.0.1) simplifies significantly and the right-hand-side depends only on the *divergent* part of the $d = 4 + 2p - 2\epsilon$-dimensional integral

$$\int \frac{\mathrm{d}^{4-2\epsilon} l}{(2\pi)^{4-2\epsilon}} (l_{-2\epsilon}^2)^p f(l) = -(p-1)!(4\pi)^p \left[ \int \frac{\mathrm{d}^{4+2p-2\epsilon} l}{(2\pi)^{4+2p-2\epsilon}} f(l) \right]_{1/\epsilon} + \mathcal{O}(\epsilon). \tag{F.0.3}$$

This is the key formula for obtaining explicit expressions for one-loop rational integrals. As we will see below the simplification arises from the fact that after Feynman parametrization the divergent part of the integral can be extracted as the trivial integration of a polynomial in Feynman parameters.

## F.1 Rational Scalar $n$-gon Integral

In this section we present the explicit calculation of the rational scalar $n$-gon integral

$$I_n^{d=4-2\epsilon}\left[(\mu^2)^n;\{p_i\}\right] \equiv \int \frac{\mathrm{d}^{4-2\epsilon}l}{(2\pi)^{4-2\epsilon}} \frac{\left(l_{-2\epsilon}^2\right)^n}{\prod_{i=1}^n \left(l-\sum_{j=1}^i p_j\right)^2}, \tag{F.1.1}$$

where the external momenta $p_i$ are massive. Using the dimension shifting formula (F.0.1) this is related to the massless scalar $n$-gon integral in $d = 4+2n-2\epsilon$ dimensions

$$= (4\pi)^n \frac{\Gamma(n-\epsilon)}{\Gamma(-\epsilon)} \int \frac{\mathrm{d}^{4+2n-2\epsilon}l}{(2\pi)^{4+2n-2\epsilon}} \frac{1}{\prod_{i=1}^n \left(l-\sum_{j=1}^i p_j\right)^2}. \tag{F.1.2}$$

The next step is to use Feynman parametrization and write the integral as

$$\begin{aligned}
&= (4\pi)^n \frac{\Gamma(n-\epsilon)}{\Gamma(-\epsilon)} (n-1)! \\
&\times \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \left[\delta\left(\sum_{i=1}^n x_i - 1\right) \int \frac{\mathrm{d}^{4+2n-2\epsilon}l}{(2\pi)^{4+2n-2\epsilon}} \frac{1}{\left[\sum_{i=1}^n x_i \left(l-\sum_{j=1}^i p_j\right)^2\right]^n}\right].
\end{aligned} \tag{F.1.3}$$

After shifting the loop momentum by $l \to l + \sum_{i=1}^{n-1}\sum_{j=1}^i x_i p_j$ the denominator of the above integrand can be written as $[l^2+\Delta]^n$ with

$$\begin{aligned}
\Delta &= \sum_{i=1}^n x_i (1-x_i) \left(\sum_{j=1}^i p_j\right)^2 - 2\sum_{i<j}^n x_i x_j \left(\sum_{k=1}^i p_k\right)\cdot\left(\sum_{k=1}^j p_k\right) \\
&= -\sum_{i=1}^n x_i (1-x_i) \left(\sum_{j=1}^i p_j\right)\cdot\left(\sum_{j=i+1}^n p_j\right) + 2\sum_{i<j}^n x_i x_j \left(\sum_{k=1}^i p_k\right)\cdot\left(\sum_{k=j+1}^n p_k\right) \\
&= -\sum_{i<j}^n p_i\cdot p_j \left(\sum_{k=i}^{j-1} x_k\right)\left(1-\sum_{k=i}^{j-1} x_k\right).
\end{aligned} \tag{F.1.4}$$

In the second line above, we used momentum conservation to write everything in terms of scalar products of two different momenta and in the third line, we rearranged the sums, writing explicitly the coefficient of each $p_i\cdot p_j$. To further simplify this, we substitute $1=\sum_{i=1}^n x_i$ and we collect the coefficients of each product $x_i x_j$,

$$\Delta = -\sum_{i<j}^{n} x_i x_j \left(\sum_{k=i+1}^{j} p_k\right) \cdot \left(\sum_{k=1}^{i} p_k + \sum_{k=j+1}^{n} p_k\right) = \sum_{i<j}^{n} x_i x_j \left(\sum_{k=i+1}^{j} p_k\right)^2 , \tag{F.1.5}$$

where in the second step we used momentum conservation to write everything in terms of Mandelstam variables of adjacent momenta. Going back to (F.1.3) and using the standard integral

$$\int \frac{\mathrm{d}^{4+2n-2\epsilon} l}{(2\pi)^{4+2n-2\epsilon}} \frac{1}{[l^2+\Delta]^n} = \frac{i}{(4\pi)^{n+2-\epsilon}} \frac{\Gamma(-2+\epsilon)}{(n-1)!} \Delta^{2-\epsilon}, \tag{F.1.6}$$

in full generality the rational integral (F.1.1) is given by the Feynman parameter integral

$$\begin{aligned} I_n^{d=4-2\epsilon}\left[(\mu^2)^n;\{p_i\}\right] &= \frac{i}{(4\pi)^{2-\epsilon}} \frac{\Gamma(n-\epsilon)\Gamma(-2+\epsilon)}{\Gamma(-\epsilon)} \\ &\times \int_0^1 \mathrm{d}x_1 \ldots \mathrm{d}x_n\, \delta\left(\sum_{i=1}^{n} x_i - 1\right) \left[\sum_{i<j}^{n} x_i x_j \left(\sum_{k=i+1}^{j} p_k\right)^2\right]^{2-\epsilon} . \end{aligned} \tag{F.1.7}$$

Only in certain special cases ($n = 2$ and $n = 3$) is this integral known to all orders in $\epsilon$ [233]. The leading $\mathcal{O}(\epsilon^0)$ contribution however, can be calculated explicitly for all $n$. It is given by

$$= -\frac{i}{32\pi^2}(n-1)! \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \delta\left(\sum_{i=1}^{n} x_i - 1\right) \left[\sum_{i<j}^{n} x_i x_j \left(\sum_{k=i+1}^{j} p_k\right)^2\right]^2 + \mathcal{O}(\epsilon) . \tag{F.1.8}$$

We now have to perform the integration over the $n$ Feynman parameters. For this we use the general formula

$$\int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \delta\left(\sum_{i=1}^{n} x_i - 1\right) x_1^{r_1} x_2^{r_2} \ldots x_n^{r_n} = \frac{\Gamma(1+r_1)\Gamma(1+r_2)\ldots\Gamma(1+r_n)}{\Gamma(n+r_1+r_2+\ldots+r_n)} . \tag{F.1.9}$$

Special instances of this formula that are relevant for the calculations of this and the next subsection are the following

$$\int_0^1 \mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n\delta\left(\sum_{i=1}^n x_i - 1\right) x_1x_2x_3x_4 = \frac{1}{(n+3)!}\,, \tag{F.1.10}$$

$$\int_0^1 \mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n\delta\left(\sum_{i=1}^n x_i - 1\right) x_1x_2x_3^2 = \frac{2}{(n+3)!}\,, \tag{F.1.11}$$

$$\int_0^1 \mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n\delta\left(\sum_{i=1}^n x_i - 1\right) x_1^2x_2^2 = \frac{4}{(n+3)!}\,, \tag{F.1.12}$$

$$\int_0^1 \mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n\delta\left(\sum_{i=1}^n x_i - 1\right) x_1^3x_2 = \frac{6}{(n+3)!}\,, \tag{F.1.13}$$

With these, we find that the integrated result takes the form

$$\begin{aligned} &I_n^{d=4-2\epsilon}\left[(\mu^2)^n;\{p_i\}\right] \\ &= -\frac{i}{32\pi^2}\frac{1}{n(n+1)(n+2)(n+3)}\sum_{i<j}^n\sum_{k<l}^n a_{ijkl}\left(\sum_{m=i+1}^j p_m\right)^2\left(\sum_{m=k+1}^l p_m\right)^2 + \mathcal{O}(\epsilon)\,, \end{aligned} \tag{F.1.14}$$

where

$$a_{ijkl} = \begin{cases} 1 & \text{if all } i,j,k,l \text{ are different} \\ 2 & \text{if exactly 2 of } i,j,k,l \text{ are identical} \\ 4 & \text{if } i=k \text{ and } j=l \end{cases}. \tag{F.1.15}$$

## F.2 Rational Rank-2 Tensor $n$-gon Integral

Similar to the case of the rational scalar $n$-gon integral, we present the explicit calculation of the rational rank-2 tensor $n$-gon integral

$$I_n^{d=4-2\epsilon}\left[\left(\mu^2\right)^{n-1}(u\cdot l)^2,\{p_i\}\right] \equiv \int \frac{\mathrm{d}^{4-2\epsilon}l}{(2\pi)^{4-2\epsilon}}\frac{\left(l_{-2\epsilon}^2\right)^{n-1}(u\cdot l)^2}{\prod_{i=1}^n\left(l-\sum_{j=1}^i p_j\right)^2}, \tag{F.2.1}$$

where $u^\mu$ is a 4-dimensional null vector. The dimension shifting formula (F.0.1) gives

$$= (4\pi)^{n-1}\frac{\Gamma(n-1-\epsilon)}{\Gamma(-\epsilon)}\int\frac{\mathrm{d}^{2+2n-2\epsilon}l}{(2\pi)^{2+2n-2\epsilon}}\frac{(u\cdot l)^2}{\prod_{i=1}^n\left(l-\sum_{j=1}^i p_j\right)^2}. \tag{F.2.2}$$

We can use the same Feynman parametrization trick as before to write the integral as

$$= (4\pi)^{n-1} \frac{\Gamma(n-1-\epsilon)}{\Gamma(-\epsilon)} (n-1)! \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \delta\left(\sum_{i=1}^n x_i - 1\right) \\ \times \int \frac{\mathrm{d}^{2+2n-2\epsilon} l}{(2\pi)^{2+2n-2\epsilon}} \frac{(u \cdot l)^2}{\left[\sum_{i=1}^n x_i \left(l - \sum_{j=1}^i p_j\right)^2\right]^n} . \tag{F.2.3}$$

After shifting the loop momentum by $l \to l + \sum_{i=1}^{n-1} \sum_{j=1}^i x_i p_j$, we get

$$= (4\pi)^{n-1} \frac{\Gamma(n-1-\epsilon)}{\Gamma(-\epsilon)} (n-1)! \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \delta\left(\sum_{i=1}^n x_i - 1\right) \\ \times \int \frac{\mathrm{d}^{2+2n-2\epsilon} l}{(2\pi)^{2+2n-2\epsilon}} \frac{(u \cdot l)^2 + \left(\sum_{i=1}^{n-1} \sum_{j=1}^i x_i (u \cdot p_j)\right)^2}{\left[l^2 + \Delta\right]^n}, \tag{F.2.4}$$

where $\Delta = \sum_{i<j}^n x_i x_j \left(\sum_{k=i+1}^j p_k\right)^2$ as before and all cross-terms have been dropped since they are odd in $l$. The first term integrates to an expression proportional to $u^2$ which is zero by assumption. The remaining terms have the form of the standard integral (F.1.6), so we can give a general expression for (F.2.1) as a integral over Feynman parameters

$$I_n^{d=4-2\epsilon}\left[\left(\mu^2\right)^{n-1} (u \cdot l)^2, \{p_i\}\right] = \frac{i}{(4\pi)^{2-\epsilon}} \frac{\Gamma(n-1-\epsilon)\Gamma(-1+\epsilon)}{\Gamma(-\epsilon)} \\ \times \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \delta\left(\sum_{i=1}^n x_i - 1\right) \left(\sum_{i=1}^{n-1} \sum_{j=1}^i x_i u \cdot p_j\right)^2 \left[\sum_{i<j}^n x_i x_j \left(\sum_{k=i+1}^j p_k\right)^2\right]^{1-\epsilon} . \tag{F.2.5}$$

As in the scalar case we can give explicit expressions for all $n$ at $\mathcal{O}\left(\epsilon^0\right)$, using the Feynman-parameter integrals (F.1.10) - (F.1.13). With these, we find that the integrated result takes the form

$$I_n^{d=4-2\epsilon}\left[\left(\mu^2\right)^{n-1} (u \cdot l)^2, \{p_i\}\right] = \frac{i}{16\pi^2} \frac{1}{(n-1)n(n+1)(n+2)(n+3)} \\ \times \sum_{i<j}^n \left(\sum_{m=i+1}^j p_m\right)^2 \left[\sum_{k<l}^n 2a_{ijkl} \left(\sum_{m=1}^k u \cdot p_m\right) \left(\sum_{m=1}^l u \cdot p_m\right) + \sum_{k=1}^n b_{ijk} \left(\sum_{m=1}^k u \cdot p_m\right)^2\right] , \tag{F.2.6}$$

where $a_{ijkl}$ is as defined above and

$$b_{ijk} = \begin{cases} 2 & \text{if } i \neq k \text{ and } j \neq k \\ 6 & \text{if } i = k \text{ or } j = k \end{cases} \quad . \tag{F.2.7}$$

# APPENDIX G

# EFT Basis and On-Shell Matrix Elements

Operator redundancies in EFTs arise due to the field reparametrization invariance of physical observables [234]. For example, in Einstein-Maxwell we consider redefinitions of the metric of the form

$$g'_{\mu\nu} \equiv g_{\mu\nu} + c_1 R_{\mu\nu} + c_2 R g_{\mu\nu} + c_3 F_{\mu\rho} F^{\rho}{}_{\nu} + \ldots \tag{G.0.1}$$

where $c_i$ are independent coefficients. In the complete effective action (including all possible terms of all mass dimensions consistent with the assumed symmetries) the effect of such a field redefinition is to shift the Wilson coefficients. By choosing $c_i$ in a particular way, certain operators can be removed from the effective action entirely; these are the so-called redundant operators. One approach to constructing a non-redundant basis of operators is to first enumerate all local operators, then use the most general field reparametrization to remove redundant operators. In this appendix we describe an alternative approach that makes use of on-shell scattering amplitudes methods.

The S-matrix corresponding to the effective action is likewise a physical observable, and independent of the choice of field parametrization. In the tree approximation, gauge invariant effective operators generate Lorentz invariant on-shell matrix elements without kinematic singularities. The on-shell method begins with the observation that there is a one-to-one correspondence between non-redundant gauge invariant local operators and Lorentz invariant local matrix elements [235]. By making use of the spinor-helicity formalism for massless on-shell states [71], it is sometimes more efficient to construct an independent set of the latter. Below we use this correspondence to construct a complete basis for operators coupling gravity to $N$ $U(1)$ gauge fields with up to four derivatives.

The on-shell matrix elements we construct are in the helicity basis. Lorentz invariance is encoded in the requirement that the expressions we construct are rational functions of spinor brackets

$$\langle ij \rangle = \epsilon^{\dot{\alpha}\dot{\beta}} \tilde{\lambda}_{i\dot{\alpha}} \tilde{\lambda}_{j\dot{\beta}}, \qquad [ij] = \epsilon_{\alpha\beta} \lambda_i^{\alpha} \lambda_j^{\beta}. \tag{G.0.2}$$

On-shell matrix elements corresponding to gauge invariant local operators are given by polynomials of spinor brackets; we first construct a basis of monomials satisfying certain physical conditions. The first condition we impose is consistency with the action of the massless little group. Such monomials must scale homogeneously with the correct little group weight determined by the helicities $h_i$ of each of the external states

$$M\left(t\lambda_i, t^{-1}\tilde{\lambda}_i\right) = t^{2h_i} M\left(\lambda_i, \tilde{\lambda}_i\right) . \tag{G.0.3}$$

Here we are scaling the spinors of particle $i$ separately, leaving the remaining spinors unchanged. Since the expressions we are constructing are simply strings of $\tilde{\lambda}$s and $\lambda$s, this constraint is equivalent to the following

$$2h_i = (\text{\# of } \lambda_i) - (\text{\# of } \tilde{\lambda}_i). \tag{G.0.4}$$

This constraint places a lower bound on the mass dimension of the monomial. The minimal dimension monomial we could construct with the correct little group weight for each state contains no anti-holomorphic spinors ($\tilde{\lambda}$) for positive helicity states, no holomorphic spinors ($\lambda$) for negative helicity states and no spinors of either chirality for helicity zero states. As an example, the schematic form of such a minimal dimension monomial

$$M_4\left(1^{+2}, 2^{+1}, 3^{-2}, 4^0\right) \sim \lambda_1^4 \lambda_2^2 \tilde{\lambda}_3^4. \tag{G.0.5}$$

As described above, we need to contract the implicit spinor indices in all inequivalent ways to form a basis of such monomials. The mass dimension of such a string is given simply by $[\lambda] = [\tilde{\lambda}] = 1/2$. In this example the minimal dimension is 5. Non-minimal monomials may be generated by introducing further pairs of spinors $\lambda_i\tilde{\lambda}_i \sim p_i$, which have zero little group weight. In general, for a monomial with $k$ photon states and $m$ graviton states the dimension of the monomial is bounded below as:

$$[M_n] \geq k + 2m. \tag{G.0.6}$$

To connect this to the EFT basis, such a monomial must correspond to the Feynman vertex rule derived from a gauge invariant local operator. Since polarization vectors for Bosonic states are dimensionless, $[\epsilon] = 0$, the mass dimension of the monomial can only arise from powers of momenta generated from derivative interactions. For a local operator with $D$ derivatives the matrix element of $k$ photons and $m$ gravitons has the schematic form

$$M_n\left(\{\epsilon, p\}\right) \sim \epsilon_\gamma^k \epsilon_h^m p^D, \tag{G.0.7}$$

and so the dimension of the monomial is simply

$$[M_n] = D. \tag{G.0.8}$$

Putting these results together we find that the number of photons and gravitons in a local matrix element is bounded above by the number of derivatives in the corresponding local operator

$$D \geq k + 2m. \tag{G.0.9}$$

This also bounds the total number of states $n = k + m$ (since both $k$ and $m$ are non-negative) as $D \geq n$. Our task is now to enumerate all inequivalent monomials for photon and gravitons with $D = 3$ and $D = 4$ and identify the corresponding local operators. Here inequivalent means constructing a basis of monomials that are not related to each other by momentum conservation

$$\sum_{j=1}^{n} \langle ij \rangle [jk] = 0, \tag{G.0.10}$$

or Schouten identities

$$\langle ij \rangle \langle kl \rangle + \langle ik \rangle \langle lj \rangle + \langle il \rangle \langle jk \rangle = 0, \qquad [ij][kl] + [ik][lj] + [il][jk] = 0. \tag{G.0.11}$$

A straightforward (though certainly not optimal) approach to this is to first generate a complete basis of monomials, and then numerically evaluate on sets of randomly generated spinors to find a linearly independent subset.

To construct local operators corresponding to the monomials we can make use of the following replacement rules, for photons:

$$\lambda_\alpha \lambda_\beta \to F^+_{\alpha\beta} \equiv \sigma^{\mu\nu}_{\alpha\beta} F_{\mu\nu}, \qquad \tilde{\lambda}_{\dot{\alpha}} \tilde{\lambda}_{\dot{\beta}} \to F^-_{\dot{\alpha}\dot{\beta}} \equiv \overline{\sigma}^{\mu\nu}_{\dot{\alpha}\dot{\beta}} F_{\mu\nu}, \tag{G.0.12}$$

and for gravitons[1]:

$$\lambda_\alpha \lambda_\beta \lambda_\gamma \lambda_\delta \to W^+_{\alpha\beta\gamma\delta} \equiv \sigma^{\mu\nu}_{\alpha\beta} \sigma^{\rho\sigma}_{\gamma\delta} W_{\mu\nu\rho\sigma}, \quad \tilde{\lambda}_{\dot{\alpha}} \tilde{\lambda}_{\dot{\beta}} \tilde{\lambda}_{\dot{\gamma}} \tilde{\lambda}_{\dot{\delta}} \to W^-_{\dot{\alpha}\dot{\beta}\dot{\gamma}\dot{\delta}} \equiv \overline{\sigma}^{\mu\nu}_{\dot{\alpha}\dot{\beta}} \overline{\sigma}^{\rho\sigma}_{\dot{\gamma}\dot{\delta}} W_{\mu\nu\rho\sigma}, \tag{G.0.13}$$

where $F^\pm$ and $W^\pm$ are the (anti-)self-dual field strength and Weyl tensors respectively. For non-minimal operators there are additional helicity spinors; these must come in pairs with zero net little group weight and so we can replace:

$$\lambda^i_\alpha \tilde{\lambda}^i_{\dot{\alpha}} \to \sigma^\mu_{\alpha\dot{\alpha}} \nabla_\mu \,, \tag{G.0.14}$$

where the derivative acts on the local operator creating state $i$. As an illustrative example, consider

[1]Here we are defining $\sigma^{\mu\nu}_{\alpha\beta} \equiv \frac{i}{4} \epsilon^{\dot{\alpha}\dot{\beta}} \left( \sigma^\mu_{\alpha\dot{\alpha}} \sigma^\nu_{\beta\dot{\beta}} - \sigma^\nu_{\alpha\dot{\alpha}} \sigma^\mu_{\beta\dot{\beta}} \right)$ and $\overline{\sigma}^{\mu\nu}_{\dot{\alpha}\dot{\beta}} \equiv \frac{i}{4} \epsilon^{\alpha\beta} \left( \sigma^\mu_{\alpha\dot{\alpha}} \sigma^\nu_{\beta\dot{\beta}} - \sigma^\nu_{\alpha\dot{\alpha}} \sigma^\mu_{\beta\dot{\beta}} \right)$. Using standard trace identities, we can rewrite the local operators we construct in the more familiar (though less compact) Lorentz vector notation.

the following matrix element

$$\begin{aligned}
&M_4\left(1^{+1},2^{+1},3^{-1},4^{-2}\right)\\
&=[12]^3\langle 34\rangle^2\langle 14\rangle\langle 24\rangle\\
&=(\lambda_1^{\alpha_1}\lambda_1^{\alpha_2})(\lambda_{2\alpha_1}\lambda_{2\alpha_2})(\tilde\lambda_{3\dot\alpha_1}\tilde\lambda_{3\dot\alpha_2})(\tilde\lambda_4^{\dot\alpha_1}\tilde\lambda_4^{\dot\alpha_2}\tilde\lambda_4^{\dot\alpha_3}\tilde\lambda_4^{\dot\alpha_4})(\tilde\lambda_{1\dot\alpha_3}\lambda_1^{\alpha_3})(\tilde\lambda_{2\dot\alpha_4}\lambda_{2\alpha_3})\,.
\end{aligned} \tag{G.0.15}$$

Using the replacement rules given above, this can be generated from the following local operator

$$[12]^3\langle 34\rangle^2\langle 14\rangle\langle 24\rangle \to \epsilon^{\dot\alpha_3\dot\alpha_4}\sigma^\mu_{\alpha_3\dot\alpha_3}\sigma^\nu_{\alpha_4\dot\alpha_4}(\nabla_\mu F^{1+\alpha_1\alpha_2})(\nabla_\nu F^{2+}_{\alpha_1\alpha_2})F^{3-}_{\dot\alpha_1\dot\alpha_2}W^{-\dot\alpha_1\dot\alpha_2\dot\alpha_3\dot\alpha_4} \tag{G.0.16}$$

Here we have used a superscript $F^i$ to indicate that the spin-1 states correspond to distinct $U(1)$ gauge groups. If two or more states with the same helicity correspond to the same $U(1)$ factor, then we must Bose symmetrize over the particle labels in the matrix elements before applying the replacement rules. This generically reduces the number of independent local operators at a given order in the derivative expansion.

Finally we must discuss the constraints of parity conservation. In the spinor-helicity formalism, parity $\mathcal{P}$ acts by interchanging the chirality of the spinors $\lambda_{i\alpha}\leftrightarrow\tilde\lambda_{i\dot\alpha}$, or equivalently interchanging angle and square spinor brackets[2]. A local operator is called parity conserving if it generates local matrix elements that satisfy

$$\mathcal{P}\cdot M_n\left(1^{h_1},2^{h_2},...,n^{h_n}\right)=M_n\left(1^{-h_1},2^{-h_2},...,n^{-h_n}\right). \tag{G.0.17}$$

This means that when constructing a basis of local operators using the method described above, in a parity conserving model the matrix elements $M_n\left(1^{h_1},2^{h_2},...,n^{h_n}\right)$ and $M_n\left(1^{-h_1},2^{-h_2},...,n^{-h_n}\right)$ should not be counted separately, while in a parity non-conserving model they should be.

## G.1 Three-Derivative Operators

In accord with the constraint (G.0.9) the possible, non-redundant, three-derivative operators that generate on-shell matrix elements with $k$-photons and $m$-gravitons have

$$(k,m)\in\{(3,0)\}. \tag{G.1.1}$$

The list of possible matrix elements modulo Schouten and momentum conservation, and the corresponding local operators is:

[2]This definition of parity makes sense only if we write the entire matrix element in terms of spinor brackets. For example, to see that local matrix elements containing a single instance of the Levi-Civita symbol are parity odd we must use the identity $\epsilon^{\mu\nu\rho\sigma}p_{1\mu}p_{2\nu}p_{3\rho}p_{4\sigma}\propto[12]\langle 23\rangle[34]\langle 41\rangle-\langle 12\rangle[23]\langle 34\rangle[41]$.

$(+1, +1, +1)$ :

$$[12][23][31] \to F^{1+}_{\alpha\beta} F^{2+\beta\gamma} F^{3+}{}_{\gamma}{}^{\alpha}. \tag{G.1.2}$$

$(-1, -1, -1)$ :

$$\langle 12\rangle\langle 23\rangle\langle 31\rangle \to F^{1-}_{\dot\alpha\dot\beta} F^{2-\dot\beta\dot\gamma} F^{3-}{}_{\dot\gamma}{}^{\dot\alpha}. \tag{G.1.3}$$

There are two independent, three-derivative local operators. Imposing parity conservation there is only a single independent local operator. Such operators vanish unless all field strength tensors are from distinct $U(1)$ factors. To preserve Bose symmetry of the matrix element we see that the associated Wilson coefficients must be totally antisymmetric in flavor indices.

An equivalent form of the three-derivative effective Lagrangian is

$$\mathcal{L}^{(3)} = a_{ijk} F^i_{\mu\nu} F^{j\nu\rho} F^{k\mu}_{\rho} + b_{ijk} F^i_{\mu\nu} F^{j\nu\rho} \tilde{F}^{k\mu}_{\rho}, \tag{G.1.4}$$

where both $a_{ijk}$ and $b_{ijk}$ are totally antisymmetric. The first operator ($a$) is parity even while the second ($b$) is parity odd.

## G.2 Four-Derivative Operators

The possible, non-redundant, four-derivative operators generate on-shell matrix elements with $k$-photons and $m$-gravitons with

$$(k, m) \in \{(2, 1), (4, 0)\}. \tag{G.2.1}$$

The list of possible matrix elements modulo Schouten and momentum conservation, and the corresponding local operators is :
$(+1, +1, +2)$ :

$$[13]^2[23]^2 \to F^{1+}_{\alpha_1\alpha_2} F^{2+}_{\alpha_3\alpha_4} W^{+\alpha_1\alpha_2\alpha_3\alpha_4}. \tag{G.2.2}$$

$(-1, -1, -2)$ :

$$\langle 13\rangle^2\langle 23\rangle^2 \to F^{1-}_{\dot\alpha_1\dot\alpha_2} F^{2-}_{\dot\alpha_3\dot\alpha_4} W^{-\dot\alpha_1\dot\alpha_2\dot\alpha_3\dot\alpha_4}. \tag{G.2.3}$$

$(+1,+1,+1,+1):$

$$\begin{aligned}
[13]^2[24]^2 &\to F^{1+}_{\alpha_1\alpha_2}F^{3+\alpha_1\alpha_2}F^{2+}_{\alpha_3\alpha_4}F^{4+\alpha_3\alpha_4}\\
[12][23][34][41] &\to F^{1+}_{\alpha_1\alpha_2}F^{2+\alpha_2\alpha_3}F^{3+}_{\alpha_3\alpha_4}F^{4+\alpha_4\alpha_1}\\
[12]^2[34]^2 &\to F^{1+}_{\alpha_1\alpha_2}F^{2+\alpha_1\alpha_2}F^{3+}_{\alpha_3\alpha_4}F^{4+\alpha_3\alpha_4}.
\end{aligned} \tag{G.2.4}$$

$(-1,-1,-1,-1):$

$$\begin{aligned}
\langle 13\rangle^2\langle 24\rangle^2 &\to F^{1-}_{\dot\alpha_1\dot\alpha_2}F^{3-\dot\alpha_1\dot\alpha_2}F^{2-}_{\dot\alpha_3\dot\alpha_4}F^{4-\dot\alpha_3\dot\alpha_4}\\
\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle &\to F^{1-}_{\dot\alpha_1\dot\alpha_2}F^{2-\dot\alpha_2\dot\alpha_3}F^{3-}_{\dot\alpha_3\dot\alpha_4}F^{4-\dot\alpha_4\dot\alpha_1}\\
\langle 12\rangle^2\langle 34\rangle^2 &\to F^{1-}_{\dot\alpha_1\dot\alpha_2}F^{2-\dot\alpha_1\dot\alpha_2}F^{3-}_{\dot\alpha_3\dot\alpha_4}F^{4-\dot\alpha_3\dot\alpha_4}.
\end{aligned} \tag{G.2.5}$$

$(+1,+1,-1,-1):$

$$[12]^2\langle 34\rangle^2 \to F^{1+}_{\alpha_1\alpha_2}F^{2+\alpha_1\alpha_2}F^{3-}_{\dot\alpha_1\dot\alpha_2}F^{4-\dot\alpha_1\dot\alpha_2}. \tag{G.2.6}$$

There are five independent, four-derivative local operators. Imposing parity conservation there are only three independent local operators. An equivalent form of the four-derivative effective Lagrangian is

$$\begin{aligned}
\mathcal{L}^{(4)} = {}&\alpha_{ijkl}F^i_{\mu\nu}F^{j\mu\nu}F^k_{\rho\sigma}F^{l\rho\sigma} + \beta_{ijkl}F^i_{\mu\nu}\tilde{F}^{j\mu\nu}F^k_{\rho\sigma}\tilde{F}^{l\rho\sigma} + \gamma_{ij}F^i_{\mu\nu}F^j_{\rho\sigma}W^{\mu\nu\rho\sigma}\\
&+ \chi_{ijkl}F^i_{\mu\nu}F^{j\mu\nu}F^k_{\rho\sigma}\tilde{F}^{l\rho\sigma} + \omega_{ij}F^i_{\mu\nu}\tilde{F}^j_{\rho\sigma}W^{\mu\nu\rho\sigma}.
\end{aligned} \tag{G.2.7}$$

The first three operators ($\alpha$, $\beta$ and $\gamma$) are parity even, while the remaining two ($\chi$ and $\omega$) are parity odd.

# APPENDIX H

# Corrections to Maxwell Equation

In this appendix we shall review the derivation of (5.2.12). Recall the corrected equation of motion for the gauge field:

$$\nabla_\mu F^{i\mu\nu} = \nabla_\mu \Big( 8\,\alpha_{ijkl} F^{j\mu\nu} F^{k}_{\alpha\beta} F^{l\alpha\beta} + 8\,\beta_{ijkl} \tilde{F}^{j\mu\nu} F^{k}_{\alpha\beta} \tilde{F}^{l\alpha\beta} + 4\,\gamma_{ij} F^{j}_{\alpha\beta} W^{\mu\nu\alpha\beta} \\ + 4\,\left( \chi_{ijkl} \tilde{F}^{j\mu\nu} F^{k}_{\alpha\beta} F^{l\alpha\beta} + \chi_{klij} F^{j\mu\nu} \tilde{F}^{k}_{\alpha\beta} F^{l\alpha\beta} \right) + 4\,\omega_{ij} \tilde{F}^{j}_{\alpha\beta} W^{\mu\nu\alpha\beta} \Big)\,. \tag{H.0.1}$$

For simplicity we label the term in the parentheses on the right-hand side of (5.2.10) by $G^{i\,\mu\nu}$. First note that the anti-symmetry of $F^{\mu\nu}$ allows us to rewrite the equation of motion as

$$\frac{1}{\sqrt{-g}} \partial_\mu \left[ \sqrt{-g}\, F^{i\mu\nu} \right] = \frac{1}{\sqrt{-g}} \partial_\mu \left[ \sqrt{-g}\, G^{i\,\mu\nu} \right]\,. \tag{H.0.2}$$

We expand this equation in power of the coefficients $\alpha,\ \dots\ \omega$. The zeroth- and first-order equations are:

$$\partial_\mu \left[ \sqrt{-g}\, F^{i\mu\nu} \right]^{(0)} = 0 \tag{H.0.3a}$$

$$\partial_\mu \left[ \sqrt{-g}\, F^{i\mu\nu} \right]^{(1)} = \partial_\mu \left[ \sqrt{-g}\, G^{i\,\mu\nu} \right]^{(1)}\,. \tag{H.0.3b}$$

The solution to the zeroth-order equation is the uncorrected Reissner-Nordström solution. We are interested in obtaining the first-order part, which represents the corrections to the background. The derivative may be removed from (H.0.3b) because an additive constant has the same fall-off in $r$ as the solution to (H.0.3a), so we may absorb it into the definition of integration constant in the zeroth-order solution, which is $q$. As a result, we have

$$\left[ \sqrt{-g}\, F^{i\mu\nu} \right]^{(1)} = \left[ \sqrt{-g}\, G^{i\,\mu\nu} \right]^{(1)}\,. \tag{H.0.4}$$

Note that $G^{\mu\nu}$ depends explicitly on $(\,\alpha, \dots, \omega\,)$, so $(G^{\mu\nu})^{(1)}$, which is first-order in the coefficients, depends only on the zeroth-order value of the fields $F^{\mu\nu}$ and $W^{\mu\nu\rho\sigma}$.

In addition to the Maxwell equation, the gauge fields must satisfy the Bianchi identity

$$\partial_\mu F^i_{\nu\rho} + \partial_\nu F^i_{\rho\mu} + \partial_\rho F^i_{\mu\nu} = 0. \tag{H.0.5}$$

Together with the assumed spherically symmetry, which imposes that only $F^i_{tr}$ and $F^i_{\theta\phi}$ are non-zero, this gives the following constraint on the *magnetic* component of the gauge field

$$\partial_r F^i_{\theta\phi} = 0. \tag{H.0.6}$$

Since the leading order magnetic field (5.2.6) is the unique spherically symmetric field with magnetic monopole moment $p^i$, and by (H.0.6) there can be no subleading $1/r$ corrections, it remains the exact solution even with the addition of higher-derivative interactions. Therefore we are only interested in the corrections to the electric fields $F^{(i)}_{tr}$. Using that $g^0_{tt} = -g^0_{rr}$, we have

$$\left[\sqrt{-g}F^{i\,tr}\right]^{(1)} \;=\; \sqrt{-g}^{(0)}\left(8\alpha_{ijkl}F^{(0)j}{}_{tr}F^{(0)k}{}_{tr}F^{(0)l}{}_{tr} + ...\right). \tag{H.0.7}$$

Now we may use this to compute the first contribution to the stress tensor corrections. This relies on the non-trivial fact that this combination of $\sqrt{-g}$ and $F$ is the only combination that appears in the corrections to the stress tensor. To see this consider the stress tensor for a Maxwell field,

$$T_{\mu\nu} = F^i_{\mu\alpha}F^{i\alpha}_{\nu} - \frac{1}{4}F^i_{\alpha\beta}F^{i\alpha\beta}g_{\mu\nu}. \tag{H.0.8}$$

We are interested only in the corrections to

$$T_t{}^t = F^i{}_{t\alpha}F^{it\alpha} - \frac{1}{4}F^i{}_{\alpha\beta}F^{i\alpha\beta}\delta_t{}^t\,. \tag{H.0.9}$$

We use the fact that only $F_{tr}$ and $F_{\theta\phi}$ are non-zero, and only the former is corrected, to write

$$\begin{aligned} T_t{}^t &= \frac{1}{2}F^i{}_{tr}F^{itr} - \frac{1}{2}F^i{}_{\theta\phi}F^{i\theta\phi} \\ &= (T^{(0)})_t{}^t - \left[\sqrt{-g}F^{itr}\right]^{(1)}\left[\sqrt{-g}F^{itr}\right]^{(0)}/(g_{\theta\theta}g_{\phi\phi}) + \mathcal{O}\left[(\alpha,...)^2\right]\,. \end{aligned} \tag{H.0.10}$$

So we have found that

$$\begin{aligned} (T^{(1)}_{Max})_t{}^t &= -\left[\sqrt{-g}F^{itr}\right]^{(1)}\left[\sqrt{-g}F^{itr}\right]^{(0)}/(g_{\theta\theta}g_{\phi\phi}) \\ &= -\sqrt{-g}^{(0)}\left(8\alpha_{ijkl}F^{(0)j}{}_{tr}F^{(0)k}{}_{tr}F^{(0)l}{}_{tr} + ...\right)\sqrt{-g}^{(0)}F^{itr\,(0)}/(g_{\theta\theta}g_{\phi\phi}) \\ &= \left(8\alpha_{ijkl}F^{(0)j}{}_{tr}F^{(0)k}{}_{tr}F^{(0)l}{}_{tr} + ...\right)F^i{}_{tr}{}^{(0)}\,. \end{aligned} \tag{H.0.11}$$

Evaluating this expression gives the result obtained in (5.2.12).

# APPENDIX I

# Variations of Four-Derivative Operators

In Section (5.2.2), we computed the shift to the geometry by first computing the shift to the stress tensor due to the presence of higher-derivative operators. One source of stress tensor corrections comes from varying the four-derivative operators with respect to the metric. The variations of each of these terms are recorded here for reference.

$$
\begin{aligned}
(F^iF^j)(F^kF^l): \quad & g_{\alpha\beta}(F^iF^j)(F^k\cdot F^l) - 4\left(F^i_{\mu\alpha}F^{j\mu}{}_{\beta}(F^kF^l) + (F^iF^j)F^k_{\mu\alpha}F^{l\mu}{}_{\beta}\right) \\
(F^i\tilde{F}^j)(F^k\tilde{F}^l): \quad & -g_{\alpha\beta}(F^i\tilde{F}^j)(F^k\tilde{F}^l) \\
WF^iF^j: \quad & g_{\alpha\beta}WF^iF^j - 3R^{\mu}{}_{\alpha\rho\sigma}(F^i_{\mu\beta}F^{j\rho\sigma} + F^{i\rho\sigma}F^j_{\mu\beta}) + 4R_{\alpha\mu}(F^i_{\beta\nu}F^{j\mu\nu} + F^{i\mu\nu}F^j_{\beta\nu}) \\
& + 4R_{\mu\nu}F^{i\mu}{}_{\alpha}F^{j\nu}{}_{\beta} - \frac{4}{3}RF^i_{\alpha\mu}F^{j\mu}_{\beta} - \frac{2}{3}R_{\alpha\beta}(F^iF^j) \\
& - 4\nabla_\mu\nabla_\nu(F^{i\mu}{}_{\alpha}F^{j\nu}{}_{\beta}) - 4\nabla_\mu\nabla_\alpha(F^{i\mu}{}_{\nu}F^{j\nu}_{\beta}) + 2g_{\alpha\beta}\nabla_\mu\nabla_\nu(F^{i\mu}{}_{\rho}F^{j\nu\rho}) \\
& + 2\Box(F^i_{\alpha\mu}F^{j\mu}_{\beta}) + \frac{2}{3}\nabla_\alpha\nabla_\beta(F^iF^j) - \frac{2}{3}g_{\alpha\beta}\Box(F^iF^j) \\
(F^i\tilde{F}^j)(F^kF^l): \quad & -4(F^i\tilde{F}^j)F^k_{\mu\alpha}F^{l\mu}{}_{\beta} \\
WF^i\tilde{F}^j: \quad & -2R^{\mu}{}_{\alpha\rho\sigma}F^i_{\mu\beta}\tilde{F}^{j\rho\sigma} + 4R_{\alpha\mu}F^i_{\beta\nu}\tilde{F}^{j\mu\nu} - \frac{2}{3}R_{\alpha\beta}(F^i\tilde{F}^j) \\
& - 4\nabla_\mu\nabla_\nu(F^{i\mu}{}_{\alpha}\tilde{F}^{j\nu}{}_{\beta}) - 4\nabla_\mu\nabla_\alpha(F^{i\mu}{}_{\nu}\tilde{F}^{j\nu}_{\beta}) + 2g_{\alpha\beta}\nabla_\mu\nabla_\nu(F^{i\mu}{}_{\rho}\tilde{F}^{j\nu\rho}) \\
& + 2\Box(F^i_{\alpha\mu}\tilde{F}^{j\mu}_{\beta}) + \frac{2}{3}\nabla_\alpha\nabla_\beta(F^i\tilde{F}^j) - \frac{2}{3}g_{\alpha\beta}\Box(F^i\tilde{F}^j)
\end{aligned}
\tag{I.0.1}
$$

Each of the terms on the left-hand side are multiplied by $\sqrt{-g}$ in the action. Note that we use the shorthand $(F^iF^j)$ to denote $F^i_{\mu\nu}F^{j\mu\nu}$, and $WAB$ to denote $W_{\mu\nu\rho\sigma}A^{\mu\nu}B^{\rho\sigma}$.

# APPENDIX J

# Proof of Convexity of the Extremality Surface

In this appendix we give a short proof of the claim made in Section 5.3, that in the perturbative regime, $Q^2 \gg 1$, the extremality surface bounds a convex region. Though convexity is a global property, we can reduce the problem to a local one through the *Tietze-Nakajima theorem* [236]: if $X \subset \mathbb{R}^n$ is closed, connected and *locally convex*, then $X$ is convex. Here local convexity means that for each $x \in X$, for some $\delta > 0$ the set $B_\delta(x) \cap X$ is convex.

Since the requirements of closure and connectedness are trivial for the kinds of regions we are considering, it remains to show that the extremality surface is the boundary of a locally convex set. The key idea of the argument is to show that on a sufficiently small neighborhood of any point, the surface is well approximated by an inverted paraboloid up to $\mathcal{O}(1/Q^2)$ corrections. Local convexity is then a consequence of the convexity of the paraboloid hypograph.

Consider a general co-dimension-1 hypersurface $X$ embedded in $\mathbb{R}^n$, defined by an equation of the form

$$\sum_{i=1}^{n} x_i^2 = 1 + T(x_i), \tag{J.0.1}$$

where $T(x_i)$ is *small* in the sense that

$$\left| \sum_{i=1}^{n} x_i^2 - 1 \right| < \epsilon, \tag{J.0.2}$$

for all points $x_i \in X$, for some arbitrarily small $\epsilon > 0$. Since this condition is preserved under orthogonal rotations, every point on $X$ can be mapped to $x_i = 0$ for $i > 1$ up to a redefinition of the function $T(x_i)$. Without loss of generality then we will study the local neighbourhood of such a point. We use the fact that we are interested in functions of the form

$$T(x_i) = \sum_{ijkl} T_{ijkl} x_i x_j x_k x_l \,. \tag{J.0.3}$$

Here the smallness condition (J.0.2) is equivalent to the statement that $|T_{ijkl}| \sim \epsilon$. To begin with we can rewrite the equation (J.0.1) in a useful form

$$x_1^2 = 1 - \sum_{i\neq 1} x_i^2 + T_{1111}x_1^4 + 4x_1^3 \sum_i T_{111i}x_i + 6x_1^2 \sum_{ij\neq 1} T_{11ij}x_ix_j$$
$$+ 4x_1 \sum_{ijk\neq 1} T_{1ijk}x_ix_jx_k + \sum_{ijkl\neq 1} T_{ijkl}x_ix_jx_kx_l. \quad \text{(J.0.4)}$$

At $x_i = 0$, $i > 1$, for small $\epsilon$ there is a single value of $x_1 > 0$ on $X$. Since we are interested in the surface on an arbitrarily small convex neighbourhood $D$ of $x_i = 0$, $i > 1$, we can construct a local parametrization of the surface as a function $x_1 : D \to \mathbb{R}$

$$x_1(x_2, ..., x_n) = 1 - \frac{1}{2}\sum_{i\neq 1} x_i^2 + \frac{1}{2}T_{1111} + \frac{1}{2}T_{1111}\sum_{i\neq 1} x_i^2 + 3\sum_i T_{111i}x_i + 3\sum_{i,j\neq 1} T_{11ij}x_ix_j + \mathcal{O}(x_i^3).$$
(J.0.5)

It is an elementary theorem that the *hypograph* of a function $f : D \to \mathbb{R}$, with $D$ a convex set in $\mathbb{R}^{n-1}$, is a convex set in $\mathbb{R}^n$ if the Hessian of $f$ is negative definite on the interior of $D$. From (J.0.5) we can read off the eigenvalues of the Hessian matrix at this point as $-1 + \mathcal{O}(\epsilon)$. Since the eigenvalues of the Hessian are continuous on $X$ they must all be negative on some neighbourhood of this point. This completes the proof that $X$ is locally convex.

# APPENDIX K

# Entropy Shifts from the On-Shell Action

In Section (6.5), we computed the constraints on the coefficients in $\mathrm{AdS}_4$. Here we will present the results of this calculation for $\mathrm{AdS}_5$ through $\mathrm{AdS}_7$ using the entropies computed in Section (6.3), which corresponds to working in the zero Casimir energy scheme. For completeness, we also present the Casimir energies for $\mathrm{AdS}_5$ and $\mathrm{AdS}_7$ that show up in the thermodynamic energy of Section (6.4) when using a minimal set of counterterms.

## K.1 $\mathrm{AdS}_5$

In $\mathrm{AdS}_5$ we find that the stability condition obtained by demanding positive specific heat and permittivity is given by $\xi < \xi^*$ for $\nu < \nu^*$, with

$$\xi^* = 1 - \sqrt{\frac{1-2\nu^2}{1+2\nu^2}}, \qquad \nu^* = \frac{1}{\sqrt{2}}\,, \tag{K.1.1}$$

and that all black holes with $\nu > \nu^*$ are stable for all values of the charge. The full entropy shift is simpler to express as a function of charge $q$ than extremality parameter $\xi$. We find

$$\begin{aligned}\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} &= \frac{\pi}{256 l^6 \nu^8 T}\Big( c_1\left(43q^4 - 24l^4q^2\nu^4(8+5\nu^2) + 32l^8\nu^8(18+41\nu^2+13\nu^4)\right) \\ &\qquad + 24c_2q^2\left(3q^2 - 8l^4\nu^4\right) + 72(2c_3+c_4)q^4\Big)\,.\end{aligned} \tag{K.1.2}$$

Note that holographic renormalization in $\mathrm{AdS}_5$ with a Riemann-squared correction yields a Casimir energy

$$E_c = \frac{\omega_3}{16\pi}\left(\frac{3}{4}l^2 - \frac{15}{4}c_1 l^2\right), \tag{K.1.3}$$

where $\omega_3 = 2\pi^2$. This Casimir energy must be removed from the thermodynamic energy in order to obtain the mass $M$ of the black hole. Alternatively, it can be cancelled right from the beginning

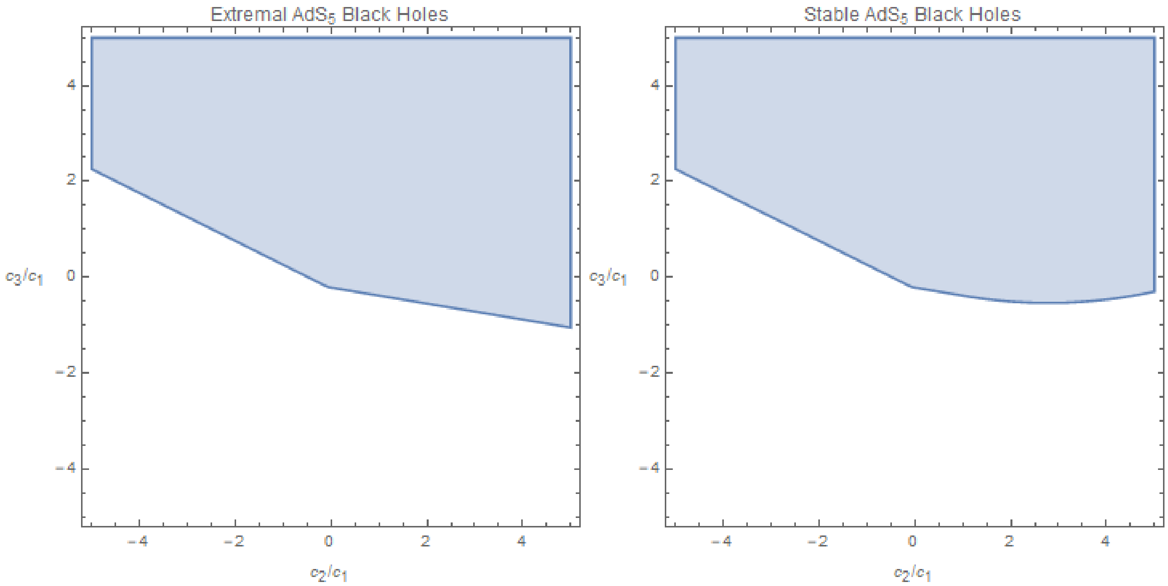



Figure K.1: Allowed regions for $AdS_5$ EFT coefficients.

by adding an appropriate finite counterterm to the action, in which case the thermodynamic energy would then correspond directly to the mass. If the Casimir energy is not removed, then the thermodynamic energy shift becomes a combination of mass shift and Casimir energy shift since $E_c$ depends explicitly on the $c_1$ Wilson coefficient.

We find the following expression for the extremal limit,

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} = \frac{\pi l^2}{16T}\left(c_1(31+128\nu^2+138\nu^4)+24c_2(1+2\nu^2)(1+6\nu^2)+72(2c_3+c_4)(1+2\nu^2)^2\right), \tag{K.1.4}$$

while in the neutral limit we have

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} = \frac{\pi l^2}{16T} c_1\left(18+41\nu^2+13\nu^4\right) . \tag{K.1.5}$$

Once again, the entropy shift is proportional to $c_1$ in this limit. It is interesting that we do not find a positivity constraint on $c_2$, as we did in $AdS_4$. There is a lower bound on $c_3/c_1$ of about -0.5339. The general constraints obtained by the Reduce function of Mathematica are extremely complicated and probably of little interest.

## K.2 AdS$_6$

In AdS$_6$ the stability condition obtained by demanding positive specific heat and permittivity is of the same general structure as in AdS$_5$, but with the following identifications:

$$\xi^* = 1 - \sqrt{\frac{3-5\nu^2}{3+5\nu^2}}, \qquad \nu^* = \sqrt{\frac{3}{5}}\,. \tag{K.2.1}$$

The entropy shift is given by:

$$\begin{aligned}\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} &= \frac{\pi}{264 l^9 \nu^{11} T}\Big( c_1\left(189 q^4 - 22 l^6 q^2 \nu^6 (36+29\nu^2) + 264 l^{12}\nu^{12}(8+17\nu^2+7\nu^4)\right) \\ &\quad + 2c_2 q^2\left(153 q^2 - 44 l^6\nu^6(9+5\nu^2)\right) + 288(2c_3+c_4)q^4\Big)\,,\end{aligned} \tag{K.2.2}$$

and in the extremal limit takes the form:

$$\begin{aligned}&\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} \\ &= \frac{2\nu\pi l^3}{99T}\left(c_1(369+1263\nu^2+1124\nu^4) + 4c_2(3+5\nu^2)(27+100\nu^2) + 96(2c_3+c_4)(3+5\nu^2)^2\right)\,.\end{aligned} \tag{K.2.3}$$

Finally, in the neutral limit we find

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} = \frac{\nu\pi l^3}{T} c_1\left(8+17\nu^2+7\nu^4\right)\,. \tag{K.2.4}$$

Note that no Casimir energy subtraction is needed in AdS$_6$. We again find that $c_1$ is positive. The other bounds are displayed in figure K.2. In AdS$_6$ and AdS$_7$, the Reduce function of Mathematica was not able to find the general constraints over all stable values of $\xi$ and $\nu$. However, we believe that the strongest constraints will come from the boundaries of the region of stable black holes. Specifically, we imposed positivity at the neutral $\xi \to 1$ limit, the extremal $\xi \to 0$ limit, the planar limit $\nu \to \infty$ limit, and at $\xi = \xi^*$. We believe this method should give the same answer, and we have checked explicitly that it does in the case for AdS$_4$ and AdS$_5$.

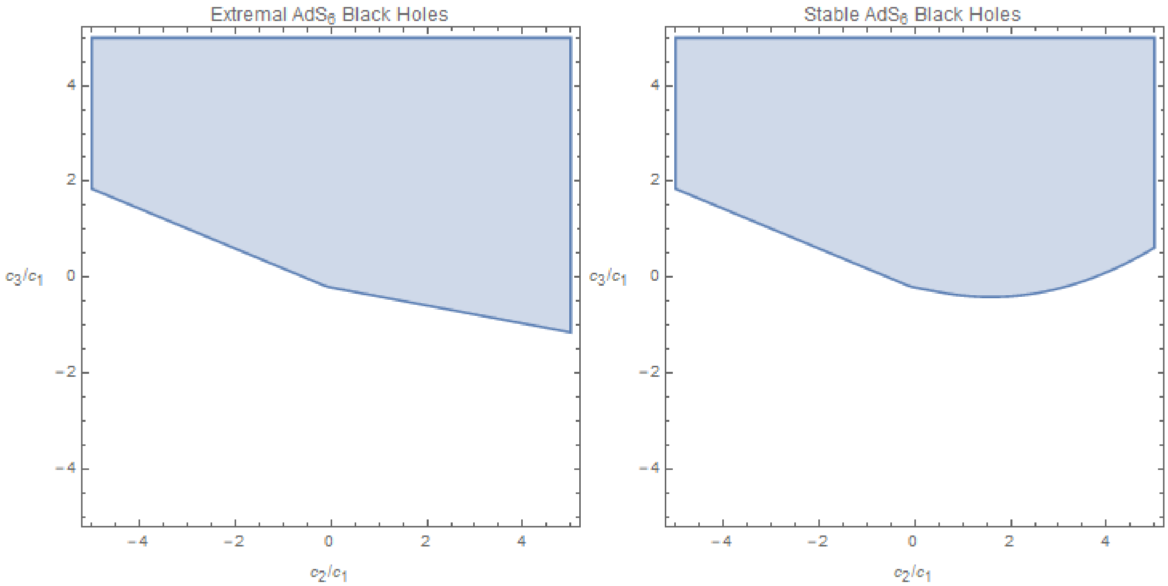



Figure K.2: Allowed regions for AdS$_6$ EFT coefficients.

## K.3 AdS$_7$

In AdS$_7$ the stability window is determined by

$$\xi^* = 1 - \sqrt{\frac{2-3\nu^2}{2+3\nu^2}}, \qquad \nu^* = \sqrt{\frac{2}{3}}, \tag{K.3.1}$$

and the entropy shift is:

$$\begin{aligned}
&\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} \\
&= \frac{\pi^2}{896 l^{12} \nu^{14} T}\Big( c_1 \left(556 q^4 - 14 q^2 l^8 \nu^8 (160 + 141\nu^2) + 56 l^{16} \nu^{16} (100 + 207\nu^2 + 8\nu^4)\right) \\
&\qquad + 80 c_2 q^2 \left(11 q^2 - 7 l^8 \nu^8 (4 + 3\nu^2)\right) + 800(2c_3 + c_4) q^4 \Big) .
\end{aligned}$$

The Casimir energy that must be removed from the thermodynamic energy in AdS$_7$ is

$$E_c = \frac{\omega_5}{16\pi}\left(-\frac{5}{8} l^4 + \frac{35}{8} c_1 l^4\right), \tag{K.3.2}$$

where $\omega_5 = \pi^3$.

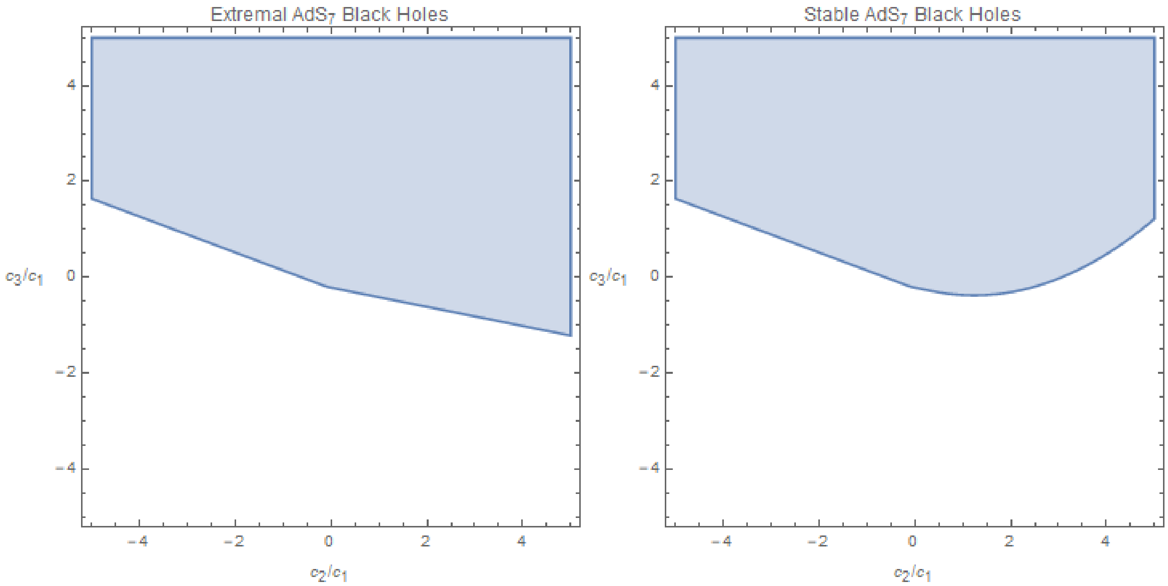



Figure K.3: Allowed regions for AdS$_7$ EFT coefficients.

We find the following expression for the extremal limit,

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} = \frac{\pi^2\nu^2 l^4}{224T}\Bigg(c_1\left(1384+4236\nu^2+3345\nu^4\right) + 40c_2(2+3\nu^2)(16+45\nu^2)+800(2c_3+c_4)(2+3\nu^2)^2\Bigg)\,, \tag{K.3.3}$$

while in the neutral limit we find

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{M,Q} = \frac{\pi^2 l^2\nu^2}{16T}c_1\left(100+207\nu^2+93\nu^4\right)\,. \tag{K.3.4}$$

Once again, $c_1$ is positive. The other bounds are displayed in figure K.3. Again, we used the method of extremizing over the boundaries of the space of stable black holes.

# APPENDIX L

# Another Proof of the Entropy-Extremality Relation

Recent work [49, 186] suggests a remarkable universal relationship between the corrections to extremality and corrections to entropy. Here we will present a simple derivation of this relation using standard thermodynamic identities, including a slight generalization of the relation away from extremality. The statement itself is not specific to black holes, and is in fact a relatively universal statement about infinitesimal deformations of thermodynamic systems.

Consider a thermodynamic system, let $E$ be the total thermal energy, $T$ the temperature, $S$ the entropy and $X$ collectively label a set of extensive thermodynamic variables (for black holes this could be the charge $Q$ and spin $J$). Now consider a small deformation of this system parametrized by a continuous parameter $\epsilon$. The only assumption we will make about this deformation is that it preserves the third law of thermodynamics in the form

$$\lim_{T\to 0} TS(T,X,\epsilon) = 0, \tag{L.0.1}$$

for all $\epsilon$ on an open neighbourhood of $\epsilon = 0$.

We begin with the first law of thermodynamics in the form

$$1 = T\left(\frac{\partial S}{\partial E}\right)_{X,\epsilon}. \tag{L.0.2}$$

Making use of the triple product identity

$$\left(\frac{\partial S}{\partial E}\right)_{X,\epsilon}\left(\frac{\partial E}{\partial \epsilon}\right)_{X,S}\left(\frac{\partial \epsilon}{\partial S}\right)_{X,E} = -1, \tag{L.0.3}$$

we have

$$\left(\frac{\partial E}{\partial \epsilon}\right)_{X,S} = -T\left(\frac{\partial S}{\partial \epsilon}\right)_{X,E}. \tag{L.0.4}$$

Formally inverting $S(T, X, \epsilon)$ gives $T(S, X, \epsilon)$. We can use this to write

$$\left(\frac{\partial E}{\partial \epsilon}\right)_{X,S} = \left(\frac{\partial E}{\partial \epsilon}\right)_{X,T} + \left(\frac{\partial E}{\partial T}\right)_{X,\epsilon} \left(\frac{\partial T}{\partial \epsilon}\right)_{X,S} . \tag{L.0.5}$$

Combining these

$$\left(\frac{\partial E}{\partial \epsilon}\right)_{X,T} = -T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,E} - \left(\frac{\partial E}{\partial T}\right)_{X,\epsilon} \left(\frac{\partial T}{\partial \epsilon}\right)_{X,S} . \tag{L.0.6}$$

Next, we use (L.0.2) again

$$\begin{aligned}
\left(\frac{\partial E}{\partial \epsilon}\right)_{X,T} &= -T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,E} - \left(\frac{\partial E}{\partial T}\right)_{X,\epsilon} \left(\frac{\partial T}{\partial \epsilon}\right)_{X,S} \\
&= -T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,E} - T \left(\frac{\partial S}{\partial E}\right)_{X,\epsilon} \left(\frac{\partial E}{\partial T}\right)_{X,\epsilon} \left(\frac{\partial T}{\partial \epsilon}\right)_{X,S} \\
&= -T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,E} - T \left(\frac{\partial S}{\partial T}\right)_{X,\epsilon} \left(\frac{\partial T}{\partial \epsilon}\right)_{X,S} ,
\end{aligned} \tag{L.0.7}$$

one final application of the triple product identity gives the *generalized Entropy-Extremality relation*

$$\left(\frac{\partial E}{\partial \epsilon}\right)_{X,T} + T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,E} = T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,T} . \tag{L.0.8}$$

Next we make use of the assumption that the deformation does not violate the third law of thermodynamics. Taylor expanding (L.0.2) we have

$$\lim_{T\to 0} \left[ TS(T, X, \epsilon = 0) + \epsilon T \left(\frac{\partial S}{\partial \epsilon}\right)_{T,X}\Bigg|_{\epsilon=0} + \mathcal{O}\left(\epsilon^2\right) \right] = 0. \tag{L.0.9}$$

By assumption this is true on an open neighbourhood of $\epsilon = 0$ and so must be true order-by-order in the expansion, this gives

$$\lim_{T\to 0} T \left(\frac{\partial S}{\partial \epsilon}\right)_{T,X}\Bigg|_{\epsilon=0} = 0\,. \tag{L.0.10}$$

Using this together with (L.0.8) gives the Entropy-Extremality (or *Goon-Penco*) relation

$$\boxed{\lim_{T\to 0} \left[ \left(\frac{\partial E}{\partial \epsilon}\right)_{X,T}\Bigg|_{\epsilon=0} + T \left(\frac{\partial S}{\partial \epsilon}\right)_{X,E}\Bigg|_{\epsilon=0} \right] = 0.} \tag{L.0.11}$$

For the specific application to black hole thermodynamics we identify $E$ with the mass $M$ of the black hole, $X$ with the black hole parameters measured at infinity such as charge $Q$ or angular momentum $J$, and $\epsilon$ with a Wilson coefficient of a four-derivative effective operator.

In Section (6.3), we have pointed out that shift in charge at fixed mass is also proportional to the entropy shift and mass shift. This statement can be derived similarly. By the triple product identity,

$$\left(\frac{\partial E}{\partial \epsilon}\right)_{X_i,T} = -\left(\frac{\partial X_i}{\partial \epsilon}\right)_{E,T}\left(\frac{\partial E}{\partial X_i}\right)_{\epsilon,T} . \tag{L.0.12}$$

This holds for any extensive quantity. Now we choose $X_i = Q$, and we may identify

$$\left(\frac{\partial E}{\partial X_i}\right)_{\epsilon,T} = \Phi. \tag{L.0.13}$$

So we find

$$\left(\frac{\partial E}{\partial \epsilon}\right)_{Q,T} = -\Phi\left(\frac{\partial Q}{\partial \epsilon}\right)_{E,T} . \tag{L.0.14}$$

For black holes, this means that the shift in charge is related to the shift in mass. Neither of them is related to the entropy except at extremality. The result of this is that the entropy shift at extremality may be related to the extremality shift at constant charge or at constant mass,

$$\lim_{T\to 0}\left(\frac{\partial E}{\partial \epsilon}\right)_{Q,T} = -\lim_{T\to 0}\Phi\left(\frac{\partial Q}{\partial \epsilon}\right)_{E,T} = -\lim_{T\to 0}T\left(\frac{\partial S}{\partial \epsilon}\right)_{Q,E} . \tag{L.0.15}$$

## L.1 Comment on $\alpha'$-Corrections to Black Holes in Heterotic String Theory

Recent work has considered the leading $\alpha'$-corrections to dyonic Reissner-Nordström black holes embedded in heterotic string theory [237, 238]. Though the four-dimensional backgrounds considered in these papers are asymptotically flat, we would like to briefly comment on them in connection with the universal entropy-extremality relationship.

From the dimensionally reduced, effective four-dimensional solutions the authors calculated explicit expressions for the Wald entropy,

$$S = \pi\left[(\mu + M)^2 + \epsilon\alpha' \frac{(18M\mu + 21\mu^2 + M^2)}{40\mu\,(\mu + M)}\right], \tag{L.1.1}$$

and Hawking temperature,

$$T = \frac{\mu}{2\pi\,(\mu + M)^2} + \epsilon\alpha' \frac{(M + 3\mu)\,(M - \mu)^2}{160\pi\mu\,(\mu + M)^5}, \tag{L.1.2}$$

where we have defined $\mu = \sqrt{M^2 - \frac{P^2}{2}}$. Here $P$ denotes the charge of the black hole, and we have adopted the same small expansion parameter $\epsilon$ as earlier. From these results it is straightforward to verify the following relation

$$T^{-1} = \left(\frac{\partial S}{\partial M}\right)_P , \tag{L.1.3}$$

up to errors of $\mathcal{O}(\epsilon^2)$. Consequently, the parameter $M$ corresponds to the thermal mass of the black hole. With these explicit expressions we can verify the entropy-extremality relation derived in [186]. The differential change in the mass at fixed temperature is given by a simple application of the triple product identity,

$$\left(\frac{\partial M}{\partial \epsilon}\right)_{T,P} = - \frac{\left(\frac{\partial T}{\partial \epsilon}\right)_{M,P}}{\left(\frac{\partial T}{\partial M}\right)_{\epsilon,P}}\Bigg|_{\epsilon=0} . \tag{L.1.4}$$

The differential change in the *extremal* mass is given by taking $T \to 0^+$. This limit must be taken indirectly since the function (L.1.2) is too complicated to be inverted directly. The zero temperature limit is then the same as taking $M \to M_{\text{ext}}$. Since (L.1.4) is already a relation between two quantities at $\mathcal{O}(\alpha')$, we only require $M \to (M_{\text{ext}})_0$, which is the leading-order extremality relationship

$$(M_{\text{ext}})_0 = \frac{|P|}{\sqrt{2}}. \tag{L.1.5}$$

The correction to the extremal mass is then found to be

$$\left(\frac{\partial M}{\partial \epsilon}\right)_{T=0,P} = - \frac{\alpha'}{40\sqrt{2}|P|}, \tag{L.1.6}$$

in agreement with [238]. To verify the entropy-extremality relation we also need the shift to the entropy at fixed charge and mass. Since the Wald entropy given above is already parametrized in terms of the thermal mass and charge, this is trivial to calculate,

$$\left(\frac{\partial S}{\partial \epsilon}\right)_{M,P} = \alpha' \pi \frac{(M^2 + 18M\mu + 21\mu^2)}{40\mu\,(\mu + M)}. \tag{L.1.7}$$

According to [186], we should take the zero temperature limit of this expression multiplied by the uncorrected temperature. This is equivalent to taking $M \to (M_{\text{ext}})_0$. Indeed, taking this limit we find

$$\lim_{T\to 0} T_0 \left(\frac{\partial S}{\partial \epsilon}\right)_{M,P} = \lim_{M\to (M_{\text{ext}})_0} \lim_{\epsilon \to 0} T(M,P,\epsilon) \left(\frac{\partial S}{\partial \epsilon}\right)_{M,P} = \frac{\alpha'}{40\sqrt{2}|P|} , \tag{L.1.8}$$

which agrees precisely with the shift to the mass in (L.1.6), verifying the entropy-extremality relationship. While (L.1.8) was also reproduced in [238], the authors suggest that it may be more convenient to work in a near-extremal regime, in which the result of [186] would be modified. In particular, in the "very near-extremal" regime discussed in [238], in which one extracts the leading order terms in the near extremal temperature and near extremal entropy (both contributions of order $\sqrt{\epsilon}$), one finds $\frac{\alpha'}{80\sqrt{2}|P|}$, which differs from (L.1.8) precisely by a factor of $1/2$.

Such issues are related to the claim of [238] that the positivity of the entropy shift doesn't necessarily imply a positive correction to the charge to mass ratio at extremality. Here we would like to further clarify the validity of the results of [186], and stress that the order of limits was *crucial* in order to obtain (L.1.8). Indeed, if the limits were taken in the other order, we would find

$$\lim_{\epsilon\to 0}\lim_{M\to M_{\text{ext}}} T(M,P,\epsilon)\left(\frac{\partial S}{\partial \epsilon}\right)_{M,P} = \frac{\alpha'}{60\sqrt{2}|P|}, \tag{L.1.9}$$

which does not agree with the correction to the extremal mass. The near-extremal computation that yields the $1/2$ factor mentioned above is yet another way to compute $T\Delta S$, which not surprisingly leads to a different value. The final result is indeed extremely sensitive to the way in which one approaches extremality. Nonetheless, the results of [186] are valid in general, provided the extremality limit is taken in a precise way, as described by (L.1.8).

# BIBLIOGRAPHY


[1] Cheung, C., Kampf, K., Novotny, J., Shen, C.-H., Trnka, J., and Wen, C., "Vector Effective Field Theories from Soft Limits," *Phys. Rev. Lett.*, Vol. 120, No. 26, 2018, pp. 261602.

[2] Feynman, R. P., "Space - time approach to quantum electrodynamics," *Phys. Rev.*, Vol. 76, 1949, pp. 769–789, [,99(1949)].

[3] 't Hooft, G. and Veltman, M. J. G., "Regularization and Renormalization of Gauge Fields," *Nucl. Phys.*, Vol. B44, 1972, pp. 189–213.

[4] Weinberg, S., "A Model of Leptons," *Phys. Rev. Lett.*, Vol. 19, 1967, pp. 1264–1266.

[5] Gross, D. J. and Wilczek, F., "Ultraviolet Behavior of Nonabelian Gauge Theories," *Phys. Rev. Lett.*, Vol. 30, 1973, pp. 1343–1346, [,271(1973)].

[6] Politzer, H. D., "Reliable Perturbative Results for Strong Interactions?" *Phys. Rev. Lett.*, Vol. 30, 1973, pp. 1346–1349, [,274(1973)].

[7] Dyson, F. J., "Divergence of perturbation theory in quantum electrodynamics," *Phys. Rev.*, Vol. 85, 1952, pp. 631–632.

[8] Dixon, L. J., "Calculating scattering amplitudes efficiently," *QCD and beyond. Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics, TASI-95, Boulder, USA, June 4-30, 1995*, 1996, pp. 539–584.

[9] Buttar, C. et al., "Les houches physics at TeV colliders 2005, standard model and Higgs working group: Summary report," *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 2-20, 2005*, 2006.

[10] Xu, Z., Zhang, D.-H., and Chang, L., "Helicity Amplitudes for Multiple Bremsstrahlung in Massless Nonabelian Gauge Theories," *Nucl. Phys.*, Vol. B291, 1987, pp. 392–428.

[11] Britto, R., Cachazo, F., Feng, B., and Witten, E., "Direct proof of tree-level recursion relation in Yang-Mills theory," *Phys. Rev. Lett.*, Vol. 94, 2005, pp. 181602.

[12] Bern, Z., Dixon, L. J., and Kosower, D. A., "One loop corrections to five gluon amplitudes," *Phys. Rev. Lett.*, Vol. 70, 1993, pp. 2677–2680.

[13] Bern, Z., Dixon, L. J., Dunbar, D. C., and Kosower, D. A., "Fusing gauge theory tree amplitudes into loop amplitudes," *Nucl. Phys.*, Vol. B435, 1995, pp. 59–101.

[14] Ossola, G., Papadopoulos, C. G., and Pittau, R., "Reducing full one-loop amplitudes to scalar integrals at the integrand level," *Nucl. Phys.*, Vol. B763, 2007, pp. 147–169.

[15] Witten, E., "Perturbative gauge theory as a string theory in twistor space," *Commun. Math. Phys.*, Vol. 252, 2004, pp. 189–258.

[16] Nair, V. P., "A Current Algebra for Some Gauge Theory Amplitudes," *Phys. Lett.*, Vol. B214, 1988, pp. 215–218.

[17] Gervais, J.-L. and Neveu, A., "Feynman rules for massive gauge fields with dual diagram topology," *Nucl. Phys.*, Vol. B46, 1972, pp. 381–401.

[18] Mangano, M. L., Parke, S. J., and Xu, Z., "Duality and Multi - Gluon Scattering," *Nucl. Phys.*, Vol. B298, 1988, pp. 653–672.

[19] Parke, S. J. and Taylor, T. R., "An Amplitude for $n$ Gluon Scattering," *Phys. Rev. Lett.*, Vol. 56, 1986, pp. 2459.

[20] Badger, S., "Automating QCD amplitudes with on-shell methods," *J. Phys. Conf. Ser.*, Vol. 762, No. 1, 2016, pp. 012057.

[21] Berends, F. A. and Giele, W. T., "Recursive Calculations for Processes with n Gluons," *Nucl. Phys.*, Vol. B306, 1988, pp. 759–808.

[22] Weinberg, S., *The Quantum theory of fields. Vol. 1: Foundations*, Cambridge University Press, 2005.

[23] Wigner, E. P., "On Unitary Representations of the Inhomogeneous Lorentz Group," *Annals Math.*, Vol. 40, 1939, pp. 149–204, [Reprint: Nucl. Phys. Proc. Suppl.6,9(1989)].

[24] Newton, I., *Philosophiae Naturalis Principia Mathematica*, apud Sa. Smith, London, 1687.

[25] Weinberg, S., "Nonlinear realizations of chiral symmetry," *Phys. Rev.*, Vol. 166, 1968, pp. 1568–1577.

[26] Weinberg, S., "Phenomenological Lagrangians," *Physica*, Vol. A96, No. 1-2, 1979, pp. 327–340.

[27] Henning, B., Lu, X., and Murayama, H., "How to use the Standard Model effective field theory," *JHEP*, Vol. 01, 2016, pp. 023.

[28] Wilson, K. G., "Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture," *Phys. Rev.*, Vol. B4, 1971, pp. 3174–3183.

[29] Poland, D., Rychkov, S., and Vichi, A., "The Conformal Bootstrap: Theory, Numerical Techniques, and Applications," *Rev. Mod. Phys.*, Vol. 91, 2019, pp. 015002.

[30] Zamolodchikov, A. B., "Irreversibility of the Flux of the Renormalization Group in a 2D Field Theory," *JETP Lett.*, Vol. 43, 1986, pp. 730–732, [Pisma Zh. Eksp. Teor. Fiz.43,565(1986)].

[31] Polchinski, J., "Scale and Conformal Invariance in Quantum Field Theory," *Nucl. Phys.*, Vol. B303, 1988, pp. 226–236.

[32] Nakayama, Y., "Scale invariance vs conformal invariance," *Phys. Rept.*, Vol. 569, 2015, pp. 1–93.

[33] Palti, E., "The Swampland: Introduction and Review," *Fortsch. Phys.*, Vol. 67, No. 6, 2019, pp. 1900037.

[34] Coleman, S. R., Wess, J., and Zumino, B., "Structure of phenomenological Lagrangians. 1." *Phys. Rev.*, Vol. 177, 1969, pp. 2239–2247.

[35] Callan, Jr., C. G., Coleman, S. R., Wess, J., and Zumino, B., "Structure of phenomenological Lagrangians. 2." *Phys. Rev.*, Vol. 177, 1969, pp. 2247–2250.

[36] Klein, R., Malek, E., Roest, D., and Stefanyszyn, D., "No-go theorem for a gauge vector as a spacetime Goldstone mode," *Phys. Rev.*, Vol. D98, No. 6, 2018, pp. 065001.

[37] Adler, S. L., "Consistency conditions on the strong interactions implied by a partially conserved axial vector current," *Phys. Rev.*, Vol. 137, 1965, pp. B1022–B1033, [,140(1964)].

[38] Cheung, C., Kampf, K., Novotny, J., and Trnka, J., "Effective Field Theories from Soft Limits of Scattering Amplitudes," *Phys. Rev. Lett.*, Vol. 114, No. 22, 2015, pp. 221602.

[39] Arkani-Hamed, N., Motl, L., Nicolis, A., and Vafa, C., "The String landscape, black holes and gravity as the weakest force," *JHEP*, Vol. 06, 2007, pp. 060.

[40] Susskind, L., "Trouble for remnants," 1995.

[41] Banks, T., Johnson, M., and Shomer, A., "A Note on Gauge Theories Coupled to Gravity," *JHEP*, Vol. 09, 2006, pp. 049.

[42] Vafa, C., "The String landscape and the swampland," 2005.

[43] Susskind, L., "Some speculations about black hole entropy in string theory," 1993, pp. 118–131.

[44] Horowitz, G. T. and Polchinski, J., "A Correspondence principle for black holes and strings," *Phys. Rev.*, Vol. D55, 1997, pp. 6189–6197.

[45] Gibbons, G. W., "Vacuum Polarization and the Spontaneous Loss of Charge by Black Holes," *Commun. Math. Phys.*, Vol. 44, 1975, pp. 245–264.

[46] Kats, Y., Motl, L., and Padi, M., "Higher-order corrections to mass-charge relation of extremal black holes," *JHEP*, Vol. 12, 2007, pp. 068.

[47] Hamada, Y., Noumi, T., and Shiu, G., "Weak Gravity Conjecture from Unitarity and Causality," *Phys. Rev. Lett.*, Vol. 123, No. 5, 2019, pp. 051601.

[48] Bellazzini, B., Lewandowski, M., and Serra, J., "Amplitudes' Positivity, Weak Gravity Conjecture, and Modified Gravity," *Phys. Rev. Lett.*, Vol. 123, No. 25, 2019, pp. 251103.

[49] Cheung, C., Liu, J., and Remmen, G. N., "Proof of the Weak Gravity Conjecture from Black Hole Entropy," *JHEP*, Vol. 10, 2018, pp. 004.

[50] Charles, A. M., "The Weak Gravity Conjecture, RG Flows, and Supersymmetry," 2019.

[51] Elvang, H., Jones, C. R. T., and Naculich, S. G., "Soft Photon and Graviton Theorems in Effective Field Theory," *Phys. Rev. Lett.*, Vol. 118, No. 23, 2017, pp. 231601.

[52] Elvang, H., Hadjiantonis, M., Jones, C. R. T., and Paranjape, S., "On the Supersymmetrization of Galileon Theories in Four Dimensions," *Phys. Lett.*, Vol. B781, 2018, pp. 656–663.

[53] Elvang, H., Hadjiantonis, M., Jones, C. R. T., and Paranjape, S., "Soft Bootstrap and Supersymmetry," *JHEP*, Vol. 01, 2019, pp. 195.

[54] Elvang, H., Hadjiantonis, M., Jones, C. R. T., and Paranjape, S., "All-Multiplicity One-Loop Amplitudes in Born-Infeld Electrodynamics from Generalized Unitarity," 2019.

[55] Jones, C. R. T. and McPeak, B., "The Black Hole Weak Gravity Conjecture with Multiple Charges," 2019.

[56] Cremonini, S., Jones, C. R. T., Liu, J. T., and McPeak, B., "Higher-Derivative Corrections to Entropy and the Weak Gravity Conjecture in Anti-de Sitter Space," 2019.

[57] Low, F. E., "Scattering of light of very low frequency by systems of spin 1/2," *Phys. Rev.*, Vol. 96, 1954, pp. 1428–1432.

[58] Gell-Mann, M. and Goldberger, M. L., "Scattering of low-energy photons by particles of spin 1/2," *Phys. Rev.*, Vol. 96, 1954, pp. 1433–1438.

[59] Low, F. E., "Bremsstrahlung of very low-energy quanta in elementary particle collisions," *Phys. Rev.*, Vol. 110, 1958, pp. 974–977.

[60] Weinberg, S., "Infrared photons and gravitons," *Phys. Rev.*, Vol. 140, 1965, pp. B516–B524.

[61] Burnett, T. H. and Kroll, N. M., "Extension of the low soft photon theorem," *Phys. Rev. Lett.*, Vol. 20, 1968, pp. 86.

[62] Gross, D. J. and Jackiw, R., "Low-Energy Theorem for Graviton Scattering," *Phys. Rev.*, Vol. 166, 1968, pp. 1287–1292.

[63] Jackiw, R., "Low-Energy Theorems for Massless Bosons: Photons and Gravitons," *Phys. Rev.*, Vol. 168, 1968, pp. 1623–1633.

[64] Cachazo, F. and Strominger, A., "Evidence for a New Soft Graviton Theorem," 2014.

[65] Kapec, D., Lysov, V., Pasterski, S., and Strominger, A., "Semiclassical Virasoro symmetry of the quantum gravity $\mathcal{S}$-matrix," *JHEP*, Vol. 08, 2014, pp. 058.

[66] He, T., Mitra, P., Porfyriadis, A. P., and Strominger, A., "New Symmetries of Massless QED," *JHEP*, Vol. 10, 2014, pp. 112.

[67] Lysov, V., Pasterski, S., and Strominger, A., "Lows Subleading Soft Theorem as a Symmetry of QED," *Phys. Rev. Lett.*, Vol. 113, No. 11, 2014, pp. 111601.

[68] Kapec, D., Mitra, P., Raclariu, A.-M., and Strominger, A., "A 2D Stress Tensor for 4D Gravity," 2016.

[69] Cheung, C., de la Fuente, A., and Sundrum, R., "4D Scattering Amplitudes and Asymptotic Symmetries from 2D CFT," 2016.

[70] Strominger, A., “Lectures on the Infrared Structure of Gravity and Gauge Theory,” 2017.

[71] Elvang, H. and Huang, Y.-t., *Scattering Amplitudes in Gauge Theory and Gravity*, Cambridge University Press, 2015.

[72] Casali, E., “Soft sub-leading divergences in Yang-Mills amplitudes,” *JHEP*, Vol. 08, 2014, pp. 077.

[73] Weinberg, S., “Photons and Gravitons in s Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass,” *Phys. Rev.*, Vol. 135, 1964, pp. B1049–B1056.

[74] Weinberg, S. and Witten, E., “Limits on Massless Particles,” *Phys. Lett.*, Vol. B96, 1980, pp. 59–62.

[75] Porrati, M., “Universal Limits on Massless High-Spin Particles,” *Phys. Rev.*, Vol. D78, 2008, pp. 065016.

[76] Cheung, C., Kampf, K., Novotny, J., Shen, C.-H., and Trnka, J., “A Periodic Table of Effective Field Theories,” *JHEP*, Vol. 02, 2017, pp. 020.

[77] Hinterbichler, K. and Joyce, A., “Hidden symmetry of the Galileon,” *Phys. Rev.*, Vol. D92, No. 2, 2015, pp. 023503.

[78] Farakos, F., Germani, C., and Kehagias, A., “On ghost-free supersymmetric galileons,” *JHEP*, Vol. 11, 2013, pp. 045.

[79] Cheung, C., Kampf, K., Novotny, J., Shen, C.-H., and Trnka, J., “On-Shell Recursion Relations for Effective Field Theories,” *Phys. Rev. Lett.*, Vol. 116, No. 4, 2016, pp. 041601.

[80] Kampf, K., Novotny, J., and Trnka, J., “Tree-level Amplitudes in the Nonlinear Sigma Model,” *JHEP*, Vol. 05, 2013, pp. 032.

[81] Luo, H. and Wen, C., “Recursion relations from soft theorems,” *JHEP*, Vol. 03, 2016, pp. 088.

[82] Nambu, Y. and Jona-Lasinio, G., “Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I,” *Phys. Rev.*, Vol. 122, Apr 1961, pp. 345–358.

[83] Elvang, H. and Huang, Y.-t., “Scattering Amplitudes,” 2013.

[84] Cheung, C., Shen, C.-H., and Trnka, J., “Simple Recursion Relations for General Field Theories,” *JHEP*, Vol. 06, 2015, pp. 118.

[85] Cheung, C. and Shen, C.-H., “Private communications,” 2017.

[86] de Rham, C. and Gabadadze, G., “Selftuned Massive Spin-2,” *Phys. Lett.*, Vol. B693, 2010, pp. 334–338.

[87] de Rham, C. and Gabadadze, G., “Generalization of the Fierz-Pauli Action,” *Phys. Rev.*, Vol. D82, 2010, pp. 044020.

[88] Hinterbichler, K., "Theoretical Aspects of Massive Gravity," *Rev. Mod. Phys.*, Vol. 84, 2012, pp. 671–710.

[89] Grisaru, M. T., Pendleton, H. N., and van Nieuwenhuizen, P., "Supergravity and the S Matrix," *Phys. Rev.*, Vol. D15, 1977, pp. 996.

[90] Grisaru, M. T. and Pendleton, H. N., "Some Properties of Scattering Amplitudes in Supersymmetric Theories," *Nucl. Phys.*, Vol. B124, 1977, pp. 81–92.

[91] Elvang, H., Freedman, D. Z., and Kiermaier, M., "Solution to the Ward Identities for Superamplitudes," *JHEP*, Vol. 10, 2010, pp. 103.

[92] Weinberg, S., "Infrared Photons and Gravitons," *Phys. Rev.*, Vol. 140, Oct 1965, pp. B516–B524.

[93] Elvang, H. and Kiermaier, M., "Stringy KLT relations, global symmetries, and $E_{7(7)}$ violation," *JHEP*, Vol. 10, 2010, pp. 108.

[94] Broedel, J. and Dixon, L. J., "R**4 counterterm and E(7)(7) symmetry in maximal supergravity," *JHEP*, Vol. 05, 2010, pp. 003.

[95] Low, I. and Yin, Z., "The Infrared Structure of Nambu-Goldstone Bosons," *JHEP*, Vol. 10, 2018, pp. 078.

[96] Zumino, B., "Supersymmetry and Kahler Manifolds," *Phys. Lett.*, Vol. 87B, 1979, pp. 203.

[97] Freedman, D. Z. and Van Proeyen, A., *Supergravity*, Cambridge Univ. Press, Cambridge, UK, 2012.

[98] Bando, M., Kuramoto, T., Maskawa, T., and Uehara, S., "Structure of Nonlinear Realization in Supersymmetric Theories," *Phys. Lett.*, Vol. 138B, 1984, pp. 94.

[99] Elvang, H., Huang, Y.-t., and Peng, C., "On-shell superamplitudes in N¡4 SYM," *JHEP*, Vol. 09, 2011, pp. 031.

[100] Laddha, A. and Mitra, P., "Asymptotic Symmetries and Subleading Soft Photon Theorem in Effective Field Theories," *JHEP*, Vol. 05, 2018, pp. 132.

[101] Dvali, G. R., Gabadadze, G., and Porrati, M., "4-D gravity on a brane in 5-D Minkowski space," *Phys. Lett.*, Vol. B485, 2000, pp. 208–214.

[102] Nicolis, A., Rattazzi, R., and Trincherini, E., "The Galileon as a local modification of gravity," *Phys. Rev.*, Vol. D79, 2009, pp. 064036.

[103] de Rham, C., Gabadadze, G., and Tolley, A. J., "Resummation of Massive Gravity," *Phys. Rev. Lett.*, Vol. 106, 2011, pp. 231101.

[104] de Rham, C. and Tolley, A. J., "DBI and the Galileon reunited," *JCAP*, Vol. 1005, 2010, pp. 015.

[105] Goon, G. L., Hinterbichler, K., and Trodden, M., "Stability and superluminality of spherical DBI galileon solutions," *Phys. Rev.*, Vol. D83, 2011, pp. 085015.

[106] Goon, G., Hinterbichler, K., and Trodden, M., "Symmetries for Galileons and DBI scalars on curved space," *JCAP*, Vol. 1107, 2011, pp. 017.

[107] Adams, A., Arkani-Hamed, N., Dubovsky, S., Nicolis, A., and Rattazzi, R., "Causality, analyticity and an IR obstruction to UV completion," *JHEP*, Vol. 10, 2006, pp. 014.

[108] Koehn, M., Lehners, J.-L., and Ovrut, B., "Supersymmetric cubic Galileons have ghosts," *Phys. Rev.*, Vol. D88, No. 2, 2013, pp. 023528.

[109] Khoury, J., Lehners, J.-L., and Ovrut, B. A., "Supersymmetric Galileons," *Phys. Rev.*, Vol. D84, 2011, pp. 043521.

[110] Khoury, J., Ovrut, B. A., and Stokes, J., "The Worldvolume Action of Kink Solitons in AdS Spacetime," *JHEP*, Vol. 08, 2012, pp. 015.

[111] Ovrut, B. A. and Stokes, J., "Heterotic Kink Solitons and their Worldvolume Action," *JHEP*, Vol. 09, 2012, pp. 065.

[112] Goon, G., Hinterbichler, K., Joyce, A., and Trodden, M., "Galileons as Wess-Zumino Terms," *JHEP*, Vol. 06, 2012, pp. 004.

[113] Roest, D., Werkman, P., and Yamada, Y., "Internal Supersymmetry and Small-field Goldstini," *JHEP*, Vol. 05, 2018, pp. 190.

[114] Bagger, J. and Galperin, A., "Matter couplings in partially broken extended supersymmetry," *Phys. Lett.*, Vol. B336, 1994, pp. 25–31.

[115] Toppan, F., "On anomalies in classical dynamical systems," *J. Nonlin. Math. Phys.*, Vol. 8, 2001, pp. 518–533.

[116] Fabri, E. and Picasso, L. E., "Quantum Field Theory and Approximate Symmetries," *Phys. Rev. Lett.*, Vol. 16, Mar 1966, pp. 408–410.

[117] Dumitrescu, T. T. and Seiberg, N., "Supercurrents and Brane Currents in Diverse Dimensions," *JHEP*, Vol. 07, 2011, pp. 095.

[118] Hinterbichler, K., Trodden, M., and Wesley, D., "Multi-field galileons and higher codimension branes," *Phys. Rev.*, Vol. D82, 2010, pp. 124018.

[119] Beltran Jimenez, J. and Heisenberg, L., "Derivative self-interactions for a massive vector field," *Phys. Lett.*, Vol. B757, 2016, pp. 405–411.

[120] Padilla, A., Stefanyszyn, D., and Wilson, T., "Probing Scalar Effective Field Theories with the Soft Limits of Scattering Amplitudes," *JHEP*, Vol. 04, 2017, pp. 015.

[121] Cachazo, F., He, S., and Yuan, E. Y., "Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM," *JHEP*, Vol. 07, 2015, pp. 149.

[122] Broedel, J. and Dixon, L. J., "Color-kinematics duality and double-copy construction for amplitudes from higher-dimension operators," *JHEP*, Vol. 10, 2012, pp. 091.

[123] Chen, G. and Du, Y.-J., "Amplitude Relations in Non-linear Sigma Model," *JHEP*, Vol. 01, 2014, pp. 061.

[124] Kampf, K., Novotny, J., and Trnka, J., "Recursion relations for tree-level amplitudes in the $SU(N)$ nonlinear sigma model," *Phys. Rev.*, Vol. D87, No. 8, 2013, pp. 081701.

[125] Carrasco, J. J. M., Mafra, C. R., and Schlotterer, O., "Abelian Z-theory: NLSM amplitudes and $\alpha$'-corrections from the open string," *JHEP*, Vol. 06, 2017, pp. 093.

[126] Born, M. and Infeld, L., "Foundations of the new field theory," *Proc. Roy. Soc. Lond.*, Vol. A144, No. 852, 1934, pp. 425–451.

[127] Fradkin, E. S. and Tseytlin, A. A., "Nonlinear Electrodynamics from Quantized Strings," *Phys. Lett.*, Vol. 163B, 1985, pp. 123–130.

[128] Polchinski, J., "Dirichlet Branes and Ramond-Ramond charges," *Phys. Rev. Lett.*, Vol. 75, 1995, pp. 4724–4727.

[129] Polchinski, J., "Tasi lectures on D-branes," *Fields, strings and duality. Proceedings, Summer School, Theoretical Advanced Study Institute in Elementary Particle Physics, TASI'96, Boulder, USA, June 2-28, 1996*, 1996, pp. 293–356.

[130] Cheung, C., Remmen, G. N., Shen, C.-H., and Wen, C., "Pions as Gluons in Higher Dimensions," *JHEP*, Vol. 04, 2018, pp. 129.

[131] Cheung, C., Shen, C.-H., and Wen, C., "Unifying Relations for Scattering Amplitudes," *JHEP*, Vol. 02, 2018, pp. 095.

[132] Shmakova, M., "One loop corrections to the D3-brane action," *Phys. Rev.*, Vol. D62, 2000, pp. 104009.

[133] Bern, Z., Dixon, L. J., Perelstein, M., and Rozowsky, J. S., "Multileg one loop gravity amplitudes from gauge theory," *Nucl. Phys.*, Vol. B546, 1999, pp. 423–479.

[134] Heisenberg, W. and Euler, H., "Consequences of Dirac's theory of positrons," *Z. Phys.*, Vol. 98, No. 11-12, 1936, pp. 714–732.

[135] Anastasiou, C., Britto, R., Feng, B., Kunszt, Z., and Mastrolia, P., "D-dimensional unitarity cut method," *Phys. Lett.*, Vol. B645, 2007, pp. 213–216.

[136] Bern, Z. and Huang, Y.-t., "Basics of Generalized Unitarity," *J. Phys.*, Vol. A44, 2011, pp. 454003.

[137] Cutkosky, R. E., "Singularities and discontinuities of Feynman amplitudes," *J. Math. Phys.*, Vol. 1, 1960, pp. 429–433.

[138] Britto, R., "Loop Amplitudes in Gauge Theories: Modern Analytic Approaches," *J. Phys.*, Vol. A44, 2011, pp. 454006.

[139] Bern, Z., Dixon, L. J., and Kosower, D. A., "One loop amplitudes for e+ e- to four partons," *Nucl. Phys.*, Vol. B513, 1998, pp. 3–86.

[140] Bern, Z. and Kosower, D. A., "The Computation of loop amplitudes in gauge theories," *Nucl. Phys.*, Vol. B379, 1992, pp. 451–561.

[141] Bern, Z., De Freitas, A., Dixon, L. J., and Wong, H. L., "Supersymmetric regularization, two loop QCD amplitudes and coupling shifts," *Phys. Rev.*, Vol. D66, 2002, pp. 085002.

[142] Nandan, D., Plefka, J., and Travaglini, G., "All rational one-loop Einstein-Yang-Mills amplitudes at four points," *JHEP*, Vol. 09, 2018, pp. 011.

[143] Cachazo, F., He, S., and Yuan, E. Y., "Scattering of Massless Particles in Arbitrary Dimensions," *Phys. Rev. Lett.*, Vol. 113, No. 17, 2014, pp. 171601.

[144] Cachazo, F., He, S., and Yuan, E. Y., "Scattering of Massless Particles: Scalars, Gluons and Gravitons," *JHEP*, Vol. 07, 2014, pp. 033.

[145] Bern, Z., Carrasco, J. J. M., and Johansson, H., "Perturbative Quantum Gravity as a Double Copy of Gauge Theory," *Phys. Rev. Lett.*, Vol. 105, 2010, pp. 061602.

[146] Bern, Z., Carrasco, J. J., Chen, W.-M., Edison, A., Johansson, H., Parra-Martinez, J., Roiban, R., and Zeng, M., "Ultraviolet Properties of $\mathcal{N} = 8$ Supergravity at Five Loops," *Phys. Rev.*, Vol. D98, No. 8, 2018, pp. 086021.

[147] Heydeman, M., Schwarz, J. H., and Wen, C., "M5-Brane and D-Brane Scattering Amplitudes," *JHEP*, Vol. 12, 2017, pp. 003.

[148] Bergshoeff, E. and De Roo, M., "D-branes and T duality," *Phys. Lett.*, Vol. B380, 1996, pp. 265–272.

[149] Badger, S. D., Glover, E. W. N., Khoze, V. V., and Svrcek, P., "Recursion relations for gauge theory amplitudes with massive particles," *JHEP*, Vol. 07, 2005, pp. 025.

[150] Forde, D. and Kosower, D. A., "All-multiplicity amplitudes with massive scalars," *Phys. Rev.*, Vol. D73, 2006, pp. 065007.

[151] Carrillo Gonzlez, M., Penco, R., and Trodden, M., "Shift symmetries, soft limits, and the double copy beyond leading order," 2019.

[152] Low, I. and Yin, Z., "Soft Bootstrap and Effective Field Theories," *JHEP*, Vol. 11, 2019, pp. 078.

[153] Rodina, L., "Uniqueness from gauge invariance and the Adler zero," *JHEP*, Vol. 09, 2019, pp. 084.

[154] Rodina, L., "Scattering Amplitudes from Soft Theorems and Infrared Behavior," *Phys. Rev. Lett.*, Vol. 122, No. 7, 2019, pp. 071601.

[155] Rosly, A. A. and Selivanov, K. G., "Helicity conservation in Born-Infeld theory," *Quarks. Proceedings, 12th International Seminar on High Energy Physics, Quarks'2002, Novgorod, Russia, June 1-7, 2002*, 2002.

[156] Boels, R., Larsen, K. J., Obers, N. A., and Vonk, M., "MHV, CSW and BCFW: Field theory structures in string theory amplitudes," *JHEP*, Vol. 11, 2008, pp. 015.

[157] Novotn, J., "Self-duality, helicity conservation and normal ordering in nonlinear QED," *Phys. Rev.*, Vol. D98, No. 8, 2018, pp. 085015.

[158] Gibbons, G. W. and Rasheed, D. A., “Electric - magnetic duality rotations in nonlinear electrodynamics,” *Nucl. Phys.*, Vol. B454, 1995, pp. 185–206.

[159] Deser, S. and Teitelboim, C., “Duality Transformations of Abelian and Nonabelian Gauge Fields,” *Phys. Rev.*, Vol. D13, 1976, pp. 1592–1597.

[160] Deser, S., “Off-Shell Electromagnetic Duality Invariance,” *J. Phys.*, Vol. A15, 1982, pp. 1053.

[161] Schwarz, J. H. and Sen, A., “Duality symmetric actions,” *Nucl. Phys.*, Vol. B411, 1994, pp. 35–63.

[162] Berman, D., “SL(2,Z) duality of Born-Infeld theory from nonlinear selfdual electrodynamics in six-dimensions,” *Phys. Lett.*, Vol. B409, 1997, pp. 153–159.

[163] Roiban, R. and Tseytlin, A. A., “On Duality Symmetry in Perturbative Quantum Theory,” *JHEP*, Vol. 10, 2012, pp. 099.

[164] Bern, Z., Parra-Martinez, J., and Roiban, R., “Canceling the U(1) Anomaly in the $S$ Matrix of $N$=4 Supergravity,” *Phys. Rev. Lett.*, Vol. 121, No. 10, 2018, pp. 101604.

[165] Bern, Z., Kosower, D., and Parra-Martinez, J., “Two-loop n-point anomalous amplitudes in $N = 4$ supergravity,” 2019.

[166] Cheung, C. and Remmen, G. N., “Naturalness and the Weak Gravity Conjecture,” *Phys. Rev. Lett.*, Vol. 113, 2014, pp. 051601.

[167] Andriolo, S., Junghans, D., Noumi, T., and Shiu, G., “A Tower Weak Gravity Conjecture from Infrared Consistency,” *Fortsch. Phys.*, Vol. 66, No. 5, 2018, pp. 1800020.

[168] Wang, F. and Qi, L., “Comments on ”Explicit criterion for the positive definiteness of a general quartic form”,” *IEEE Trans. Automat. Contr.*, Vol. 50, 2005, pp. 416–418.

[169] Heisenberg, W. and Euler, H., “Folgerungen aus der Diracschen Theorie des Positrons,” *Zeitschrift für Physik*, Vol. 98, No. 11, Nov 1936, pp. 714–732.

[170] Gaillard, M. K. and Zumino, B., “Duality Rotations for Interacting Fields,” *Nucl. Phys.*, Vol. B193, 1981, pp. 221–244.

[171] Cheung, C. and Remmen, G. N., “Infrared Consistency and the Weak Gravity Conjecture,” *JHEP*, Vol. 12, 2014, pp. 087.

[172] Colwell, K. and Terning, J., “S-Duality and Helicity Amplitudes,” *JHEP*, Vol. 03, 2016, pp. 068.

[173] Kovalevich, S. G., Osland, P., Shnir, Ya. M., and Tolkachev, E. A., “The Effective Lagrangian of QED with a magnetic charge and dyon mass bounds,” *Phys. Rev.*, Vol. D55, 1997, pp. 5807–5812.

[174] Camanho, X. O., Edelstein, J. D., Maldacena, J., and Zhiboedov, A., “Causality Constraints on Corrections to the Graviton Three-Point Coupling,” *JHEP*, Vol. 02, 2016, pp. 020.

[175] Arkani-Hamed, N., Cachazo, F., and Kaplan, J., "What is the Simplest Quantum Field Theory?" *JHEP*, Vol. 09, 2010, pp. 016.

[176] Huang, Y.-t., McGady, D. A., and Peng, C., "One-loop renormalization and the S-matrix," *Phys. Rev.*, Vol. D87, No. 8, 2013, pp. 085028.

[177] Kinoshita, T., "Mass singularities of Feynman amplitudes," *J. Math. Phys.*, Vol. 3, 1962, pp. 650–677.

[178] Deser, S. and van Nieuwenhuizen, P., "One Loop Divergences of Quantized Einstein-Maxwell Fields," *Phys. Rev.*, Vol. D10, 1974, pp. 401.

[179] Deser, S., Tsao, H.-S., and van Nieuwenhuizen, P., "One Loop Divergences of the Einstein Yang-Mills System," *Phys. Rev.*, Vol. D10, 1974, pp. 3337.

[180] Charles, A. M., Larsen, F., and Mayerson, D. R., "Non-Renormalization For Non-Supersymmetric Black Holes," *JHEP*, Vol. 08, 2017, pp. 048.

[181] Dunbar, D. C. and Norridge, P. S., "Infinities within graviton scattering amplitudes," *Class. Quant. Grav.*, Vol. 14, 1997, pp. 351–365.

[182] Agullo, I., del Rio, A., and Navarro-Salas, J., "Electromagnetic duality anomaly in curved spacetimes," *Phys. Rev. Lett.*, Vol. 118, No. 11, 2017, pp. 111301.

[183] Marcus, N., "Composite Anomalies in Supergravity," *Phys. Lett.*, Vol. 157B, 1985, pp. 383–388.

[184] Carrasco, J. J. M., Kallosh, R., Roiban, R., and Tseytlin, A. A., "On the U(1) duality anomaly and the S-matrix of N=4 supergravity," *JHEP*, Vol. 07, 2013, pp. 029.

[185] Norridge, P. S., "Recovering infinities in graviton scattering amplitudes using Cutkosky rules," *Phys. Lett.*, Vol. B387, 1996, pp. 701–706.

[186] Goon, G. and Penco, R., "A Universal Relation Between Corrections to Entropy and Extremality," 2019.

[187] Nakayama, Y. and Nomura, Y., "Weak gravity conjecture in the AdS/CFT correspondence," *Phys. Rev.*, Vol. D92, No. 12, 2015, pp. 126006.

[188] Harlow, D., "Wormholes, Emergent Gauge Fields, and the Weak Gravity Conjecture," *JHEP*, Vol. 01, 2016, pp. 122.

[189] Montero, M., Shiu, G., and Soler, P., "The Weak Gravity Conjecture in three dimensions," *JHEP*, Vol. 10, 2016, pp. 159.

[190] Montero, M., "A Holographic Derivation of the Weak Gravity Conjecture," *JHEP*, Vol. 03, 2019, pp. 157.

[191] Myers, R. C., Paulos, M. F., and Sinha, A., "Holographic Hydrodynamics with a Chemical Potential," *JHEP*, Vol. 06, 2009, pp. 006.

[192] Cremonini, S., Liu, J. T., and Szepietowski, P., "Higher Derivative Corrections to R-charged Black Holes: Boundary Counterterms and the Mass-Charge Relation," *JHEP*, Vol. 03, 2010, pp. 042.

[193] Balasubramanian, V. and Kraus, P., "A Stress tensor for Anti-de Sitter gravity," *Commun. Math. Phys.*, Vol. 208, 1999, pp. 413–428.

[194] Loukas, O., Orlando, D., Reffert, S., and Sarkar, D., "An AdS/EFT correspondence at large charge," *Nucl. Phys.*, Vol. B934, 2018, pp. 437–458.

[195] Wald, R. M., "Black hole entropy is the Noether charge," *Phys. Rev.*, Vol. D48, No. 8, 1993, pp. R3427–R3431.

[196] Hawking, S. W. and Israel, W., *General Relativity*, Univ. Pr., Cambridge, UK, 1979.

[197] York, Jr., J. W., "Role of conformal three geometry in the dynamics of gravitation," *Phys. Rev. Lett.*, Vol. 28, 1972, pp. 1082–1085.

[198] Gibbons, G. W. and Hawking, S. W., "Action Integrals and Partition Functions in Quantum Gravity," *Phys. Rev.*, Vol. D15, 1977, pp. 2752–2756.

[199] Hawking, S. W. and Ross, S. F., "Duality between electric and magnetic black holes," *Phys. Rev.*, Vol. D52, 1995, pp. 5865–5876.

[200] Henningson, M. and Skenderis, K., "The Holographic Weyl anomaly," *JHEP*, Vol. 07, 1998, pp. 023.

[201] Emparan, R., Johnson, C. V., and Myers, R. C., "Surface terms as counterterms in the AdS / CFT correspondence," *Phys. Rev.*, Vol. D60, 1999, pp. 104001.

[202] Skenderis, K., "Lecture notes on holographic renormalization," *Class. Quant. Grav.*, Vol. 19, 2002, pp. 5849–5876.

[203] Liu, J. T. and Sabra, W. A., "Hamilton-Jacobi Counterterms for Einstein-Gauss-Bonnet Gravity," *Class. Quant. Grav.*, Vol. 27, 2010, pp. 175014.

[204] Reall, H. S. and Santos, J. E., "Higher derivative corrections to Kerr black hole thermodynamics," *JHEP*, Vol. 04, 2019, pp. 021.

[205] Cheung, C., Liu, J., and Remmen, G. N., "Entropy Bounds on Effective Field Theory from Rotating Dyonic Black Holes," *Phys. Rev.*, Vol. D100, No. 4, 2019, pp. 046003.

[206] Loges, G. J., Noumi, T., and Shiu, G., "Thermodynamics of 4D Dilatonic Black Holes and the Weak Gravity Conjecture," 2019.

[207] Chamblin, A., Emparan, R., Johnson, C. V., and Myers, R. C., "Charged AdS black holes and catastrophic holography," *Phys. Rev.*, Vol. D60, 1999, pp. 064018.

[208] Chamblin, A., Emparan, R., Johnson, C. V., and Myers, R. C., "Holography, thermodynamics and fluctuations of charged AdS black holes," *Phys. Rev.*, Vol. D60, 1999, pp. 104026.

[209] Shen, J.-y., Cai, R.-G., Wang, B., and Su, R.-K., "Thermodynamic geometry and critical behavior of black holes," *Int. J. Mod. Phys.*, Vol. A22, 2007, pp. 11–27.

[210] Kats, Y. and Petrov, P., "Effect of curvature squared corrections in AdS on the viscosity of the dual gauge theory," *JHEP*, Vol. 01, 2009, pp. 044.

[211] Cheung, C. and Remmen, G. N., "Positivity of Curvature-Squared Corrections in Gravity," *Phys. Rev. Lett.*, Vol. 118, No. 5, 2017, pp. 051601.

[212] Nojiri, S. and Odintsov, S. D., "On the conformal anomaly from higher derivative gravity in AdS / CFT correspondence," *Int. J. Mod. Phys.*, Vol. A15, 2000, pp. 413–428.

[213] Blau, M., Narain, K. S., and Gava, E., "On subleading contributions to the AdS / CFT trace anomaly," *JHEP*, Vol. 09, 1999, pp. 018.

[214] Buchel, A., Myers, R. C., and Sinha, A., "Beyond eta/s = 1/4 pi," *JHEP*, Vol. 03, 2009, pp. 084.

[215] Hofman, D. M. and Maldacena, J., "Conformal collider physics: Energy and charge correlations," *JHEP*, Vol. 05, 2008, pp. 012.

[216] Policastro, G., Son, D. T., and Starinets, A. O., "The Shear viscosity of strongly coupled N=4 supersymmetric Yang-Mills plasma," *Phys. Rev. Lett.*, Vol. 87, 2001, pp. 081601.

[217] Buchel, A. and Liu, J. T., "Universality of the shear viscosity in supergravity," *Phys. Rev. Lett.*, Vol. 93, 2004, pp. 090602.

[218] Cremonini, S., "The Shear Viscosity to Entropy Ratio: A Status Report," *Mod. Phys. Lett.*, Vol. B25, 2011, pp. 1867–1888.

[219] Kovtun, P. and Ritz, A., "Universal conductivity and central charges," *Phys. Rev.*, Vol. D78, 2008, pp. 066009.

[220] Cremonini, S. and Szepietowski, P., "Generating Temperature Flow for eta/s with Higher Derivatives: From Lifshitz to AdS," *JHEP*, Vol. 02, 2012, pp. 038.

[221] Cremonini, S., Gursoy, U., and Szepietowski, P., "On the Temperature Dependence of the Shear Viscosity and Holography," *JHEP*, Vol. 08, 2012, pp. 167.

[222] Hellerman, S., Orlando, D., Reffert, S., and Watanabe, M., "On the CFT Operator Spectrum at Large Global Charge," *JHEP*, Vol. 12, 2015, pp. 071.

[223] Gubser, S. S., "Breaking an Abelian gauge symmetry near a black hole horizon," *Phys. Rev.*, Vol. D78, 2008, pp. 065034.

[224] Hartnoll, S. A., Herzog, C. P., and Horowitz, G. T., "Holographic Superconductors," *JHEP*, Vol. 12, 2008, pp. 015.

[225] Hartnoll, S. A., Herzog, C. P., and Horowitz, G. T., "Building a Holographic Superconductor," *Phys. Rev. Lett.*, Vol. 101, 2008, pp. 031601.

[226] Denef, F. and Hartnoll, S. A., "Landscape of superconducting membranes," *Phys. Rev.*, Vol. D79, 2009, pp. 126008.

[227] Bhattacharyya, S., Minwalla, S., and Papadodimas, K., "Small Hairy Black Holes in $AdS_5xS^5$," *JHEP*, Vol. 11, 2011, pp. 035.

[228] Dias, O. J. C., Monteiro, R., Reall, H. S., and Santos, J. E., "A Scalar field condensation instability of rotating anti-de Sitter black holes," *JHEP*, Vol. 11, 2010, pp. 036.

[229] Crisford, T., Horowitz, G. T., and Santos, J. E., "Testing the Weak Gravity - Cosmic Censorship Connection," *Phys. Rev.*, Vol. D97, No. 6, 2018, pp. 066005.

[230] Horowitz, G. T. and Santos, J. E., "Further evidence for the weak gravity  cosmic censorship connection," *JHEP*, Vol. 06, 2019, pp. 122.

[231] Horowitz, G. T., Santos, J. E., and Way, B., "Evidence for an Electrifying Violation of Cosmic Censorship," *Class. Quant. Grav.*, Vol. 33, No. 19, 2016, pp. 195007.

[232] Bern, Z. and Morgan, A. G., "Massive loop amplitudes from unitarity," *Nucl. Phys.*, Vol. B467, 1996, pp. 479–509.

[233] Davydychev, A. I., "Explicit results for all orders of the epsilon expansion of certain massive and massless diagrams," *Phys. Rev.*, Vol. D61, 2000, pp. 087701.

[234] Arzt, C., "Reduced effective Lagrangians," *Phys. Lett.*, Vol. B342, 1995, pp. 189–195.

[235] Henning, B., Lu, X., Melia, T., and Murayama, H., "Operator bases, $S$-matrices, and their partition functions," *JHEP*, Vol. 10, 2017, pp. 199.

[236] Bjorndahl, C. and Karshon, Y., "Revisiting TietzeNakajima: Local and Global Convexity for Maps," *Canadian Journal of Mathematics*, Vol. 62, No. 5, 2010, pp. 975993.

[237] Cano, P. A., Ortin, T., and Ramirez, P. F., "On the extremality bound of stringy black holes," 2019.

[238] Cano, P. A., Chimento, S., Linares, R., Ortin, T., and Ramirez, P. F., "$\alpha'$ corrections of Reissner-Nordstrm black holes," *JHEP*, Vol. 02, 2020, pp. 031.