ON-SHELL SCATTERING AND TEMPERATURE-REFLECTIONS

David A. McGady

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics
Adviser: Nima Arkani-Hamed

March 2015

Abstract

Merging Einstein's relativity with quantum mechanics leads almost inexorably to relativistic quantum field theory (QFT). Although relativity and non-relativistic quantum mechanics have been on solid mathematical footing for nearly a century, many aspects of relativistic QFT remain elusive and poorly understood. In this thesis, we study two fundamentally important objects in quantum field theory: the S-matrix of a given QFT, which quantifies how quantum states interact and scatter off of each other, and the partition function, a quantity which defines observables of a given QFT.

In chapters 2–6, we study generic properties of the S-matrix between on-shell massless states in four- and six-dimensions. S-matrix analysis performed with on-shell probes differ from more conventional analysis, with Feynman diagrams, in two important ways: (1) on-shell calculations are automatically gauge-invariant from start to finish, and (2) on-shell probes are inherently delocalized through all of space and time, i.e. they are "long distance" probes. In chapter 2, we note that one-loop scattering amplitudes have ultraviolet divergences that dictate how coupling constants in QFT "run" and evolve at finite distance. We show that on-shell techniques, which use exclusively long distance probes, are nevertheless sensitive to these important finite distance effects. In chapters 3–4, we show how the manifestly gauge-invariant on-shell S-matrix can be sensitive to what are called "gauge anomalies" in more conventional discussions of relativistic quantum systems, i.e. in local formulations of quantum field theory. In chapter 5 we use the basic tools of the analytic S-matrix program in an exhaustive study of the simplest non-trivial scattering processes in massless theories in four-dimensions. From the most basic incarnations of locality and unitarity, we derive many classic results, such as the Weinberg–Witten theorem, the equivalence theorem, supersymmetry, and the exclusion of "higher spin" S-matrices. Finally, in chapter 6, we inductively prove that the entire tree-level S-matrix of Einstein gravity in four dimensions can be recursively constructed through on-shell means. This implies, as a corollary, that the Einstein-Hilbert action may be completely excised from studies of tree-level/semi-classical scattering in General Relativity.

In chapters 7–9, we note a surprising property of statistical mechanical partition functions for many exactly solved quantum field theories, and explore a basic corollary. In chapter 7, we note that many partition functions, $Z(T) = \sum_n e^{-E_n/T}$, that can be exactly computed and re-summed into closed-form expressions, are surprisingly self-similar under temperature reflection (T-reflection). In short, we find that $Z(+T) = e^{i\gamma} Z(-T)$, where γ is a real number that is independent of temperature. This T-reflection symmetry only exists for a unique value of the ground state energy, often given by the naive quantization of classical potentials/Hamiltonians. In chapter 8, we note that a certain

limit of quantum chromodynamics (QCD) is symmetric under T-reflections only when its vacuum energy vanishes. We verify that this is indeed the correct vacuum energy of this theory through two calculations that are independent of each other and independent of the presence or absence of T-reflection symmetry. In chapter 9, we use this T-reflection symmetry to obtain a detailed understanding of the presence/absence of Hagedorn phase transitions in a certain calculable limit of QCD.

Acknowledgements

My gratitude extends at all levels, from the very broad, down to the extremely personal. Broadly speaking, I am thoroughly and deeply grateful to both Princeton University and the Institute for Advanced Study for providing such a wonderful environment for my studies. I have been deeply honored to participate in such a diverse, active, and interesting community of friends and scholars. Further, without the value and monetary support that the government and people of the United States of America have given to science, none of this would be possible. For all of these broad things and themes, I am and will remain ever grateful.

On a more personal level, I have so many individuals to thank that any list would almost equal the length of a chapter. I cannot do this. As such, this list is glaringly incomplete. Virtually without exception, it has been a pleasure to meet and know each and every individual during my time at Princeton.

In my time here, I have learned from so many wonderful scientists. None of this would have been possible without the unique combination of guidance, support, and freedom that my advisor, Nima Arkani-Hamed, has given me. Thank you so much, Nima. Further, I would like to thank both the other members of my committee, Lyman Page and Herman Verlinde for their interest, support, and guidance in research throughout my time in Princeton, and to Simone Giombi for being the second reader of this thesis, for their time and input into making this thesis what it is today.

Science is unthinkable without the phrases "I don't know" and "I don't understand". Collectively, there are things science does not know or understand. But we are all individuals, each with our own flawed understanding and knowledge of the field. And acknowledging one's own ignorance and stupidity is so hard. I would like to thank my collaborators, Aleksey Cherman, Yu-tin Huang, Laurentiu Rodina, Masahito Yamazaki, Gokce Basar, Wei-Ming Chen, and Cheng Peng, for providing safe havens and space to make all of the silly errors necessary to do serious work. Additional thanks is due to the entire high energy theory group in Princeton (University and Institute). I would like to particularly thank the students and postdocs who shared so many wonderful and inspiring conversations. Thank you Guilherme Pimentel, Jaroslav Trnka, Sasha Zhiboedov, Michael Kiermaier, Song He, Ben Safdi, Thomas Dumitrescu and Johannes Henn. In truth, the central ideas for chapters 7–9 came grew out of one of the silliest mistakes imaginable. Without the comfort to make such mistakes on the fly, and the support to understand and pursue this accidental truth, which, as time goes on, seems to me more and more exciting, my work would have suffered considerably. I lack the vocabulary to articulate the depths of gratitude I have to all of you.

Just as importantly, I could not have managed without my many friends. You know who you are; please be understanding if I have accidentally forgotten to list you. I guarantee I meant to put you there. The following list is the surely incomplete, but most of you are (I hope?) there: Blake Sherwin, Jon Gudmundsson, Dora Sigurdardottir, Guilherme Pimentel, Jaroslav Trnka, Sasha Zhiboedov, Andrew Hartnett, Michael Mooney, Doug Swanson, Anushya Chandran, Halil Saka, Nikhil Deshmuk, Anne Gambril, Thomas Dumetriscu, Ben Safdi, Ed Young, Chris Maxey, Dave Witt, Mark Brooks, Iris Chan, Colin Hill, Maryam Patton, Bhadri Visweswaran, Sonika Johri, Carole Dalin and Quentin Berthet, Matt Kosmer, Krysten Connon, Kaitlin LaPallo, Natalie Berger, Allison Chang, John Klopfer, Jess Lueders-Dumont, Margaret Lyford, Peng Zhao, Cara Brook, and everybody on the Big Bangers and the GC Skating Chimps. Further, I would like to thank all of the ladies and gentlemen who I've shared the ice with at Noontime Hockey (NTH). Without lunchtime hockey, things would have been much harder. That group is truly unique, and not only for hockey: one of the pivotal moments in my own personal history of T-reflections (the primary content of chapters 7–9) came when Mikko Haataja and I were talking about physics on the bench, between shifts.

Truly, the people I've met at Princeton have enriched my life and my time here beyond my power to put into words. It has been an honor and a pleasure to be here, these long years.

Finally, to my family, I can say nothing more nor less than this: without you I would be nowhere. I love you all, and could not ask for any better, more loving, or more nurturing family.

To all I have loved who are no longer with us.

Contents

Chapter 1

Why Quantum Field Theory?

Notions of space and of time, and the closely related concepts of locality and causality (cause preceding effect), are some of the oldest in physics. Everyday observations and experience lend us very natural intuition about space and time. Newton codified this intuition into his spectacularly successful laws of mechanics and gravity: time marches ever onwards in the same way and at the same speed for all observers, regardless the state of the various observers (e.g. relative motion). Classical physics, built on these ideas about space and time, grew ever more sophisticated and successful. However, the nearly simultaneous discoveries of relativity and quantum mechanics in twentieth century forced fundamental changes of the physical meaning of space and time.

Relativity forces us to modify our understanding of space and time. Because the speed of light is constant in all frames of reference, we must modify our notions of space and time, to weave the disparate classical notions of space and time into the unified fabric of space-time. For example, if a rocket (or a tau-lepton) flies towards me at half the speed of light, i.e. $v = c/2$, and turns on a flashlight pointed (or radiatively decays to a muon and a pair of photons which fly) towards me, the light will fly towards me at the speed of light, $v_{\text{light}} = c$. Classically, this could not be so. Classical notions of space and time require velocities to combine *linearly*; relativistically this fails to hold:

$$\text{Classical} : v \oplus w = v + w \tag{1.1}$$

$$\text{Relativistic} : v \oplus w = \frac{v + w}{1 + vw/c^2} \,. \tag{1.2}$$

Clearly, classical addition of velocities gives $c/2 \oplus c = 1.5c$, while the relativistic velocity addition formula yields $c/2 \oplus c = c$. Straightforward algebraic and geometrical reasoning leads from the invariance of the speed of light in all inertial frames of reference, exemplified by Eq. (1.2), to the conclusion that time and space must be put on equal footing. Just as rotations mix x and y, "boosting" to relativistic speeds must mix space with time.[1] The essence of relativity is a democracy between space and time.

Simultaneously, quantum mechanics and the inherent wave-particle duality of the fundamental constituents of matter force us to relinquish the primary role of position as *the* central object in a dynamical treatment of nature. Though matter is corpuscular in nature—we can count electrons, we can count photons—it also has inherently wave-like properties. Reconciling these two seemingly opposing properties forces a probabilistic understanding for the position of quantum matter. Indeed, the very physical and mathematical act of measuring positions of particles governed by quantum mechanical interactions acts to change their state: if a given particle is in a state of well-defined momentum, it is necessarily delocalized throughout space. Mathematically, measuring the position of an electron in a delocalized plane-wave state, a state with definite momentum that is spread out over space, literally amounts to a projection of the delocalized wave onto one isolated point in space. While one may ask for positions of particles in quantum mechanics, position is but a shadow, a projection of the full quantum state of the system.

Each pillar of twentieth century physics, built up from a bedrock footing of solid experimental measurements, enforces a subtle yet dramatic retooling of the very notion of space and time. Nature at the atomic and nuclear scales is both fundamentally relativistic and fundamentally quantum mechanical. Space and time must be treated democratically, and they must be inherently quantum in nature. Separately, one can easily incorporate relativistic effects into classical mechanics. Similarly, it is not overly difficult to incorporate quantum effects into classical (i.e. non-relativistic) physics: positions of fundamental degrees of freedom may fluctuate quantum mechanically, but the overall quantum mechanical state of a system evolves in time according to the well-known Schrodinger equation of motion, $\widehat{H}|\psi(t)\rangle = i\hbar\partial_t|\psi(t)\rangle$.

Such treatment however does inherent violence to the democracy between space and time mandated by Einstein's relativity. Put more plainly, non-relativistic quantum mechanics has position as an observable property of the quantum state, yet the state explicitly depends on time. Time is a parameter, but position is observable. To put space and time on equal footings—mandated by relativity—one of two options is must be taken:

1. Promote time to an observable of quantum mechanical world-lines. Because relativistic parti-

[1] To my mind, geometric constructions using space-time diagrams, rather than algebraic manipulations, powerfully illustrate the road from the fundamental observation that the speed of light is invariant under "boosts" to the mathematical relations in special relativity, e.g. Eq. (1.2). I highly recommend Sander Bias' book, *Very Special Relativity*, for a readable and exciting introduction to the lovely structure of relativity and space-time.

cles sweep out world-lines in space-time that are naturally endowed with a proper time, τ, an affine parameter which increases along the world line, one can construct equations of motion which evolve a quantum state along its *proper* time.

2. Demote position to a parameter of the quantum mechanical system. In this situation, to allow an amplitude for the fundamental quantum degrees of freedom to take on fluctuating values at a given position in space and time, one is forced to allow the quantum variable to exist at all points in space time.

Of the two options, the latter has been found to be of more practical use for unions of special relativity and quantum mechanics. In it, the quantum degrees of freedom have non-zero amplitude at all points in space time; in analogy to electric and magnetic fields which are defined over all of space-time, these quantum degrees of freedom are known as quantum field theory (QFT).

This thesis concerns itself with a study of various aspects of QFT. Before going into a more detailed introduction to the various aspects of quantum field theory of interest to us in this thesis, we would like to pause to make two remarks on the interplay between the two options. First, while these two options appear to be logically distinct, it is entirely possible that they are simply different ways of discussing the same physics. Certainly, the world-line formalism has played a crucial role throughout the history of quantum field theory, notably in Schwinger's proper time formalism.

Second, while the union of quantum mechanics and special relativity following option 2 has lead to the most successful, predictive, mathematical theories ever constructed (see below), option 1 is not without its merits. Specifically, if one were to endow the quantum world-histories with one more internal degree of freedom, call it a proper length, then rather than world-lines, these relativistic quantum states would sweep-out world sheets. Pursuits along this direction lead to string theory, which is widely regarded as the best candidate for a consistent theory of quantum gravity.

1.1 Local QFT and the Analytic S-matrix

Quantum field theory has many incarnations and formulations. By far the most powerful and ubiquitous formulation is through using path integrals over all possible quantum field configurations on the given (non-dynamical) spacetime manifold, exponentially weighted by the action associated with the given quantum field configuration. These path integrals, which take the schematic form,

$$Z[\varphi] = \int [D\varphi(x)] e^{\frac{i}{\hbar}\int d^d x \mathcal{L}[\varphi(x)]}\,, \tag{1.3}$$

where $\int d^d x$ is implicitly over the entire space-time manifold, are often taken to *define* a quantum field theory on a given space-time manifold.

Coupling the action to sources, $\mathcal{L} \to \mathcal{L} + J\varphi$, for the dynamical quantum field $\varphi(x)$, one can take functional derivatives with respect to the source J to derive correlation functions of the quantum fields whose interactions and dynamics are governed by the Lagrangian (density) $\mathcal{L}[\varphi(x)]$. The mechanical process of relating the various terms brought down by the functional derivatives is, essentially, a derivation of the Feynman rules for a theory. With the Feynman rules, one can in principle calculate all observables of a given QFT (so long as it has a Lagrangian description). What is more, one can recover the observed fact from classical mechanics that classical motion extremizes the action, by noting that when $\hbar \to 0$ quantum fluctuations which fail to extremize the action are infinitely penalized and thus have negligible contributions to measurable quantities. In short, the path integral allows one to manifestly see how quantum field theory reduces to classical, relativistic, physics as $\hbar \to 0$.

Path integrals defined in this way are manifestly quantum objects that are both relativistic and manifestly local. Each of these features are as plain as day: as emphasized above, relativity is ensured from the fact that the action integrals run over all points in space-time; quantum mechanics is ensured from the fact that all possible configurations of the quantum field have been considered; locality is of course guaranteed from the outset, as both the fields and the action depend explicitly on position. However, not all field theories can be described in this way. For example not all quantum field theories are endowed with Lagrangians. Correlation functions in such field theories are not found through straightforward manipulations of the path integral; in distinction to correlation functions in QFTs with known Lagrangian descriptions, which do derive from path integrals in straightforward ways.

Perhaps more surprisingly, an increasing body of evidence has found that scattering matrix elements, even in very well-known and theories that have valid actions and known Feynman rules, are heavily distorted when studied through the lens of local QFT. The classic example of this is the so-called "maximal helicity violating" or "MHV" scattering of n gluons off of each other. In the mid 1980s, Parke and Taylor found that if all gluons are treated as incoming and all but two of them have positive helicities, then all of the terms in the roughly $\mathcal{O}(n!)$ Feynman diagrams, each of which has $\sim n^{\#}$ distinct terms, collapse into one, single, term:

$$\left|A_n(1^+ \cdots a^- \cdots b^- \cdots n^+)\right|^2 = g^{2(n-2)} \frac{(p_a \cdot p_b)^4}{(p_1 \cdot p_2) \cdots (p_i \cdot p_{i+1}) \cdots (p_n \cdot p_1)} \,. \tag{1.4}$$

Since this seminal observation, a body of techniques have been steadily built-up to understand more complicated scattering processes in many diverse QFTs from totally different starting points.

One of the main techniques of the S-matrix is to complexify the momenta of scattering states. Through deforming the momenta of states participating in a scattering process in a complex direction, one promotes the scattering amplitude to a function of a complex variable. Exploring the plane of complex deformations of the amplitude probes the analytic structure of the amplitude; in the on-shell framework, the analytic structure of scattering amplitudes are paramount. Locality and unitarity imply, among other things, that tree amplitudes are rational functions of the external kinematic invariants (see discussion below Eq. (1.6) for details). Through clever use of locality and unitarity and other known properties of QFT such as spin-statistics, the value the deformed amplitude at special specific points in the complex plane are uniquely fixed in terms of more primitive information (coming from simpler scattering processes). Using these known values for the deformed amplitude for special values of the complex deformation, one exploits residue theorems to express the original un-deformed amplitude recursively, in terms of more primitive processes, without referring inherently local objects, such as interaction Lagrangians, in quantum field theory.

In the next few paragraphs, we consider aspects of the prototypical example system: on-shell gluon scattering. First, note that on-shell gluons only have physical, transverse, polarization states. Crucially, through interaction Lagrangians' point-like nature, interactions between gluons localizes them at a specific point in space-time. This, in turn, forces them off their mass-shells, and thereby introduces unphysical longitudinal polarization states. These unphysical polarization states that are introduced through off-shell tools, pollute and complicate intermediate steps in calculations of scattering amplitudes. Gauge-invariance constraints are needed to project-out these intermediate unphysical states.

It is worth emphasizing that, in calculations of perturbative scattering amplitudes, the gauge-invariance condition does nothing more than to project out these unphysical polarization states from the final result. Manifest locality, and its attendant gauge-invariance constraints, more than anything else, complicate scattering amplitude computations, and obscure the physical simplicity of the S-matrix, e.g. as in Eq. (1.4). Put another way, on-shell calculations of scattering amplitudes, i.e. calculations which do not ever use local interaction Lagrangians and which therefore never contain unphysical polarization states during intermediate steps, are manifestly physical and manifestly gauge-invariant throughout the calculation. This one important fact underlies a good deal of the simplicity manifested in the on-shell S-matrix. Again, for n-point MHV gluon scattering in Eq. (1.4), on-shell residue theorems replace the $\mathcal{O}(n!)$ Feynman diagrams with a handful of terms with clear physical interpretations.

Britto, Cachazo, Feng and Witten (BCFW), in a series of papers from late 2004 to early 2005, introduced the prototypical on-shell deformation for massless particles. Consider two gluon momenta, say p_1^μ and p_2^μ. By definition, $p_1^2 = -m^2 = 0$. Similarly $p_2^2 = 0$. It can be easily shown that in four-dimensions (and higher), one can always derive a third *complex* vector q^μ such that $q^2 = p_1 \cdot q = p_2 \cdot q = 0$.[2] The BCFW shift is simply this,

$$\begin{aligned} p_1^\mu \to p_1(z)^\mu &= p_1^\mu + zq^\mu \Longrightarrow (p_1(z))^2 = p_1^2 + 2z(q \cdot p_1) + z^2 q^2 = 0 \\ p_2^\mu \to p_2(z)^\mu &= p_2^\mu - zq^\mu \Longrightarrow (p_2(z))^2 = p_2^2 - 2z(q \cdot p_2) + z^2 q^2 = 0 \,. \end{aligned} \tag{1.5}$$

Deformations of this kind allow one to promote scattering amplitudes A_n between on-shell asymptotic states to functions of the complex BCFW deformation parameter: $A_n \to A_n(z)$. This promotion is more than cosmetic. If we consider the residue theorem,

$$\oint \frac{dz}{z} A_n(z) = 0 \Longrightarrow A_n(0) = -\sum_{z_P} \frac{A_n(z_P)}{z_P}\,, \tag{1.6}$$

where z_P is shorthand for the value of z where $A(z)$ develops a pole. So, if we know the rough analytic structure of scattering amplitudes, we can determine the nature of these poles. As mentioned above, tree amplitudes in QFT are rational functions of their kinematic invariants. Loop amplitudes however, have branch cuts. BCFW recursion relations are ideally suited to constructing tree-level scattering amplitudes.

If we consider the analytic structure of tree-level amplitudes in QFTs with local formulations, such as quantum chromodynamics and Yang-Mills theories, we know that the interaction Lagrangians contain either constants, inner products between polarization vectors, or derivatives of field variables. Similarly, the only source of momenta in denominators of scattering amplitudes comes from propagators, which have the generic form $1/P^2$, where P^2 is a simple kinematic invariant of the scattering process. When translated into the on-shell language of kinematic invariants, this corresponds to polynomial dependence on invariants in the numerator from interactions, and linear dependence on invariants in the denominator from propagators. And so tree amplitudes are simply, in the on-shell language, rational functions of their kinematic invariants with at most simple poles. Loop amplitudes are constructed from products of tree amplitudes where a subset of their momenta, the loop momenta, are internal and un-fixed by the external data. As these loop momenta are un-fixed,

[2] It is simple to construct such a vector. Boost to a frame where $p_1 = p(1, 0, 0, \cdots, 1)$ and $p_2 = p(1, 0, 0, \cdots, -1)$. Clearly $p_1^2 = p_2^2 = 0$. Note that $q = (0, 1, \pm i, 0, \cdots, 0)$ is perpendicular to both p_1 and p_2, and further is null.

they must be integrated over. In this way we understand that loop amplitudes, which are nothing other than integrals of products of rational functions (tree amplitudes), develop branch cuts.

Returning to (the simplest incarnation of) BCFW recursion, we see that if the amplitude in question is a tree amplitude, then it is a rational function of the the BCFW deformation parameter z. Poles in BCFW-deformed amplitudes in the complex z-plane of correspond to the BCFW-deformed kinematic invariants which enter into internal propagators in the scattering process going on-shell. Physically, when a particle goes onto its mass-shell, this corresponds to the particle propagating long distances. By unitarity, when this happens, we know the n-point amplitude factorizes into a product of two lower-point on-shell amplitudes.

Equipped with this physical understanding of the terms which occur in the residue theorem in Eq. (1.6), we see that by exploiting the shift in Eq. (1.5), we can recursively break-down higher-point tree-amplitudes into sums of products of lower-point tree amplitudes between gluons. As discussed extensively below, in particular in chapter 5, the simplest amplitudes between gluons involve three gluons. Their analytic forms are uniquely fixed by relativistic invariance. Through exploiting the BCFW recursion relations, one can construct the entire tree-level S-matrix between gluons, beginning from these primitive and uniquely fixed seeds. For example, through exploiting a few more techniques, BCFW-based analysis lends itself to a full reconstruction of the MHV amplitude in Eq. (1.4).

This is just one example of on-shell techniques which offer an alternative lens into the inner workings and structure of quantum field theory. One of the central motivations for the research that constitutes the bulk of this thesis is to understand the limitations of the on-shell program, specifically in regards to its ability to see: (1) the inherently finite-distance effect whereby electric charges (and other couplings in other theories) are modified at short, finite, distances; (2) to understand whether-or-not the manifestly gauge-invariant understanding of scattering from on-shell techniques can capture the "gauge anomalies" observed in local QFT; (3) understanding how to re-derive many classic results and consistency conditions on how massless objects interact in quantum field theory; and (4) a possible way to recast perturbative quantum gravity amplitudes without making any reference to space-time.

Formulations of QFT that keep locality manifest throughout, tend to obscure the physical properties of the S-matrix. The work in chapters 2–6 is a partial extension the on-shell formulation of QFT to scattering processes centrally involve of phenomena that, naively, are inherently "off-shell" and/or local. The general motivation here is to test the extent to which the on-shell program really furnishes an all-encompassing reformulation of QFT, in such a way that explicit reference to space and time is not built-in, while nevertheless describing relativistic quantum physics.

1.2 Partition Functions, QFT, and a Symmetry

If the on-shell language of the analytic S-matrix truly is rich enough to describe all of (perturbative) QFT, this could teach us a good deal about the character of relativistic quantum mechanics. What is simple in one language, e.g. S-matrix elements in on-shell formulations, tends to be difficult in the other. The S-matrix is a prominent observable which provides an excellent example that is natural to describe on-shell, but difficult to understand (in full detail) in local formulations of QFT. However, some phenomena which occur in quantum mechanical systems with relativistic degrees of freedom are more easily understood through local formulations of quantum field theory.

Phase transitions in quantum field theory and statistical physics, for example, are extremely important and are much better understood in local QFT than in the on-shell program. Indeed, one of the hallmarks of phase transitions is that there are qualitatively different effective degrees of freedom on either side of the transition. Let us keep QCD in mind, in this discussion. At low temperatures, quarks and gluons are confined into colorless bound states, called hadrons. The lightest hadron is the (isoscalar) pion; low energy dynamics of QCD are dominated by pion interactions. However, at sufficiently high temperatures, the pions (and all other hadrons) are "boiled" apart, and the fundamental constituents of the theory dictate the dynamics. Descriptions of physics in terms of the effective degrees one side of the phase transition break down in the vicinity of this phase transition for the simple reason that on the other side of the transition, the effective degrees of freedom change. It is not totally obvious how to put treatments of phase transitions on-shell.

There is a well-known, deep, and poorly understood connection between quantum mechanics and statistical phenomena. Specifically, it is known that path integrals of $d+1$ dimensional quantum field theory, when one of the spatial dimensions is a circle with periodicity $\tau \sim \tau + 2\pi\beta$, are equivalent to partition functions for quantum statistical mechanical systems in d-dimensions.[3] In more detail, a QFT on $d+1$-dimensional manifold $\mathcal{M}_d \times S^1_\beta$ is equivalent to quantum statistical mechanics, i.e. a quantum mechanical system coupled to a finite temperature bath with temperature $T = 1/\beta$, on the d-dimensional manifold $\mathcal{M}_d$. Phase transitions (amongst other phenomena) in statistical mechanics and in quantum field theories are understood using the same language.

[3] Another, simpler example—although in a slightly different vein—is between the heat equation and the Schrodinger equation. The heat equation which describes how heat/die propagates and diffuses, $\partial_t \phi(x,t) \propto \frac{d^2}{dx^2}\phi(x,t)$, is related to the Schrodinger equation which describes how wave-functions of massive quantum states behave in the absence of a potential, $i\hbar\partial_t \psi(x,t) \propto \frac{d^2}{dx^2}\psi(x,t)$.

In statistical physics, this breakdown of our description of physics, framed in terms of the effective degrees of freedom on one side of a phase transition approaches, a phase transition has a very clear signature. As the thermal statistical mechanical system approaches the phase transition, for instance as temperature $T = 1/\beta$ approaches some critical value $T_c = 1/\beta_c$, its partition function diverges. Schematically,

$$\lim_{\beta\to\beta_c} Z(\beta) = \lim_{\beta\to\beta_c} \sum_n d_n e^{-\beta E_n} \to \left(\frac{1}{\beta-\beta_c}\right)^{\#}, \tag{1.7}$$

where $\{E_n\}$ is the set of quantum mechanical energies in the system, d_n counts the number of distinct states at energy-level E_n, and $\#$ is a positive number.

One particular type of phase transition, discussed in this thesis (see specifically chapter 9), occurs when the number of states at a given energy grows exponentially with the energy, i.e. if $d_n \sim e^{\beta_H E_n}$ for some positive β_H. For such systems, when the inverse-temperature approaches this value, i.e. when $\beta \to \beta_H$, then the partition function diverges:

$$\lim_{\beta\to\beta_H} \sum_n d_n e^{-\beta E_n} \sim \lim_{\beta\to\beta_H} \sum_n e^{(\beta_H-\beta)E_n} \to \infty\,. \tag{1.8}$$

These phase transitions are known as Hagedorn phase transitions. QFTs with stringy behavior, such as quantum chromodynamics aka QCD (the theory of the strong force), are widely believed to have d_ns which indeed grow with energy. Such growth is called Hagedorn growth.

The aim of the second half of the thesis, chapters 7–9, is to study a surprising symmetry of partition functions in statistical mechanics and QFT. It is important to stress that this symmetry, which was first reported in the research papers which constitute this second part of the thesis, is in almost every sense of the word, not well understood. But, first, the symmetry. In chapter 7, we note that for partition functions of many diverse quantum systems,

$$Z(\beta) = \sum_n d_n e^{-\beta E_n}\,, \tag{1.9}$$

where all energies and degeneracies, i.e. $\{E_n, d_n\}$, are exactly known, we find a new symmetry:

$$Z(-\beta) = e^{i\gamma} Z(+\beta)\,. \tag{1.10}$$

Crucially, $e^{i\gamma}$ is independent of temperature, and is a complex number of modulus one.

This exact symmetry of partition functions under reflections of temperature has been dubbed temperature-reflection or T-reflection symmetry. We do not know its fundamental physical origin, or even if there is a simple, physical, origin of this symmetry. Nor are we aware of all of its broader consequences for physics at large, although it seems to be sensitive to vacuum energies of quantum systems (even in the absence of quantum gravity) as explored, chiefly, in chapters 7–8. Further, it implies that the certain toy model of QCD known to have Hagedorn behavior, the subject of chapter 9, have partition functions that are naturally written in terms of modular forms. However, as emphasized above, this T-reflection symmetry is not well understood. So more detailed discussion of the origin, structure, and consequences of T-reflection symmetry is deferred to the later, more technical, chapters.

1.3 Overview of this Thesis and Relation to Past Work

This thesis consists of nine chapters. The current chapter is introductory, and the only original work is the presentation. All of the facts are known. The research contained in this thesis is divided into eight chapters. Chapters 2–6 constitute a study of the on-shell S-matrix in massless theories in four- and six-dimensions. Chapters 7–9 constitute an exploration of partition functions and of T-reflection symmetry, in a variety of important exactly solved quantum systems.

Chapter 2: One-loop renormalization in the on-shell S-matrix

The research which underlies this chapter was conducted in collaboration with Y.-t. Huang, and C. Peng. It has been published, in slightly modified form, in Physical Review D. See Ref. [1] for details.

The project was suggested by Nima Arkani-Hamed. I performed the calculations of the bubble coefficients for non-supersymmetric gauge theories which constitute a bulk of the results in the paper. In particular, I noticed the pattern of cancellation of common collinear divergences (CCP) in both adjacent and non-adjacent MHV amplitudes; I performed the explicit analytic calculation for the case of non-adjacent MHV loop amplitudes which confirmed that CCP leaves only the terminal, double-forward poles, which are in turn trivially proportional to tree amplitudes in gauge theories; I noticed that CSW allows an inductive proof that CCP nullifies all contributions to the bubble coefficient save the same class of terminal double-forward poles, for the case of split-helicity N^{k}MHV amplitudes in Yang-Mills in four dimensions.

Chapter 3: Gauge-anomalies on-shell (short)

The research which underlies this chapter was conducted in collaboration with Y.-t. Huang. It has been published, in slightly modified form, in Physical Review Letters. See Ref. [2] for details.

I independently proposed this project, and developed it, through discussions, with Nima Arkani-Hamed and Yu-tin Huang. Yu-tin Huang proposed and performed the bulk of the original calculations which went into this work.

Chapter 4: Gauge-anomalies on-shell (long)

The research which underlies this chapter was conducted in collaboration with W.-m. Cheng, and Y.-t. Huang. It has been submitted, in slightly modified form, to Physical Review D. See Ref. [3] for details.

This project was an outgrowth of the previous project, and was concerned with more details of the structure of parity-odd loop amplitudes in chiral gauge theories in even dimensions, and in potentially anomalous quantum gravity theories in six- and ten-dimensions. I performed the bulk of the calculations which extracted the unique, gauge-invariant, rational terms in six-, eight-, and ten-dimensional chiral gauge theories, and also which extracted the parity-odd integral coefficients for these theories. I cross-checked results in all sections of the paper.

Chapter 5: Consistency conditions on four point amplitudes

The research which underlies this chapter was conducted in collaboration with L. Rodina. It has been published, in slightly modified form, in Physical Review D. See Ref. [4] for details.

The idea for this arose in conversations with Nima Arkani-Hamed and Laurentiu Rodina. I elucidated the organizing principle which guides the structure of the entire paper (the "(H, A)" classification of primitive three-point amplitudes between high-spin states); performed the bulk of the initial studies which used the guiding principle in conjunction with necessary corollaries of locality and unitarity used to rule-out all but one of the of three-point amplitudes excluded in the paper.

Chapter 6: Gravitons, Permutation Invariance, and Bonus-Scaling

The research which underlies this chapter was conducted in collaboration with L. Rodina. It has been submitted, in slightly modified form, to Physical Review Letters. See Ref. [5] for details.

The idea for this arose in conversations with Nima Arkani-Hamed and Laurentiu Rodina. Laurentiu and I played equal roles in all elements of the analysis.

Chapter 7: T-reflections

The research which underlies this chapter was conducted in collaboration with G. Basar, A. Cherman, and M. Yamazaki. It has been submitted, in slightly modified form, to Physical Review D. See Ref. [6] for details.

The idea for this project arose do to a mistaken check of modular invariance for partition functions of the toy model for QCD studied extensively in chapters 8 and 9. I took the mistake, i.e. invariance under T-reflections, seriously. After many conversations with many, many members of the hep-th group in Princeton, and after surveying and showing that many exactly solved model systems, culminating with all Virasoro minimal model 2d CFTs in which Masahito Yamasaki and I played an equal role, we wrote the paper. Other than noticing the fundamental interest and ubiquity of this symmetry and my joint role in the minimal model calculation with Masahito Yamazaki, all authors were of equal importance.

Chapter 8: Casimir energy of confining large N gauge theories

The research which underlies this chapter was conducted in collaboration with G. Basar, A. Cherman, and M. Yamazaki. It has been submitted, in slightly modified form, to Physical Review Letters. See Ref. [7] for details.

The idea for this project arose from conversations concerning T-reflection in the large N gauge theories considered in the T-reflection paper. Because T-reflection symmetry is sensitive to the vacuum energy of a theory, if it is invariant under T-reflection, this fixes the vacuum energy. A result found in the original T-reflection paper is that these theories are T-symmetric only when their vacuum energies vanish—even thought they do not have any supersymmetry. We confirmed this suggestive indication arising from T-reflection with two independent calculations. I played an equal role in confirming these calculations.

Chapter 9: Fermionic symmetries in QCD[Adj]

The research which underlies this chapter was conducted in collaboration with G. Basar, and A. Cherman. It has been submitted, in slightly modified form, to the Journal of High Energy Physics (JHEP). See Ref. [8] for details.

I joined this project after conversations with Gokce Basar and Aleksey Cherman, who published an initial shorter paper on emergent femionic symmetries in Summer of 2013. I noticed the T-reflection symmetry of the partition functions in QCD[Adj]—the central objects in this publication—and exploited their invariance under T-reflection to both (a) show that they are naturally written in terms of modular forms, and (b) to analytically calculate their Hagedorn phase transition temperatures.

Chapter 2

One-loop renormalization in the on-shell S-matrix

2.1 Introduction and summary of results

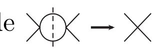
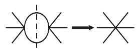
In four spacetime dimensions, integral reduction techniques [9, 10, 11] allow one to express one-loop gauge theory amplitudes in terms of rational functions and a basis of scalar integrals that includes boxes I_4, triangles I_3 and bubbles I_2 [10, 12, 13]:

$$A^{1-\text{loop}} = \sum_i C_4^i I_4^i + \sum_j C_3^j I_3^j + \sum_k C_2^k I_2^k + \text{rationals} \,. \tag{2.1}$$

Here the index i (j or k) labels the distinct integrals categorized by the set of momenta flowing into each corner of the box (triangle, or bubble). In this basis, the scalar bubble integrals, I_2^i, are the only ultraviolet divergent integrals in four dimensions. Moreover, the UV divergences of the bubble integrals take the universal form:

$$I_2^i = \frac{1}{(4\pi)^2}\frac{1}{\epsilon} + \mathcal{O}(1) \tag{2.2}$$

for all i. Thus the *sum of bubble coefficients* contains information on the ultraviolet behavior of the theory at one-loop.

In field theory, renormalizability requires that the ultraviolet divergences of the theory at one-loop can be removed by inserting a finite number of counterterms to the corresponding tree diagrams for the same process. We can also understand this renormalizability from the amplitude point of view. In terms of amplitudes, renormalizability implies that the ultraviolet divergence at one-loop must be proportional to the tree-amplitude. As we will see in detail below, this proportionality between tree amplitudes and the bubble coefficients, which encapsulate UV behavior of the theory, in renormalizable theories is cleanly illustrated in pure-scalar QFTs. In ϕ^4 theory, the bubble coefficient of the 4-point one-loop amplitude evaluates to the 4-point tree amplitude  . However, in ϕ^5 theory, the bubble coefficient of the simplest 1-loop amplitude evaluates to a new 6-point amplitude  . Similarly, this new 6-point tree amplitude will generate higher-point tree structures at higher loops, which is the trademark of a non-renormalizable theory.

This observation connects renormalizability with the 1-loop bubble coefficient: in a renormalizable theory, the *sum of bubble coefficients* is proportional to the tree amplitude

$$\mathcal{C}_2 \equiv \sum_i C_2^i \propto A^{\text{tree}} \,. \tag{2.3}$$

where the sum i runs over all distinct bubble cuts, and we use the calligraphic $\mathcal{C}_2$ to denote the *sum* of the bubble coefficients. This proportionality relation takes a very simple form in (super) Yang-Mills theory with all external lines being gluons [14, 15, 16] (see [17] for detailed discussion)

$$\mathcal{C}_2 = -\beta_0 A_n^{\text{tree}} \,, \quad \beta_0 = -\left(\frac{11}{3}n_v - \frac{2}{3}n_f - \frac{1}{6}n_s\right) . \tag{2.4}$$

where β_0 is the coefficient of the one-loop beta function and n_v, n_f, n_s are numbers of gauge bosons, fermions and scalars respectively. From the amplitude point of view, Eq. (2.4) appears to be a miraculous result as each individual bubble coefficient is now a complicated rational function of Lorentz invariants. For example, it is shown in [18, 17] that for the helicity amplitude $A_4(1^+2^-3^+4^-)$ in $\mathcal{N}$-fold super Yang-Mills theory, the bubble coefficients of the two cuts are:

$$\begin{aligned} C_2^{(23,41)} &= -(\mathcal{N}-4)\frac{\langle 12\rangle\langle 34\rangle}{\langle 13\rangle\langle 24\rangle}A_4^{\text{tree}}(1^+2^-3^+4^-) \,, \text{ and} \\ C_2^{(12,34)} &= -(\mathcal{N}-4)\frac{\langle 14\rangle\langle 23\rangle}{\langle 13\rangle\langle 24\rangle}A_4^{\text{tree}}(1^+2^-3^+4^-) \,. \end{aligned}$$

However the sum of these two bubble coefficients is exactly proportional to the tree amplitude: $C_2^{(23,41)} + C_2^{(12,34)} = -(\mathcal{N}-4)A_4^{\text{tree}}(1^+2^-3^+4^-) = -\beta_0 A_4^{\text{tree}}(1^+2^-3^+4^-)$ by the Schouten identity. For an arbitrary n-point amplitude, Eq. (2.4) implies cancellation among a large number of these rational functions, in the end yielding a simple constant multiplying A_n^{tree}. The fact that the proportionality in Eq. (2.4) holds for any renormalizable theory, hints at possible hidden structures in the sum of the bubble coefficients. Note that for gauge theories with non-adjoint matter fields, the individual bubble coefficients will also depend on higher order Casimir invariants [18]. Renormalizability then requires all the higher order invariants to cancel in the sum, leaving behind only the quadratic Casimir $tr_R(T^aT^b)$. In this chapter, we seek to partially expose hidden structure of the bubble coefficients that leads to the proportionality to the tree amplitude.

Following [19, 16], we extract the bubble coefficient by identifying it as the contribution from the pole at infinity in the complex z-plane of a BCFW-deformation [20] of the two internal momenta in the two-particle cut, where the complex deformation is introduced on the internal momenta.[1] We begin with scalar theories as a warm up. Here the contributions to bubble coefficients are tractable using Feynman diagrams in the two-particle cut. For scalar ϕ^n theories, we demonstrate that the bubble coefficient only receives contributions from one-loop diagrams that have exactly two loop-propagators. For each diagram, the contribution is proportional to a tree diagram with a new $2(n-2)$-point interaction vertex. Renormalizability requires $n = 2(n-2)$, so this implies the familiar result, $n = 4$.

Feynman diagrams become intractable in gauge theories and it is simpler to use helicity amplitudes in the cut. In (super) Yang-Mills theory, we study general MHV n-point amplitudes and find that for each 2-particle cut, the bubble coefficient can be separated into four separate terms. Each term stems from the four distinct singularities which appear as the loop momenta become collinear to one of the adjacent external legs, indicated in Fig. 2.1 (a). We show that these singularities localize the Lorentz invariant phase space (dLIPS-) integral to residues at four separate poles. Once given in this form, we find:

- *For each collinear residue in a generic cut, there is a residue in the adjacent cut that has the same form but with opposite sign.* When we sum over all channels, residues stemming from common collinear poles (CCP) in *adjacent* channels cancel pair-wise, as indicated in Fig. 2.1.The sum therefore telescopes to four unique poles that come from four distinct "terminal cuts". Here we define a terminal cut as the two particle cut which contains at least one 4-point tree amplitude on one side of the cut. The poles of interest correspond to the point in the phase-space where the two on-shell loop momenta become collinear with the two external scattering states in the 4-point sub-amplitude. We will refer to these poles as "terminal poles".

- Focusing on the terminal poles we find that their contributions to the bubble coefficients are non-trivial only if the helicity configuration of the particles crossing the cut is "preserved", i.e. the loop helicity configuration is the same as the external lines on the 4-point tree amplitude

[1] An example of a BCFW-deformation is in Eq. (2.6).

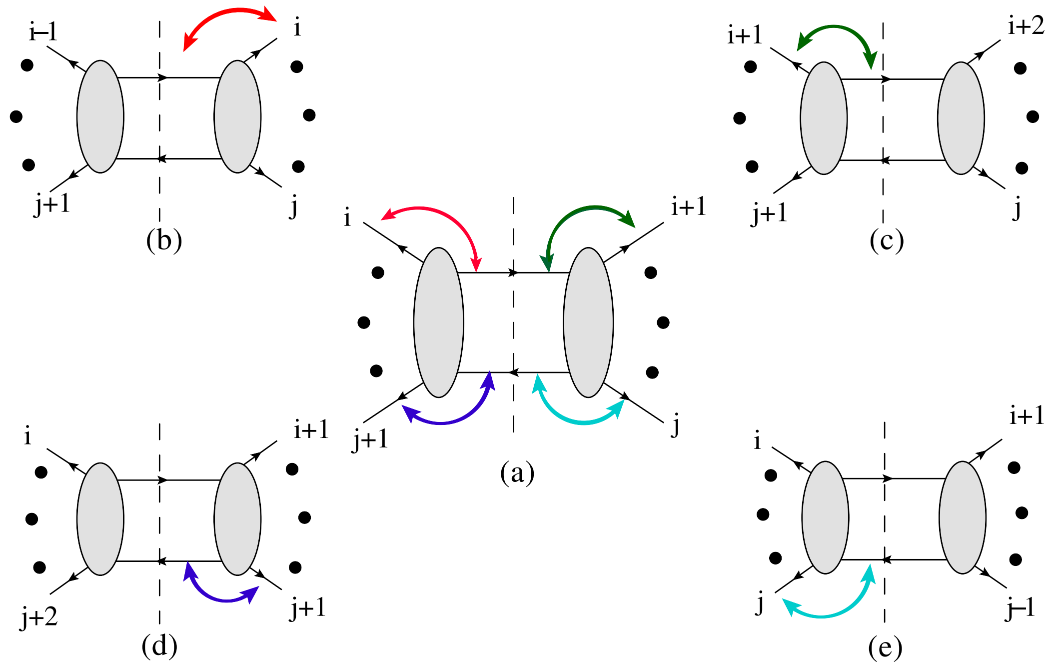


Figure 2.1: A schematic representation of the cancellation of common collinear poles (CCP). The bubble coefficient of the cut in figure (a), receives contributions from the four collinear poles indicated by colored arrows. Each collinear pole is also present, with the opposite sign residues, in the corresponding adjacent cut indicated in figures (b), (c), (d), and (e) respectively. In the sum of bubble coefficients such contributions cancel in pairs.

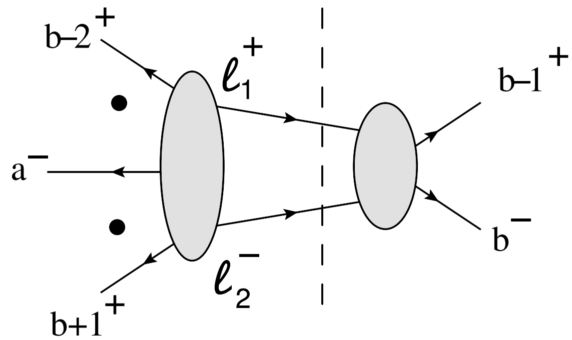


Figure 2.2: The terminal channels that gives non-trivial contribution to the sum of bubble coefficients. Note the helicity configurations of the loop legs of the n-point tree amplitude is identical with the two external legs on the 4-point tree amplitude in the cut.

> as shown in Fig. 2.2. Thus the beta function of (super) Yang-Mills theory is given by the residues of the helicity conserving terminal poles.

For MHV amplitudes, we show that there are two non-vanishing terminal poles whose residues are identical and equal to $11/6A_n^{\rm tree}$. Summing the two then gives the desired result, $\mathcal{C}_2 = 11/3A_n^{\rm tree}$ for the pure Yang-Mills theory, in agreement with Eq. (2.4). The relation (2.4) is also derived in the super Yang-Mills theory where $\mathcal{C}_2 = -(\mathcal{N}-4)\mathcal{A}_n^{\rm tree}$ for $\mathcal{N} = 1, 2$.

For general N^{k}MHV split-helicity amplitudes in pure Yang-Mills theory, we also show that the residue of each helicity conserving terminal pole give $11/6\,A_n^{\rm tree}$. We demonstrate this by using the CSW [21] representation for/expansion of the N^{k}MHV tree amplitudes appearing in the two-particle cut.[2] The fact that these terminal cuts give the correct proportionality factor indicates that these are indeed the only non-trivial contributions to the sum of bubble coefficients. This also hints at systematic cancellation in the sum of bubble coefficients should be a property of Yang-Mills amplitude for general helicity configuration. We give supporting evidence by using the collinear splitting function to show that the residues of CCP in a two particle cut for generic gauge theories are indeed identical with opposite signs.

This remainder of this chapter is organized as follows. In section 2.2, we compute the bubble coefficients for theories of self-interacting scalar fields, and rederive the well-known renormalizability conditions. We proceed to analyze (super) Yang-Mills theories with emphasis on the cancellation of common collinear poles (CCP) in section 2.4. We will use super Yang-Mills MHV amplitudes as the simplest demonstration of such cancellation. Similar results occur for MHV amplitudes in Yang-Mills as well. In section 2.5, we give an argument for the cancelation of CCP for generic external

[2] An example of a CSW expansion of an NMHV amplitude is given in Eq. (2.14).

helicity configurations by showing, using splitting functions of the tree amplitude in the cut, that the residue of collinear poles of the entire cut is indeed shared with an adjacent channel. We present further evidence in section 2.6 by explicitly proving that the forward limit poles for split-helicity N^kMHV amplitudes indeed give the complete RHS of Eq. (2.4), implying complete cancellation of all other contributions.

2.2 Bubble coefficients in scalar field theories

As a toy model, we consider scalar theories with single interaction vertex $\alpha_k \phi^k$ in this section. It was shown in [16], following previous work in [19], that the bubble coefficient for a given two-particle cut can be calculated as:

$$C_2^{(\mathrm{i,j})} = \frac{1}{(2\pi i)^2} \int d\mathrm{LIPS}[l_1, l_2] \int_{\mathcal{C}} \frac{dz}{z}\, \widehat{S}_n^{(\mathrm{i,j})} \tag{2.5}$$

where (i,j) indicates the momentum channel $P = p_{i+1} + \cdots + p_j$ of the cut as shown in Fig. 2.7, $\widehat{S}_n^{(\mathrm{i,j})} = \hat{A}_L^{\mathrm{tree}}(|\hat{l}_1\rangle, |\hat{l}_2]) \, \hat{A}_R^{\mathrm{tree}}(|\hat{l}_1\rangle, |\hat{l}_2])$, and dLIPS$=$ $d^4l_1 d^4l_2\, \delta^{(+)}(l_1^2)\, \delta^{(+)}(l_2^2)\, \delta^4(l_1 + l_2 - P)$. Here $\hat{A}_{L,R}^{\mathrm{tree}}$ in (2.5) are the amplitudes on either side of the cut; hats in (2.5) indicate a BCFW shift [20] of the two cut loop momenta:

$$\widehat{l}_1(z) = l_1 + qz, \qquad \widehat{l}_2(z) = l_2 - qz\,, \qquad \text{with} \;\; q\cdot q = q\cdot l_1 = q \cdot l_2 = 0\,. \tag{2.6}$$

We integrate the shift parameter z along a contour $\mathcal{C}$ that goes around infinity, which evaluates to the residue at the $z = \infty$ pole of the integrand.[3]

In a scalar theory, the only z dependence in BCFW-shifted tree-amplitudes comes from propagators which depend on one of the two loop momenta. Under BCFW-deformations, propagators of this type scales as $\sim 1/z$ for large-z. Diagrams containing such propagators die-off as $1/z$ or faster. The only non-vanishing contribution to the bubble coefficient comes from diagrams with the two shifted lines on the same vertex [22]. In this case there is neither z-dependence nor dependence on l_1, or l_2 in the double-cut and (2.5) evaluates to

$$C_2^{(\mathrm{i,j})} \;\; = \;\; -\frac{1}{2\pi i} \int d\mathrm{LIPS}\; A_L^{\mathrm{tree}} A_R^{\mathrm{tree}} = A_L^{\mathrm{tree}} A_R^{\mathrm{tree}}\,, \tag{2.7}$$

[3]The BCFW shifts of the two-particle cut allows one to explore all possible on-shell realizations of a double-cut for a given set of kinematics. The presence of finite-z poles indicates the existence of additional propagators, which are the contributions of box or triangle integrals to the double-cut. The contribution from the bubble integrals then correspond to poles at $z = \infty$, hence the choice of contour.

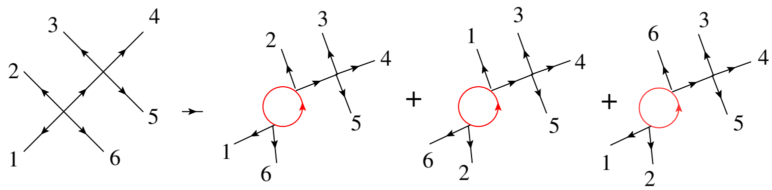


Figure 2.3: For any given tree-diagram in the ϕ^4 theory, each vertex can be blown up into 4-point one-loop subdiagrams in three distinct ways, while preserving the tree graph propagators. Each case contributes a factor of α_4 times the original tree diagram to the bubble coefficient. In this figure we show the example of 6-point amplitude.

where $A^{\text{tree}}_{L,R}$ are the unshifted amplitudes on either side of the cut as in Fig. 2.4, and we have used $\frac{1}{2\pi i}\int d\text{LIPS}(1) = -1$ (discussed in the following section, section 2.3). The bubble coefficient (2.3) is a sum over all cuts.

At 4-point, the tree amplitude is $A^{\text{tree}}_4 = \alpha_4$. There are two cuts of the 1-loop 4-point amplitudes, namely the s and t channels. Then (2.7) gives

$$\begin{aligned}\mathcal{C}_2 &= A^{\text{tree}}_4(1,2,\hat{l}_1,\hat{l}_2)\times A^{\text{tree}}_4(-\hat{l}_2,-\hat{l}_1,3,4) + A^{\text{tree}}_4(4,1,\hat{l}_1,\hat{l}_2)\times A^{\text{tree}}_4(-\hat{l}_2,-\hat{l}_1,2,3)\\ &= \alpha_4^2+\alpha_4^2 = 2\alpha_4 A^{\text{tree}}_4\,. \end{aligned}\tag{2.8}$$

This analysis extends to all-n in ϕ^4-theory: each bubble-cut with a non-vanishing large-z pole will be precisely of this form. Evaluating the pole at infinity, and integrating over phase-space reproduces a tree-diagram. A semi-detailed analysis reveals that, somewhat unsurprisingly, the tree diagrams generated in this way correctly reproduce (a result which is) proportional to the original tree amplitude,

$$C_2^{(n)}|_{\phi^4} = \frac{3(n-2)}{2}\alpha_4 A^n_{\text{tree}}\ . \tag{2.9}$$

A sketch of how this procedure is implemented, in practice, is depicted in Figure 2.3.

One can do a similar analysis to the Yukawa theory with complex scalars. Here, Yukawa theory has asymptotic states of non-zero helicity; the analysis is somewhat aided through explicit use of tree-level helicity amplitudes, as opposed to an approach based solely on Feynman diagrams. Similar results hold. Details of this analysis are omitted here as Yukawa theory is well-understood.

Crucial new aspects of, and critical uses for, integral reduction in conjunction with spinor-helicity technology manifest themselves in purest form in (S)YM. Use of spinor-helicity technology to describe tree amplitudes on either side of the bubble-cut forces all gluons (and their supersymmetric cousins)

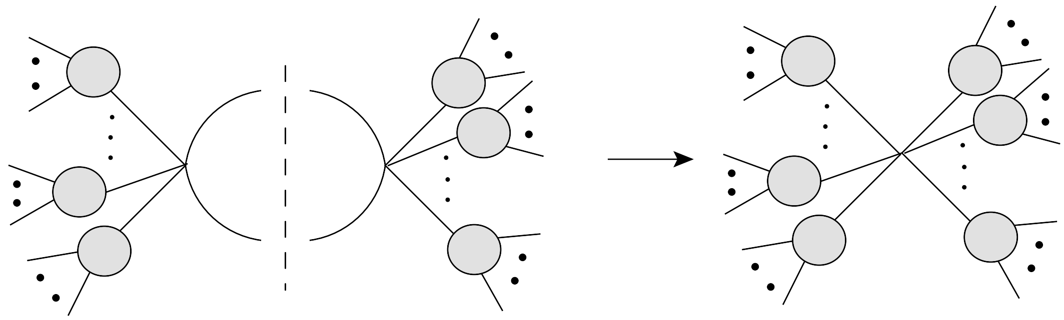


Figure 2.4: For pure ϕ^k theory, the only one-loop diagrams that gives non-trivial contribution to the bubble integrals are those with only two loop propagator. The contribution to the bubble coefficient is simply the product of the tree diagrams on both side of the cut, connected by a new $2(k-2)$ vertex.

to be on-shell, and eliminates unphysical degrees of freedom from the calculation. As we shall see presently, this vastly simplifies calculations of the one-loop beta-function in QCD (YM and SYM as well).

2.3 dLIPS integrals, via the holomorphic anomaly in 4-dimensions

In this brief technical section, we introduce the techniques used in this thesis to evaluate $\int d\mathrm{LIPS}$ for on-shell one-loop amplitudes in four-dimensional massless S-matrix elements. It uses the techniques of the holomorphic anomaly, closely follows the original discussions in the literature in Refs. [16, 17, 21]. These techniques allow us to calculate integrals of the following form,

$$\oint_{\tilde{l}=\bar{l}} P^2 \frac{\langle \lambda, dl\rangle [\tilde{l}, d\tilde{l}]}{\langle l|P|\tilde{l}]^2} \frac{\prod_{i=1}^{n}[a_i, \tilde{l}]}{\langle l|P|\tilde{l}]^n} g(l) \text{ , where } g(l) = \frac{\prod_{j=1}^{m}\langle b_j, \lambda\rangle}{\prod_{k=1}^{m}\langle c_k, \lambda\rangle} \tag{2.10}$$

where the integral over phase-space (dLIPS integral) is really a contour integral over two complex numbers. Two cases are important here: $n = 0$ for scalar- and Yukawa-theory, and $n = 2$ for gauge theories. Note:

$$P^2 \frac{[\tilde{l}, d\tilde{l}]}{\langle l|P|\tilde{l}]^2} = -d\tilde{l}^{\dot{\alpha}} \frac{\partial}{\partial \tilde{l}^{\dot{\alpha}}} \left(\frac{[\tilde{l}, \tilde{\eta}]P^2}{\langle l|P|\tilde{l}]\langle l|P|\tilde{\eta}]} \right) = -d\tilde{l}^{\dot{\alpha}} \frac{\partial}{\partial \tilde{l}^{\dot{\alpha}}} \left(\frac{[\tilde{l}|P|\alpha\rangle}{\langle l|P|\tilde{l}]\langle \lambda, \alpha\rangle} \right). \tag{2.11}$$

where we have introduced reference spinors $|\tilde{\eta}] = P|\alpha\rangle$ in order to express the dLIPS integration measure as a total derivative. We further note that integrands of the form (2.10) can be reduced to this basic measure through repeated differentiation. Concretely, for $n = 2$:

$$\frac{[I,\tilde{l}][J,\tilde{l}]}{\langle l|P|\tilde{l}]^4} \quad = \quad \frac{1}{6}\tilde{I}^{\dot{\gamma}}\tilde{J}^{\dot{\beta}}\frac{\partial^2}{\partial(\langle l|P)^{\dot{\beta}}\partial(\langle l|P)^{\dot{\gamma}}}\left\{\frac{1}{\langle l|P|\tilde{l}]^2}\right\}. \tag{2.12}$$

For the case of MHV bubble integrands, the only reference spinors are of the form $|I] = P|i\rangle$. Combining (2.11) and (2.12), and interchanging the order of differentiation, one can re-write the "$n = 2$" integrand as a total derivative:

$$\begin{aligned}
&\oint_{\tilde{l}=\bar{l}} P^2\frac{\langle\lambda, dl\rangle[\tilde{l}, d\tilde{l}]}{\langle l|P|\tilde{l}]^4}\langle i|P|\tilde{l}]\langle j|P|\tilde{l}]g(l) \\
&= \oint_{\tilde{l}=\bar{l}}\langle\lambda, dl\rangle\Bigg(-d\tilde{l}^{\dot{\gamma}}\frac{\partial}{\partial\tilde{l}^{\dot{\gamma}}}\Bigg[\frac{[\tilde{l}|P|\alpha\rangle\ g(l)}{\langle l|P|\tilde{l}]\langle l,\alpha\rangle}\frac{1}{3}\Bigg\{\frac{\langle i|P|\tilde{l}]\langle j|P|\tilde{l}]}{\langle l|P|\tilde{l}]^2} + \frac{\langle i,\alpha\rangle\langle j,\alpha\rangle}{\langle\lambda,\alpha\rangle^2} \\
&\qquad\qquad + \frac{1}{2}\frac{\langle i|P|\tilde{l}]\langle j,\alpha\rangle + \langle i,\alpha\rangle\langle j|P|\tilde{l}]}{\langle l|P|\tilde{l}]\langle\lambda,\alpha\rangle}\Bigg\}\Bigg]\Bigg).
\end{aligned} \tag{2.13}$$

The last form for the integrand, re-written as a total derivative, vanishes at all points save when it hits a simple pole. This is because along the integration contour $\tilde{l} = \bar{l}$, one has [21],

$$-d\tilde{l}^{\dot{\alpha}}\frac{\partial}{\partial\tilde{l}^{\dot{\alpha}}}\frac{1}{\langle\lambda,\xi\rangle} = -2\pi\bar{\delta}\left(\langle\lambda,\xi\rangle\right), \tag{2.14}$$

$$\tag{2.15}$$

Thus the dLIPS integral is localized to the poles $1/\langle\lambda,\xi\rangle$ of the integrand.[4] Each term in the integrand (2.13) has potential collinear divergences coming from the spinor brackets in the denominator of $g(l)$ and that of the reference spinor $\langle l,\alpha\rangle \to 0$.

Through (2.11) and (2.14), we see that the simple bubble integrals in scalar QFT in section 2.2 simply evaluate to $\int d\text{LIPS}(1) = -2\pi i$. For the MHV bubble integrands which we will encounter below, such as Eq. (2.31), there is always a choice of the reference spinor, such as $|\alpha\rangle = |a\rangle$, which eliminates the unphysical $1/\langle\lambda,\alpha\rangle$ pole.

2.4 Bubble coefficients for MHV (super) Yang-Mills amplitude

When we consider the (super) Yang-Mills theory, the proportionality between the sum of bubble coefficients and the tree amplitude becomes extremely non-trivial. Here, individual bubble coeffi-

[4] Note that $1/\langle l|P|\tilde{l}]$ is not a simple pole on the contour $\tilde{\lambda} = \bar{\lambda}$.

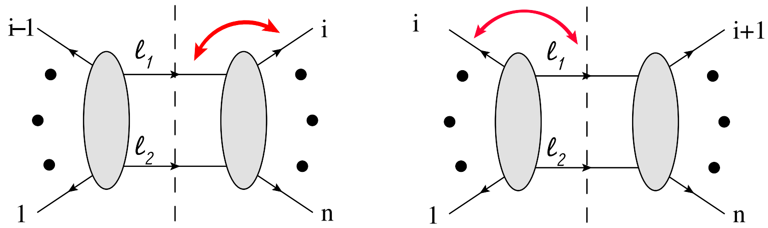


Figure 2.5: An illustration of the cancellation between adjacent channels. The contribution to the bubble coefficient coming from the dLIPS integral evaluated around the collinear pole $\langle l_1 i\rangle \to 0$, indicated by the (red) arrows, of the two diagrams cancels as indicated in Eq. (2.33).

cients are generically complicated rational functions of spinor inner products as illustrated for the $\langle\Phi_1\Psi_2\Phi_3\Psi_4\rangle$ case in the introduction. In general, only after summing all the bubble coefficients and repeated use of Schouten identities, will the result reduce to a simple constant times $\mathcal{A}_n^{\text{tree}}$. Thus from the amplitude point of view, this proportionality is a rather miraculous result.

In this section we show that the cancellation is in fact systematic. To see this, we show that for MHV amplitudes, the dLIPS integration will be localized by the collinear poles of the tree amplitude on both sides of the two-particle cut. For a generic cut, there are four distinct collinear poles involving the loop legs, each of which is also present in an adjacent cut, as illustrated in Fig. 2.5. It can be shown that the residues of these two adjacent cuts on their common collinear pole, $\langle\lambda, i\rangle \to 0$, are exactly equal and with opposite sign. By separating the bubble coefficient into four different terms, corresponding to contributions from four different poles, the cancellation between common collinear poles (CCP) in the sum of bubble coefficients is manifest.

Cancellation stops at "terminal cuts" where a 4-point tree and an n-point tree appear on opposites sides of the cut. The uncancelled terms in these terminal cuts correspond to the residues of collinear poles where the two loop-momenta become collinear with the external momenta of the two external legs on the 4-point amplitude, as illustrated in Fig. 2.6. Explicitly, for *adjoint* fields (vectors, fermions and scalars), we see the sum of these "terminal poles" is

$$-\beta_0 A_n^{\text{tree}}(1^+\cdots a^-,\lambda\cdots,b^-\cdots n^+) = \left(\frac{11}{3}n_v - \frac{2}{3}n_f - \frac{1}{6}n_s\right)\frac{\langle a,b\rangle^4}{\langle 1,2\rangle ... \langle i,i+1\rangle ... \langle n,1\rangle}\,, \qquad (2.16)$$

for MHV amplitudes with $n-2$ positive-helicity gluons and negative-helicity gluons a and b [23]. In the following, we will demonstrate this for n-point MHV amplitudes in $\mathcal{N}=1,2$ super Yang-Mills theory. This systematic cancellation is also present for pure Yang-Mills MHV amplitudes (explicitly shown in section 2.7).

Before going further, we pause to note an important distinguishing feature between the bubble

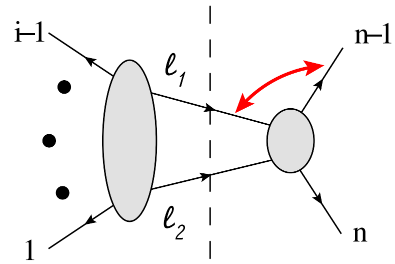


Figure 2.6: The "terminal" pole that contributes to the bubble coefficient. Such poles appear in the two particle cuts that have two legs on one side of the cut and one of the legs has to be a minus helicity. Note that, at this point in phase-space, $l_1 = -p_{n-1}$ and $l_2 = -p_n$.

coefficients in scalar QFT and in (S)YM. Specifically, the proportionality constant in (super) Yang-Mills is independent of the number of external legs: it is just -β_0, the coefficient of the one-loop beta function. To see this note that for (super)Yang-Mills theory, there are diagrams with one-loop bubbles in the external legs. These massless bubbles do not appear in Eq. (2.1), as they are set to zero in dimensions regularization, reflecting the cancellation between collinear IR and UV divergences. However, when one is only considering the pure UV divergence of the amplitude, one must take into the account the existence of the UV divergences in the external bubble diagrams, which are simply the same as that of the infrared divergences in the external bubbles, but with a relative minus sign. Thus we have

$$\begin{aligned} A_n|_{\text{UV-div.}} &= \Big(\sum C_{\text{bubble}} I_{\text{bubble}}\Big)_{\text{UV}} + \text{UV}_{\text{ext. bubbles}} \\ &= \Big(\sum C_{\text{bubble}} I_{\text{bubble}}\Big)_{\text{UV}} - \text{IR}_{\text{ext. bubbles}} \,. \end{aligned} \tag{2.17}$$

For n-gluon 1-loop amplitudes, the collinear IR divergences take the form [14]

$$\text{IR:} \qquad A^{\text{1-loop}}_{n,\text{collinear}} = -\frac{g^2}{(4\pi)^2}\frac{1}{\epsilon}\,\frac{n}{2}\,\beta_0 A_n^{\text{tree}} \,. \tag{2.18}$$

At leading order in $\epsilon \to 0$, the UV divergence is [14]

$$\text{UV:} \qquad A^{\text{1-loop}}_{n,\text{UV}} = +\frac{g^2}{(4\pi)^2}\frac{1}{\epsilon}\left(\frac{n-2}{2}\right)\beta_0 A_n^{\text{tree}} \,. \tag{2.19}$$

Thus the bubble coefficients (total UV divergence) in *purely* gluonic one-loop amplitudes are

$$\sum_i C_2^i = A^{\text{1-loop}}_{n,\text{UV}} + A^{\text{1-loop}}_{n,\text{collinear}} = -\beta_0\, A_n^{\text{tree}} = \frac{11}{3}\, A_n^{\text{tree}} \,. \tag{2.20}$$

At one loop, ϕ^4 scalar field theory lacks these collinear divergences on external legs, and no UV/IR mixing occurs, hence pure scalar bubble coefficients scale with $\frac{n-2}{2}$, the number of interaction vertices.

2.4.1 Extracting bubble coefficients in ($\mathcal{N}=0,1,2$ super) Yang-Mills

The bubble coefficient for a given two-particle cut of a one-loop (S)YM amplitude is computed in essentially the same way as for scalar field theory. However, as emphasized in the introduction, unlike the case for scalar QFT extracting this through Feynman diagrams is rather intractable. Roughly in YM this is because BCFW shifts of the two internal on-shell gluon lines in the double-cut introduces z-dependence in local interaction vertices *and* polarization vectors. These difficulties are only amplified in ($\mathcal{N}\neq 0$) SYM.

It is more efficient to directly express the LH- and RH- amplitudes as entire on-shell objects through use of the spinor-helicity formalism. Here the dLIPS integration over allowed on-shell momenta is conveniently converted into an integration over spinor variables which automatically solve the delta functions,

$$\int d^4l_1 d^4l_2\, \delta^{(+)}(l_1^2)\, \delta^{(+)}(l_2^2)\, \delta^4(l_1+l_2-P) g(|l_1\rangle,|l_2\rangle) = \int_{\tilde{l}=\bar{l}} P^2 \frac{\langle\lambda, dl\rangle[\tilde{l}, d\tilde{l}]}{\langle l|P|\tilde{l}]^2} g(|l\rangle, P|\tilde{l}])\,, \qquad (2.21)$$

where we have identified $|l_1\rangle = |l\rangle$, $|l_2\rangle = P|\tilde{l}]$, and $\int_{\tilde{l}=\bar{l}}$ indicates we are integrating over the real contour (real momenta).[5] The $\widehat{S}_n^{(\mathrm{i,j})}$ in (2.5) takes the form

$$\widehat{S}_n^{(i,j)} = \widehat{\mathcal{S}}_{n,\,0}^{(i,j)} \equiv \sum_{\text{state sum}} \hat{A}_L^{\text{tree}}(|\hat{l}_1\rangle, |\hat{l}_2]) \, \hat{A}_R^{\text{tree}}(|\hat{l}_1\rangle, |\hat{l}_2])\,, \qquad (2.22)$$

in the Yang-Mills theory. Note that to fully integrate out the bubble coefficients' dependence on the internal lines, we sum over all possible states in the loop.

Further, extraction of simple bubble coefficients is aided by on-shell SUSY.[6] Here amplitudes and state sums are promoted to superamplitudes and Grassmann integrals

$$\widehat{S}_n^{(i,j)} = \hat{\mathcal{S}}_{n,\mathcal{N}}^{(i,j)} \;\equiv\; \sum_{\sigma} \int d^{\mathcal{N}}\eta_{l_1} d^{\mathcal{N}}\eta_{l_2}\; \hat{\mathcal{A}}_{L\sigma}^{\text{tree}}\, \hat{\mathcal{A}}_{R\bar{\sigma}}^{\text{tree}}\,, \qquad \mathcal{N}=1,2\,, \qquad (2.23)$$

where σ labels the different pairs of multiplets that the loop legs, l_1 and l_2, belong to. Following [17],

[5] The explicit evaluation of integrals in Eq. (2.21) using the holomorphic anomaly [21, 16, 17] is reviewed in the previous section, section 2.3.

[6] The calculations for the simplest bubble coefficients are simpler in $\mathcal{N}=1,2$ SYM than in YM. To see this, compare non-adjacent MHV bubble computations in SYM (subsection 2.4.2) and in YM (section 2.7).

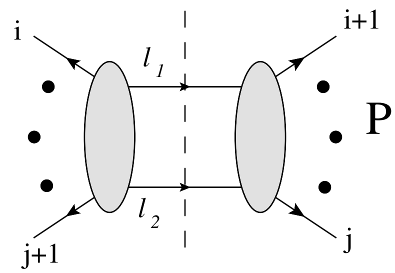


Figure 2.7: A two-particle cut for a generic n-point amplitude.

on-shell states are encoded into two separate on-shell superfields, Φ and Ψ, that contain states in the 'positive' and 'negative-helicity' sectors. In this language, $\{\sigma\} = \{(\Phi,\Psi),(\Psi,\Phi),(\Phi,\Phi),(\Psi,\Psi)\}$. The $\bar{\sigma}$ is the conjugate configuration of σ.

Crucially, to preserve SUSY the *bosonic* BCFW shift (2.6) must be combined with a *fermionic* shift of the Grassmann variables η^a [24, 16]

$$|\widehat{l_1}(z)\rangle \quad = |l_1\rangle + z|l_2\rangle, \qquad\qquad |\widehat{l_2}(z)] = |l_2] - z|l_1]\,, \tag{2.24a}$$

$$\hat{\eta}_{l_2 a} \quad = \eta_{l_2 a} + z\eta_{l_1 a}, \qquad\qquad a = 1, \lambda \cdots, \mathcal{N}\,. \tag{2.24b}$$

Note the bosonic shift (2.24a) is identical to the shift (2.6), when cast in terms of the spinor-helicity variables; it is referred to as an $[l_2,\, l_1\rangle$-shift.

Combined super-shifts (2.24), of any tree amplitude of the $\mathcal{N}=4$ SYM fall-off as $1/z$ for large-z. In (S)YM theory with $\mathcal{N}=0,1,2$ supersymmetry, it was shown [17] that the super-BCFW shifts $[\Phi,\Phi\rangle$, $[\Psi,\Phi\rangle$ and $[\Psi,\Psi\rangle$ fall off as $1/z$ at large z while the $[\Phi,\Psi\rangle$ super-shift grows as $z^{3-\mathcal{N}}$ for large-z. For $\mathcal{N}=0$ pure Yang-Mills, this reduces to the familiar observation that for shifts $[-,-\rangle$, $[-,+\rangle$, and $[+,+\rangle$ the amplitudes fall off as $1/z$, while the $[+,-\rangle$ shifts grow as z^3 [20, 25].

Carrying out the z integral gives

$$C_2^{(i,j)} \quad = \quad -\frac{1}{2\pi i}\int d\text{LIPS}[l_1,l_2]\,\left[\widehat{\mathcal{S}}^{(i,j)}_{n,\mathcal{N}}\right]_{O(1)\text{ as } z\to\infty}\,,\ \mathcal{N}=0,1,2\,, \tag{2.25}$$

where $\left[\widehat{\mathcal{S}}^{(i,j)}_{n,\mathcal{N}}\right]_{O(1)\text{ as } z\to\infty}$ is the residue of $\widehat{\mathcal{S}}^{(i,j)}_{n,\mathcal{N}}$ at the $z\to\infty$ pole.

Double-cuts with internal states $\sigma\in\{(\Phi,\Phi),(\Psi,\Psi)\}$, shown in cut (a) of Fig. 2.8, scale as

$$\widehat{\mathcal{S}}^{[\text{Cut (a)}]}_{n,\mathcal{N}} \ \sim\ \frac{1}{z}\times\frac{1}{z}\ \sim\ \frac{1}{z^2}\quad\text{as}\quad z\to\infty\,. \tag{2.26}$$

Cuts of this type do not contribute to the bubble coefficient. On the other hand, cuts with internal

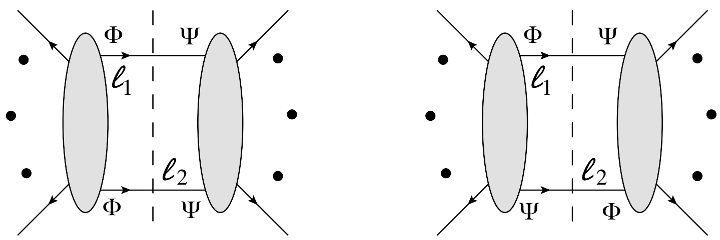


Figure 2.8: Illustration of two internal helicity configurations of the cut loop momenta.

states $\sigma \in \{(\Phi, \Psi), (\Psi, \Phi)\}$, such as cut (b) in Fig. 2.8 , always involve a shift that acts as $[\Psi, \Phi\rangle$ on one sub-amplitude and as $[\Phi, \Psi\rangle$ on the other. This gives

$$\widehat{\mathcal{S}}_{n,\mathcal{N}}^{[\text{Cut (b)}]} \sim \frac{1}{z} \times z^{3-\mathcal{N}} \sim z^{2-\mathcal{N}} \quad \text{as} \quad z \to \infty\,. \tag{2.27}$$

Note that (2.27) indicates that there can be non-vanishing $O(1)$-terms and hence contributions to the sum of the bubble coefficients in $\mathcal{N} = 0, 1, 2$ SYM but not in the $\mathcal{N} = 4$ SYM theory. This is consistent with the known non-vanishing 1-loop β-functions in $\mathcal{N} = 0, 1, 2$ SYM theories as well as the UV-finiteness of $\mathcal{N} = 4$ SYM.

We can also investigate the contribution to the bubble coefficients from particular states crossing the two-particle cut. These contributions can be projected out by acting with appropriate Grassmann integrations/derivatives on the above superamplitudes. By analyzing the large-z behavior of the integration measure, as shown in [17], we get the large-z behavior of the bubble coefficient of certain internal states. The implication for QCD with one flavor of fermions can simply be deduced from $\mathcal{N} = 1$ super Yang-Mills, where there are no scalars in the multiplet. For example, a simple computation shows that a double-cut of an internal negative-helicity gluino and an internal negative helicity gluon exiting (entering) one of the two tree-amplitudes does not contribute to the bubble coefficients in $\mathcal{N} = 1$ SYM theory. Then following [26], this result is also true in QCD with one flavor of fermions. Note that the difference between fundamental and adjoint fermions is irrelevant for this analysis since we are interested in color-ordered amplitudes and the large-z behavior of the cut holds for individual internal helicity configurations and not the sum.

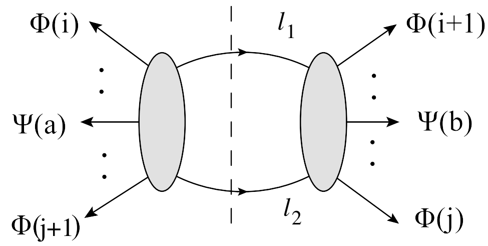


Figure 2.9: The two-particle cut that gives $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}$.

2.4.2 MHV bubble coefficients in $\mathcal{N}=1,2$ super Yang-Mills theory

It was shown in ref. [17], that for the MHV amplitudes in $\mathcal{N}=1,2$ super Yang-Mills theory, the $\mathcal{O}(z^0)$ part of the BCFW-shifted two-particle cut $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}$, depicted in Fig. 2.9 is given by:

$$\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}\Big|_{\mathcal{O}(z^0)} = (\mathcal{N}-4)\mathcal{A}^{\rm tree}_n \frac{\langle i,i+1\rangle\langle j,j+1\rangle}{\langle a,b\rangle^2}\frac{\langle a,\lambda\rangle^2\langle b,\lambda\rangle^2}{\langle j,\lambda\rangle\langle j+1,\lambda\rangle\langle i,\lambda\rangle\langle i+1,\lambda\rangle}, \tag{2.28}$$

where $\{a,b\}$ indicates the positions of the two sets of external negative-helicity states (within the Ψ multiplets). We have set $|l_2\rangle = |l\rangle$. Since $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}$ is purely holomorphic in l, we can straightforwardly use Eq. (2.11) to rewrite the dLIPS integral (2.21) as a total derivative, and the bubble coefficient is given as:

$$\begin{aligned} C^{(i,j)}_{2\ \{a,b\}} &= \frac{-1}{2\pi i}\oint_{\tilde{l}=\bar{l}} P^2_{i+1,j}\frac{\langle\lambda,dl\rangle[\tilde{l},d\tilde{l}]}{\langle l|P_{i+1,j}|\tilde{l}]^2}\ \widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}\Big|_{\mathcal{O}(z^0)} \\ &= \frac{1}{2\pi i}\oint_{\tilde{l}=\bar{l}}\langle\lambda,dl\rangle d\tilde{l}_{\dot\alpha}\frac{\partial}{\partial\tilde{l}_{\dot\alpha}}\left[\frac{[\tilde{l}|P_{i+1,j}|\alpha\rangle}{\langle l|P_{i+1,j}|\tilde{l}]\langle\lambda,\alpha\rangle}\ \widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}\Big|_{\mathcal{O}(z^0)}\right], \end{aligned} \tag{2.29}$$

where $|\alpha\rangle$ is an auxiliary reference spinor. In this section, for convenience, we label the bubble coefficients $C^{(i,j)}_{2\ \{a,b\}}$ in the same way as the two-particle cut $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}$. There are two kinds of poles inside the total derivative, the four collinear poles of $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}$ in Eq. (2.28) and the spurious pole $1/\langle\lambda,\alpha\rangle$. The spurious pole can be simply removed by the $\langle a,\lambda\rangle^2\langle b,\lambda\rangle^2$ factor in the numerator of Eq. (2.28) if we choose the auxiliary spinor $|\alpha\rangle$ to be $|a\rangle$ or $|b\rangle$. Thus with this choice of reference spinor, the contributions to the bubble coefficient come solely from the collinear poles in $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}|_{\mathcal{O}(z^0)}$.

From Eq. (2.28) we see that there are four collinear poles in $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}|_{\mathcal{O}(z^0)}$, each corresponding to λ becoming collinear with the adjacent external lines of the cut. Careful readers might find this puzzling, as the MHV tree amplitudes on both side of the cut only have collinear poles of the form $\langle l_1,i\rangle$, $\langle l_1,i+1\rangle$, $\langle l_2,j\rangle$ and $\langle l_2,j+1\rangle$. Recalling that here $|l\rangle = |l_2\rangle$, one would instead expect collinear poles of the form, $[l|P_{i+1,j}|i\rangle$, $[l|P_{i+1,j}|i+1\rangle$, $\langle\lambda,j\rangle$ and $\langle\lambda,j+1\rangle$. The resolution is that

Eq. (2.28) is obtained by shifting $\langle l_1, i\rangle \to \langle l_1, i\rangle + z\langle l_2, i\rangle$ and expanding around $z \to \infty$, thus introducing the $\langle l_2, i\rangle$ poles:

$$\frac{1}{\langle l_1(z), i\rangle}\bigg|_{z\to\infty} = \frac{1}{z\langle l_2, i\rangle} + \mathcal{O}\left(\frac{1}{z^2}\right). \tag{2.30}$$

Since these poles originated from $\langle l_1(z), i\rangle$, we will abuse the terminology, as well as the figures, and still refer to them as collinear poles.[7]

To better track the contributions of the collinear poles, we rewrite the integrand as follows:

$$\begin{aligned}\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}\bigg|_{\mathcal{O}(z^0)} &= (\mathcal{N}-4)\mathcal{A}_n^{\text{tree}}\frac{\langle a,\lambda\rangle\langle b,\lambda\rangle^2\langle i,i+1\rangle}{\langle a,b\rangle^2\langle i,\lambda\rangle\langle i+1,\lambda\rangle}\left(\frac{\langle a,j+1\rangle}{\langle j+1,\lambda\rangle} - \frac{\langle a,j\rangle}{\langle j,\lambda\rangle}\right)\\ &= (\mathcal{N}-4)\mathcal{A}_n^{\text{tree}}\frac{\langle a,\lambda\rangle\langle b,\lambda\rangle^2\langle j,j+1\rangle}{\langle a,b\rangle^2\langle j,\lambda\rangle\langle j+1,\lambda\rangle}\left(\frac{\langle a,i+1\rangle}{\langle i+1,\lambda\rangle} - \frac{\langle a,i\rangle}{\langle i,\lambda\rangle}\right),\end{aligned} \tag{2.31}$$

where the two equivalent representations focus on different adjacent collinear poles in the parentheses. The representation in Eq. (2.31) allows us to compute the bubble coefficient in a manner that manifests the relation between collinear poles in adjacent channels. With auxiliary spinor $|\alpha\rangle$ in Eq. (2.29) chosen to be $|a\rangle$, the bubble coefficient is

$$C^{(i,j)}_{2\,\{a,b\}} = C^{(i,j)}_{2\,\{a,b\}}(l\sim j+1) \;+\; C^{(i,j)}_{2\,\{a,b\}}(l\sim j) \;+\; C^{(i,j)}_{2\,\{a,b\}}(l\sim i+1) \;+\; C^{(i,j)}_{2\,\{a,b\}}(l\sim i)\,.$$

Here we have used $(l\sim j)$ to indicate the contribution from the collinear pole $\langle\lambda, j\rangle$. For convenience, we will refer to $(l\sim j)$ and $(l\sim j+1)$ collinear poles as "j-channel poles", and $(l\sim i)$ and $(l\sim i+1)$ poles as "i-channel poles". A graphical illustration of Eq. (2.33) is given in Fig. 2.10

Before proceeding, we point out a very important observation. Comparing the first line of Eq. (2.31) for $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}|_{\mathcal{O}(z^0)}$ with that for $\widehat{\mathcal{S}}^{(i,j-1)}_{\{a,b\},\mathcal{N}}|_{\mathcal{O}(z^0)}$,

$$\widehat{\mathcal{S}}^{(i,j-1)}_{a,b}\bigg|_{\mathcal{O}(z^0)} = (\mathcal{N}-4)\mathcal{A}_n^{\text{tree}}\frac{\langle a,\lambda\rangle\langle b,\lambda\rangle^2\langle i,i+1\rangle}{\langle a,b\rangle^2\langle i,\lambda\rangle\langle i+1,\lambda\rangle}\left(\frac{\langle a,j\rangle}{\langle j,\lambda\rangle} - \frac{\langle a,j-1\rangle}{\langle j-1,\lambda\rangle}\right),$$

we immediately see that *terms containing the common collinear pole of the two adjacent cuts, i.e.* $1/\langle j,\lambda\rangle$*, are exactly the same but, crucially, have opposite signs.* This applies to all of the other terms in Eq. (2.31): each residue in the sum having a counterpart in the adjacent channel, as illustrated in Fig. 2.1.

[7] In fact, this is not as much of an abuse as it may seem. Note that evaluating the pole at $z\to\infty$ is equivalent to evaluating the pole at the origin minus the poles at finite z. The former would be a true collinear pole, while the latter would be a collinear pole with shifted l_1.

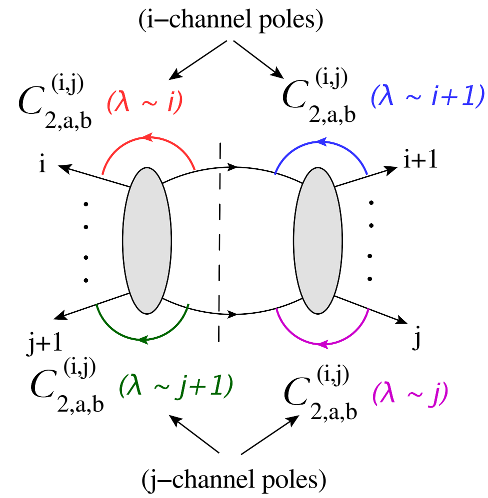


Figure 2.10: A graphical representation of Eq. (2.33). The bubble coefficient of a given channel is separated into four terms, each having a different collinear pole as the origin of the holomorphic anomaly that gives a non-zero dLIPS integral. The 4-contributions can be grouped into two channels, the i- and the j-channel.

At this point, one is tempted to conclude that the contributions to the bubble coefficient from common collinear channels cancel. However there is one subtlety. In Eq. (2.29), besides $\widehat{\mathcal{S}}^{(i,j)}_{\{a,b\},\mathcal{N}}|_{\mathcal{O}(z^0)}$, there is an extra factor in the total derivative that depends on the total momentum of the two-particle cut, $P_{i+1,j}$, which will be distinct for the adjacent cuts. Luckily these distinct factors become identical on the common collinear pole:

$$\frac{[\tilde{l}|P_{i+1,j}|a\rangle}{\langle l|P_{i+1,j}|\tilde{l}]\langle\lambda,a\rangle}\Big|_{\langle\lambda,j\rangle=[\tilde{l},j]=0}=\frac{[\tilde{l}|P_{i+1,j-1}|a\rangle}{\langle l|P_{i+1,j-1}|\tilde{l}]\langle\lambda,a\rangle}\Big|_{\langle\lambda,j\rangle=[\tilde{l},j]=0}, \tag{2.32}$$

where $|_{\langle\lambda,j\rangle=[\tilde{l},j]=0}$ indicates that the loop momentum is evaluated in the limit where it is collinear with j.[8] Because the extra factors are identical on the common collinear pole (CCP), we now conclude that the contribution of the CCP to the bubble coefficient indeed cancels between adjacent channels. This can also be concretely checked against the result from the direct evaluation of the dLIPS integral: [9]

$$\begin{aligned}
C^{(i,j-1)}_{2\,\{a,b\}}(l\sim j) &= -(\mathcal{N}-4)\mathcal{A}^{\text{tree}}_n\frac{\langle i,i+1\rangle}{\langle a,b\rangle^2}\frac{\langle a|P_{i+1,j-1}|j]}{\langle j+1|P_{i+1,j-1}|j]}\frac{\langle a,j\rangle\langle b,j\rangle^2}{\langle i,j\rangle\langle i+1,j\rangle},\\
C^{(i,j)}_{2\,\{a,b\}}(l\sim j) &= (\mathcal{N}-4)\mathcal{A}^{\text{tree}}_n\frac{\langle i,i+1\rangle}{\langle a,b\rangle^2}\frac{\langle a|P_{i+1,j}|j]}{\langle j|P_{i+1,j}|j]}\frac{\langle a,j\rangle\langle b,j\rangle^2}{\langle i,j\rangle\langle i+1,j\rangle}.
\end{aligned}$$

[8]Since the contour of the dLIPS integral is taken to be real, $\tilde{\lambda}=\bar{\lambda}$, the collinear pole $1/\langle\lambda,j\rangle$ freezes the loop momenta to satisfy $\langle\lambda,j\rangle=[\tilde{l},j]=0$.

[9]Explicit evaluation of this integral, via the holomorphic anomaly [21, 16, 17], is reviewed in section 2.3.

Adding these two equations, we find

$$
\begin{aligned}
&C_{2\,\{a,b\}}^{(i,j)}(l \sim j) + C_{2\,\{a,b\}}^{(i,j-1)}(l \sim j) \\
&= (4-\mathcal{N})\mathcal{A}_n^{\text{tree}} \frac{\langle i, i+1\rangle}{\langle a,b\rangle^2} \frac{\langle a,j\rangle\langle b,j\rangle^2}{\langle i,j\rangle\langle i+1,j\rangle}\left[-\frac{\langle a|P_{i+1,j}|j]}{\langle j|P_{i+1,j}|j]} + \frac{\langle a|P_{i+1,j-1}|j]}{\langle j|P_{i+1,j-1}|j]}\right] = 0\,, \qquad (2.33)
\end{aligned}
$$

thus verifying our claim.

Since the four collinear poles for a generic two-particle cut are shared by four different adjacent channels as shown in Fig. 2.1, this immediately leads to the result that although the bubble coefficient for a generic two-particle cut is given by complicated rational functions, as shown in Eq. (2.33), *in summing over all two-particle cuts there is a pairwise cancellation of CCP, and thus a majority of bubble coefficients do not contribute to the final result.*

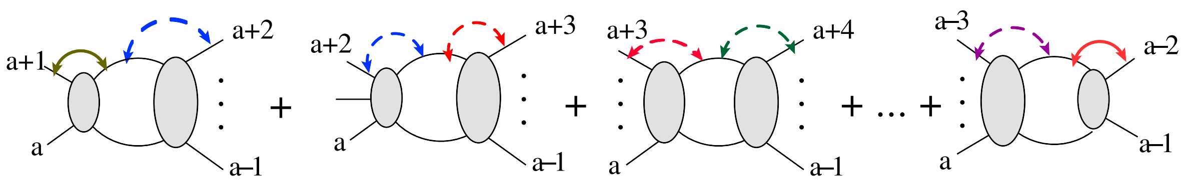


Figure 2.11: A schematic representation of the cancellation of CCP for adjacent MHV amplitude for SYM. Each colored arrow represents a collinear pole that contributes to the bubble coefficient. Pairs of dashed arrows in the same color cancel. Only those represented by the solid arrows one on the two ends remain; they are the only non-trivial contribution to the overall bubble coefficient.

The cancellation of CCP in adjacent channels leads to systematic cancellation in the sum of bubble coefficients, and the sum telescopes. However, there are "terminal cuts" which contain unique poles that are not cancelled. Below, we demonstrate that these so-called "terminal poles" constitute the sole contribution to the overall bubble coefficient. First we focus on the simplest case, namely the two external "negative helicity" Ψ-lines a, b are adjacent. The general case is treated below, first for SYM in this section, and then for YM in section 2.7.

Adjacent MHV amplitudes

We consider split-helicity MHV amplitudes where the Ψ lines a, b are adjacent, i.e. $b = a - 1$. The systematic cancellation is illustrated in Fig. 2.11, where the dashed lines indicate pairs of CCP that cancel in the sum. Note that there are no contributions from the collinear poles where the loop leg is collinear with the Ψ-lines, a and $a-1$. This is because the residues of such poles are zero, as can be seen in Eq. (2.31) and explicitly checked in Eq. (2.33). One immediately sees that the summation is reduced to the two terminal poles. These are identified as poles in two-particle cuts with a 4-point tree amplitude on one side (and an n-point tree on the other), where the two loop momenta become collinear with the two external legs of the 4-point tree amplitude. A straightforward evaluation of the contribution of these two terminal poles yields the result for the sum of all bubble coefficients:

$$\begin{aligned} \mathcal{C}_{2\,\{a,a-1\}} &= C^{(a-3,a-1)}_{2\,\{a,a-1\}}(l\sim a-2) + C^{(a+1,a-1)}_{2\,\{a,a-1\}}(l\sim a+1) \\ &= -(\mathcal{N}-4)\mathcal{A}^{\text{tree}}_n + 0 = -\beta_0 \mathcal{A}^{\text{tree}}_n\,, \end{aligned} \tag{2.34}$$

with $\beta_0 = (\mathcal{N}-4)$. Note that $C^{(a+1,a-1)}_{2\,\{a,a-1\}}(l\sim a+1) = 0$ is a result of our choice of reference spinor $|\alpha\rangle = |a\rangle$ in deriving Eq. (2.33). Were we to make the other natural choice, $|\alpha\rangle = |b\rangle = |a-1\rangle$, we would instead have $C^{(a-3,a-1)}_{2\,\{a,a-1\}}(l\sim a-2) = 0$ and $C^{(a+1,a-1)}_{2\,\{a,a-1\}}(l\sim a+1) = -\beta_0\mathcal{A}^{\text{tree}}_n$.

For example take the six-point MHV amplitude with legs 1 and 6 to be negative-helicity lines. The sum of bubble coefficients is given as

$$\begin{aligned} \mathcal{C}_{2\,\{1,6\}} &= C^{(2,6)}_{2\,\{1,6\}}(l\sim 2) + C^{(2,6)}_{2\,\{1,6\}}(l\sim 3) + C^{(3,6)}_{2\,\{1,6\}}(l\sim 3) + C^{(3,6)}_{2\,\{1,6\}}(l\sim 4) \\ &\quad + C^{(4,6)}_{2\,\{1,6\}}(l\sim 4) + C^{(4,6)}_{2\,\{1,6\}}(l\sim 5) \\ &= C^{(2,6)}_{2\,\{1,6\}}(l\sim 2) + C^{(4,6)}_{2\,\{1,6\}}(l\sim 5) = -\beta_0\mathcal{A}^{\text{tree}}_n\,. \end{aligned} \tag{2.35}$$

We see that there are two pairs of common collinear poles, $l\sim 3$ and $l\sim 4$. The pairs cancel each other in the sum and one arrives at the two terminal poles which evaluate to the desired result. The cancellation is illustrated in Fig. 2.12.

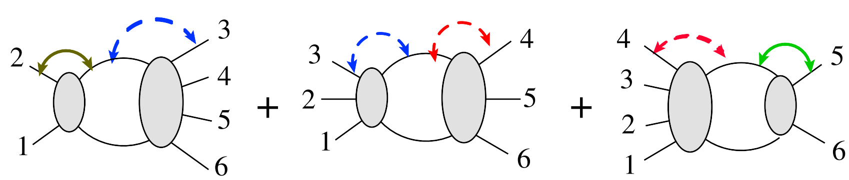


Figure 2.12: A schematic representation of the cancellation of CCP for adjacent six-point MHV amplitude. The dashed lines are common collinear poles, which cancel pairwise.

Thus we have demonstrated that one of the terminal poles vanishes and *the sum of bubble coefficients for adjacent MHV amplitudes, for arbitrary n, are given by a single terminal pole!*

Non-adjacent MHV amplitudes

The above case with the two Ψ-lines a, b being adjacent is simple because the j-channel poles (see below (2.31)) were absent. For MHV amplitudes with a, b being non-adjacent, the j-channel poles are now non-zero, and all the four collinear poles contribute in Eq. (2.33). The sum of bubble coefficients can be conveniently separated into a summation of the i-channel poles, and a summation of the j-channel poles. Cancellation of CCP in both channels again reduces the summation to the terminal poles. For simplicity, we set $a = 1$ and $1 < b$. We denote the terminal cut in the summation of i-channel poles by i_t, j, and similarly for the terminal cut of the j-channel poles by i, j_t. The contribution of these uncancelled poles are identified as:

- i-channel:

$$C_{2\ \{1,b\}}^{(i_t,j)}(\lambda \sim i_t) \qquad \begin{array}{ll} \text{for } \ j = n, & i_t = 2 \\ \text{for } \ b \le j < n, & i_t = 1 \end{array} \quad ,$$

$$C_{2\ \{1,b\}}^{(i_t,j)}(\lambda \sim i_t + 1) \qquad \begin{array}{ll} \text{for } \ j = b, & i_t = b-2 \\ \text{for } \ b < j \le n, & i_t = b-1 \end{array} \quad ,$$

- j-channel:

$$C_{2\ \{1,b\}}^{(i,j_t)}(\lambda \sim j_t) \qquad \begin{array}{ll} \text{for } \ i = b-1, & j_t = b+1 \\ \text{for } \ 1 \le i < b-1, & j_t = b \end{array} \quad ,$$

$$C_{2\ \{1,b\}}^{(i,j_t)}(\lambda \sim j_t + 1) \qquad \begin{array}{ll} \text{for } \ i = 1, & i_t = n-1 \\ \text{for } \ 1 < i \le n-1, & j_t = n \end{array} \quad .$$

In identifying the terminal poles, one has to take into account that, when summing over the i-channel poles, the value of j affects the possible values that i can take (and vice versa for the summation of j-channel poles). For a detailed discussion of the above result, we refer the reader to section 2.7 where we perform a similar analysis for non-adjacent MHV amplitudes in pure Yang-Mills. As discussed in section 2.4.2, the collinear poles where the loop momenta becomes collinear with a Ψ-line have vanishing residues. In the present context, this refers to $(\lambda \sim 1)$ and $(\lambda \sim b)$. Thus there are only four contributing terms in the sum of bubble coefficients

$$\mathcal{C}_{2\,\{1,b\}} = C_{2\ \{1,b\}}^{(2,n)}(\lambda \sim 2) + C_{2\ \{1,b\}}^{(b-2,b)}(\lambda \sim b-1) + C_{2\ \{1,b\}}^{(b-1,b+1)}(\lambda \sim b+1) + C_{2\ \{1,b\}}^{(1,n-1)}(\lambda \sim n)\,. \quad (2.36)$$

Extracting the corresponding expressions from Eq. (2.33), one finds that the first and last terms vanish. This is again due to the choice of reference spinor $|\alpha\rangle = |a\rangle$.[10] Thus the only contributions

[10] If we were to use the other natural choice, $|\alpha\rangle = |b\rangle$, we would arrive at the result that the second and third terms of Eq. (2.36) vanish. This apparent dependence of a particular double-cut on the reference spinor is illusory: with care, one can cancel the full $|\alpha\rangle$-dependence from each individual bubble coefficient. However, this cancellation comes at the expense of the manifest $a \leftrightarrow b$ symmetry present in the uncancelled form. This asymmetry causes one term to seemingly vanish while the other gives the full bubble-coefficient.

to the sum of bubble coefficients come from $C^{(b-1,b+1)}_{2\ \{1,b\}}(l \sim b+1)$ and $C^{(b-2,b)}_{2\ \{1,b\}}(l \sim b-1)$, which sum to

$$\begin{aligned} \mathcal{C}_{2\,\{1,b\}} &= C^{(b-1,b+1)}_{2\ \{1,b\}}(l \sim b+1) + C^{(b-2,b)}_{2\ \{1,b\}}(l \sim b-1) \\ &= (4-\mathcal{N})\mathcal{A}^{\mathrm{tree}}_n \frac{\langle 1, b-1\rangle\langle b, b+1\rangle + \langle b-1, b\rangle\langle 1, b+1\rangle}{\langle b-1, b+1\rangle\langle 1, b\rangle} = -(\mathcal{N}-4)\mathcal{A}^{\mathrm{tree}}_n\,. \end{aligned} \tag{2.37}$$

This agrees with Eq. (2.4) with $\beta_0 = (\mathcal{N}-4)$.

In conclusion, for both adjacent and non-adjacent MHV amplitudes in $\mathcal{N} = 1, 2$ super Yang-Mills theory, the cancellation of CCP in the sum of bubble coefficients implies that *for n-point (non-)adjacent MHV amplitudes, only (two) one term in the sum of bubble coefficients gives a non-trivial contribution $\beta_0\mathcal{A}^{\mathrm{tree}}_n$.* Thus the on-shell formalism achieves Eq. (2.4) in a systematic and simple way.

2.4.3 MHV bubble coefficients for pure Yang-Mills

The observed structure of cancellations for $\mathcal{N} = 1, 2$ super Yang-Mills theory is present in pure Yang-Mills as well. However, it is more involved to derive this since the $\mathcal{O}(z^0)$ part of the BCFW-shifted two-particle cut contains higher-order collinear poles. Nevertheless, adjacent channels again share these higher-order CCP, and their contribution to the sum of bubble coefficient also cancels. The cancellation of CCP renders the summation down to the terminal poles, which evaluate to $(11/6)A^{\mathrm{tree}}_n$. We present the detailed derivation of this in section 2.7. Here we would like to give a brief discussion on the nature of the terminal poles in pure Yang-Mills theory.

As discussed above, the terminal cuts are those where there is a 4-point tree amplitude on one side of the two-particle cut. The uncancelled terminal poles can be identified as the poles that arise when the loop momenta become collinear with the pair of external legs of this 4-point tree amplitude. For pure Yang-Mills, summing over the internal helicity configurations *before* taking the dz and dLIPS integrals obscures the nature of the cancellation.

Additional structure reveals itself if we *first* evaluate the contributions to the bubble-coefficient for a given set of internal states, aka gluon helicity configuration, and *then* sum over internal states/helicities. Specifically, these double-forward terminal poles are non-zero only when the internal helicities of the loop legs leaving the n-point tree on one side of the cut, match with the helicities of the pair of external lines in the 4-point tree on the other side of the cut (see Fig. 2.13). These "helicity preserving" double-poles (which will henceforth be called "double-forward poles") give the entire bubble coefficient.

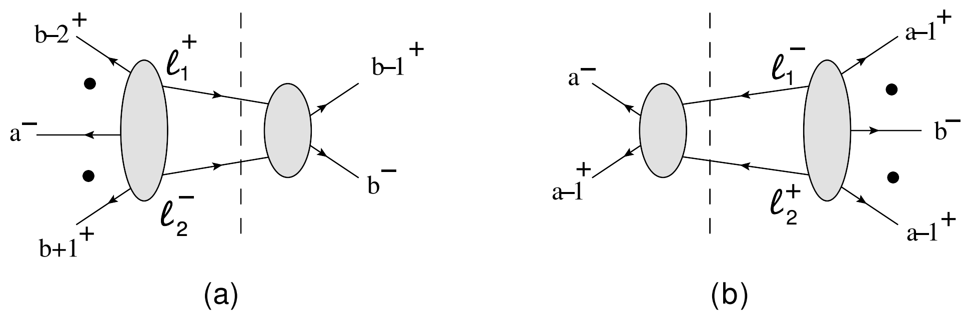


Figure 2.13: The two terminal cuts for a given helicity configuration for the loop legs. For the choice of reference spinor $|\alpha\rangle = |a\rangle$ only diagram (a) is non vanishing. If one instead choose $|\alpha\rangle = |b\rangle$, then it is diagram (b) that gives the non-trivial contribution.

Consider the internal helicity configuration (l_1^+, l_2^-) as shown in Figs. 2.4.1 and 2.13. There are two "helicity preserving" terminal-cuts: diagram (a) and (b) in Fig.2.13. Choosing the reference spinor $|\alpha\rangle = |a\rangle$, diagram (b) vanishes, and diagram (a) evaluates to $11/6A_n^{\rm tree}$, see Eq. (2.76). If one were to make the other choice for the reference spinor, $|\alpha\rangle = |b\rangle$, we would instead have diagram (a) in Fig 2.13 vanishing, and diagram (b) giving $11/6A_n^{\rm tree}$. In fact, the helicity preserving property of the contributing poles can also be seen for the $\mathcal{N} = 1, 2$ super Yang-Mills theory, where one simply substitute the $+$ and $-$ helicity in the previous discussions with Ψ and Φ lines. This fact was obscured previously as the different internal multiplet configurations were summed to obtain the simple form of the two-particle cut in Eq. (2.28).

Thus we conclude that in the pure Yang-Mills theory *the sum of bubble coefficients is simply given by the contribution of terminal poles where the helicity configuration is preserved, and where the loop momenta become collinear with the pair of external legs within the 4-point amplitude.* For simplicity we will call these double-forward poles, due to the nature of the kinematics. In section 2.6 we will show that for split-helicity N^{k}MHV bubble coefficients, these double-forward poles again produce the correct sum for the bubble coefficient, thus indicating complete cancellation among the remaining contributions. But before indulging in that story let us present a general argument for the cancellation of CCP.

2.5 Towards general cancellation of common collinear poles

In the above, we have shown that Eq. (2.4) can be largely attributed to the fact that the bubble coefficient for a given cut, secretly shares the same terms with its four adjacent cuts, leading to systematic cancellations between them. We have proven this for n-point MHV amplitudes in both

$\mathcal{N} = 1, 2$ super Yang-Mills as well as pure Yang-Mills theories (section 2.7). One can also consider adjoint scalars and fermions minimally coupled to gluons. Since at one-loop, we can separate contributions from different spins inside the loop as

$$\begin{aligned} \text{fermions} \quad &\rightarrow \quad \big(\, \mathcal{N}=1 \text{ SYM} \,\big) \;-\; \big(\, \text{YM} \,\big) \\ \text{scalar} \quad &\rightarrow \quad \big(\, \mathcal{N}=2 \text{ SYM} \,\big) \;-\; \big(\, \mathcal{N}=1 \text{ SYM} \,\big) \;-\; \big(\, \text{fermions} \,\big)\,, \end{aligned} \tag{2.38}$$

proof of cancellation of CCP for each of the theories (for MHV scattering) on the RHS of Eq. (2.38), implies that such cancellation occurs for each spin individually.

We would like to show this holds for N^{k}MHV amplitudes. Unfortunately, for N^{k}MHV amplitudes, multi-particle poles of tree amplitudes on either side of the cut contribute to the bubble coefficient, and the analysis becomes more complicated: cancellation of CCP is no longer sufficient to show terminal poles dominate the bubble coefficient. Nonetheless, we believe that the cancellation between CCP persists for arbitrary helicity configuration. As an indication, we demonstrate that the residues of CCP for adjacent cut always have the same form and opposite signs, for any helicity configuration.

Collinear limits of tree level amplitudes in Yang-Mills theory, with $k_a = zk_P,\ k_b = (1-z)k_P$, factorize as

$$A_n^{\text{tree}}\left(..., a^{\lambda_a}, b^{\lambda_b}, ...\right) \rightarrow \sum_{\lambda=\pm} \text{Split}_{-\lambda}^{\text{tree}}\left(z, a^{\lambda_a}, b^{\lambda_b}\right) A_{n-1}^{\text{tree}}\left(..., P^{\lambda}, ...\right) \tag{2.39}$$

where the factor $\text{Split}_{-\lambda}^{\text{tree}}\left(z, a^{\lambda_a}, b^{\lambda_b}\right)$ is the gluon splitting amplitude. Its form for various helicity configurations are given by [23, 27]:

$$\begin{aligned} \text{Split}_{-}^{\text{tree}}\left(a^-, b^-\right) &= 0 \\ \text{Split}_{-}^{\text{tree}}\left(a^+, b^+\right) &= \frac{1}{\sqrt{z(1-z)}\langle ab\rangle} \\ \text{Split}_{+}^{\text{tree}}\left(a^+, b^-\right) &= \frac{(1-z)^2}{\sqrt{z(1-z)}\langle ab\rangle} \\ \text{Split}_{-}^{\text{tree}}\left(a^+, b^-\right) &= -\frac{z^2}{\sqrt{z(1-z)}[ab]}\,. \end{aligned} \tag{2.40}$$

Without loss of generality, we focus on the common collinear pole, depicted in Fig. 2.5, in adjacent cuts $(1...i-1|i...n)$ and $(1...i|i+1...n)$. In other words, we study the collinear region with $l_1^{(i)} = \tau^{(i)} k_i$

and $l_1^{(i+1)} = \tau^{(i+1)} k_i$.[11] The two integrands become

$$\begin{aligned}
&\text{Cut}_{(1..i-1|i..n)}|_{\langle l_1 i\rangle} \qquad (2.41)\\
&= A_{i+1}\left(1,...,i-1,\tau^{(i)}i,l_2^{(i)}\right)\sum_{\lambda=\pm}\text{Split}^{\text{tree}}_{-\lambda}A_{n-i+2}\left(-l_2^{(i)},(1-\tau^{(i)})i,i+1,...,n\right)\ ,
\end{aligned}$$

for cut $(1...i-1|i...n)$, and

$$\begin{aligned}
&\text{Cut}_{(1..i|i+1...n)}|_{\langle l_1 i\rangle} \qquad (2.42)\\
&= \sum_{\lambda=\pm}\text{Split}^{\text{tree}}_{-\lambda}A_{i+1}\left(1,...,(1+\tau^{(i+1)})i,l_2^{(i+1)}\right)A_{n-i+2}\left(-l_2^{(i+1)},-\tau^{(i+1)}i,i+1,...,n\right)\ ,
\end{aligned}$$

for cut $(1...i|i+1...n)$. The parameter $\tau^{(i)}$ can be fixed by the on-shell condition on $l_2^{(i)}$ since in the cut $(1...i-1|i...n)$, $l_2^{(i)} = P_{i-1} + \tau^{(i)} k_i$. Similar constraints from the cut $(1...i|i+1...n)$ fix $\tau^{(i+1)}$. This leads to

$$\begin{aligned}
\tau^{(i)} &= \frac{P_{i-1}^2}{2k_i\cdot P_{i-1}} = \tau^{(i+1)} + 1\\
\to\quad l_2^{(i)} &= P_{i-1} + \tau^{(i)}k_i = P_i + \tau^{(i+1)}k_i = l_2^{(i+1)}\ .
\end{aligned}$$

Substituting these results back into Eq. (2.41) and Eq. (2.42), we see that the product of tree amplitudes are identical at their common collinear pole. Furthermore, identifying the kinematic variables in the splitting amplitudes for each cut as:

$$\begin{aligned}
(1...i-1|i...n):&\qquad k_a = k_i,\ k_b = -\tau^{(i)}k_i,\ z = \frac{1}{1-\tau^{(i)}}\\
(1...i|i+1...n):&\qquad k'_a = k_i,\ k'_b = \tau^{(i+1)}k_i,\ z' = \frac{1}{1+\tau^{(i+1)}} = \frac{1}{\tau^{(i)}}
\end{aligned}$$

we see that the splitting amplitudes for the two cuts are identical with a relative minus sign.[12]

The above analysis confirms that Eq. (2.41) and Eq. (2.42) are indeed identical up to a minus sign. Thus the residue of the *entire two-particle cut* on the common collinear poles, are identical and with opposite sign. This, however, does not directly lead to a proof of cancellation of CCP for bubble coefficients. This is because to extract the bubble coefficient, the two-particle cut must be translated into a total derivative, in order for one to use holomorphic anomaly generated by the collinear poles to isolate the dLIPS integral. It is not guaranteed that after translating the two cuts

[11]Strictly speaking, the condition $\langle l_1 i\rangle = 0$ only requires $\lambda_{l_1} \sim \lambda_i$. However since the dLIPS integration contour is along $\tilde\lambda = \bar\lambda$, the condition is equivalent to $l_1 \sim k_i$.

[12]For consistency, we take the positive branch of the square root.

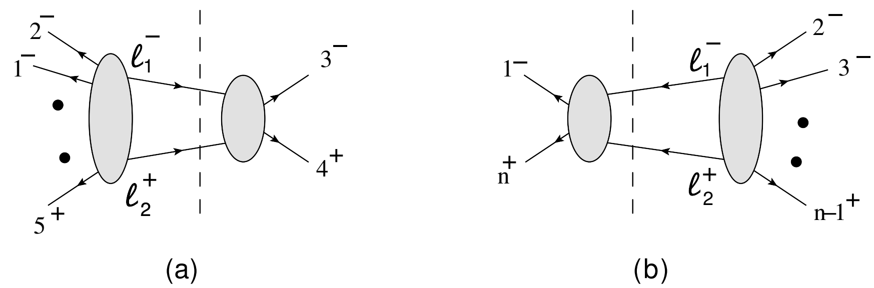


Figure 2.14: The terminal cuts of the split helicity NMHV amplitude that contain the two helicity-preserving double-forward poles.

into a total derivative form, the residues on the CCP are still equal and opposite.

2.6 N^{k}MHV bubble coefficients

The cancellation of CCP, even if it holds for generic helicity configurations, is clearly not sufficient for simplifying the sum of bubble coefficients for N^{k}MHV amplitudes. The complications arise from the presence of multi-particle poles of the tree amplitudes in the two-particle cut. It is conceivable that there exists a similar cancellation of common multi-particle poles, since a trivial example would be the cancellation of CCP, considering the fact that collinear poles are secretly multi-particle poles via momentum conservation. Here we instead ask a more direct question: does the contribution of the double forward poles to the bubble coefficient, directly give $(11/6)A^{\text{tree}}$ for n-point N^{k}MHV amplitude.

To facilitate our analysis, we will use the CSW representation [21, 28, 29] for the split helicity NMHV tree amplitude $(- - - + \cdots +)$.[13] We will show that at the double forward poles, the contribution from each individual CSW diagram evaluates to 11/6 times the original CSW diagram whose loop momenta is replaced by the corresponding external lines. Summing the different diagrams one simply recovers 11/6 times the CSW representation of the tree amplitude. Using induction, we prove that this is still true for all n-point split-helicity N^{k}MHV amplitudes.

2.6.1 Double forward poles in terminal cuts of $A_n^{1-\text{loop}}(- - - + \ldots +)$

The split helicity configuration for NMHV amplitudes is the simplest to analyze. The CSW form for n-particle NMHV scattering, with adjacent negative-helicity gluons is given by the following $2(n-3)$ terms [21]:

[13]NMHV is an abbreviation for "next to maximal helicity violating". These are the next-to-simplest tree amplitudes in (S)YM.

$$A(1^-,2^-,3^-,4^+,...,n^+) = \sum_{i=4}^{n} \frac{\langle 1,2\rangle^3}{\langle P_{3,i},i+1\rangle\cdots\langle P,2\rangle}\frac{1}{P_{3,i}^2}\frac{\langle P_{3,i},3\rangle^3}{\langle 3,4\rangle\cdots\langle i,P_{3,i}\rangle} + \sum_{i=4}^{n}\frac{\langle 1,P_{2,i-1}\rangle^3}{\langle P_{2,i-1},i\rangle\cdots\langle n,1\rangle}\frac{1}{P_{2,i-1}^2}\frac{\langle 2,3\rangle^3}{\langle P_{2,i-1},2\rangle\cdots\langle i-1,P_{2,i-1}\rangle} \tag{2.43}$$

where $|P_{i,j}\rangle \equiv P_{i,j}|\tilde{\eta}]$, and $\tilde{\eta}$ is an auxiliary spinor.

For helicity configuration (l_1^-, l_2^+), the terminal cuts are shown in Fig. 2.14. We first focus on cut (a), which is given by $A_4(\hat{l}_1^+,3^-,4^+,\hat{l}_2^-)\times A_n(\hat{l}_2^+,5^+,...,n^+,...,1^-,2^-,\hat{l}_1^-)$. Cut (b) in Fig. 2.14 evaluates in exactly the same way. We use the CSW expansion on the n-particle NMHV subamplitude as indicated in Fig. 2.15:

$$[\ \mathrm{cut}(a)\ ]\Big|_{df} = [\ \mathrm{diag}(a)+\mathrm{diag}(b)+\mathrm{diag}(c)\ ]\Big|_{df} \tag{2.44}$$

Notice for diagram (a) and (b) of Fig. 2.15, the loop legs are on the same MHV vertex and hence the CSW propagator $1/P^2$ does not depend on z. This implies that from the point of view of extracting the pole at $z\to\infty$ and performing the dLIPS integration, only the MHV vertex on which the loop legs are attached are relevant. In other words, the CSW propagator attached to the n-point amplitude on the LH-side of the cut, does not modify the dLIPS integrand on the pole at large-z. Hence evaluation of the double forward poles on diagrams (a) and (b) of Fig. 2.15 simply returns ($11/6\times$) those original CSW tree-diagrams/amplitudes. This realization makes the computation trivial, as we know the double forward pole contributes 11/6 times the tree amplitude. This implies that here the result would simply be 11/6 times the corresponding CSW tree diagram. One can thus straightforwardly obtain:

$$[\ \mathrm{diag}(a)+\mathrm{diag}(b)\ ]\Big|_{df} = \frac{11}{6}\Bigg(\sum_{i=4}^{n} \frac{\langle 1,2\rangle^3}{\langle P_{3,i},i+1\rangle\cdots\langle P,2\rangle}\frac{1}{P_{3,i}^2}\frac{\langle P_{3,i},3\rangle^3}{\langle 3,4\rangle\cdots\langle i,P_{3,i}\rangle} + \sum_{i=5}^{n}\frac{\langle 1,P_{2,i-1}\rangle^3}{\langle P_{2,i-1},i\rangle\cdots\langle n,1\rangle}\frac{1}{P_{2,i-1}^2}\frac{\langle 2,3\rangle^3}{\langle P_{2,i-1},2\rangle\cdots\langle i-1,P_{2,i-1}\rangle}\Bigg), \tag{2.45}$$

where $|_{df}$ indicates the contribution from the double forward pole. There is another type of contribution, as shown in diagram (c) of Fig. 2.15, where the CSW propagator depends on z and we need a careful analysis as following.

Denoting $|\hat{P}\rangle = \hat{P}|\tilde{\eta}]$, which accounts for the z-dependence of the CSW propagator, the cut

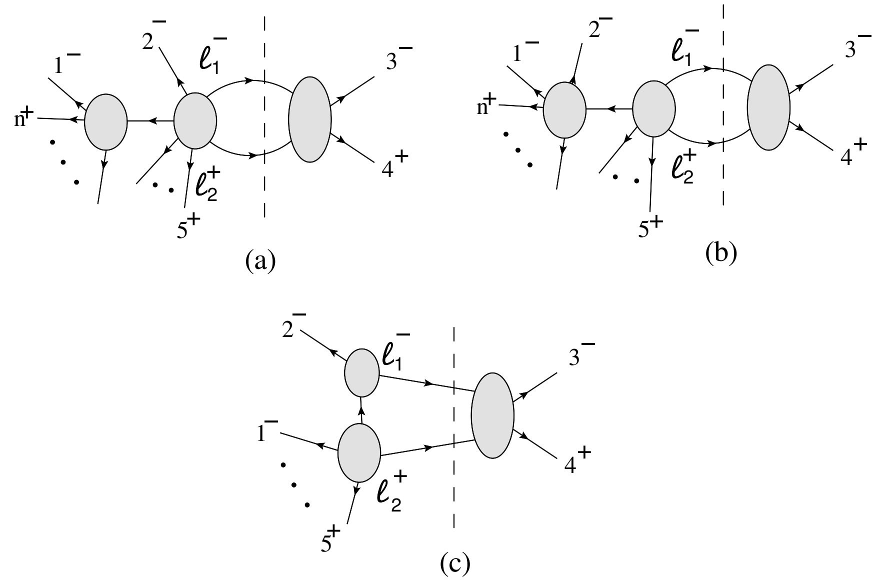


Figure 2.15: Representation of the CSW expansion of the tree amplitude within the terminal cuts. Note that the loop legs are attached to the same MHV vertex for diagram (a) and (b). The dLIPS integration for these diagrams are exactly the same as that computed for adjacent MHV bubbles. Diagram (c), the "hard" term, is more subtle.

integrand is given by:

$$\text{diag}(c) = \frac{\langle 1, \hat{P}\rangle^3}{\langle \hat{P}, l_2\rangle\langle \hat{l}_2, 5\rangle ... \langle n, 1\rangle} \frac{1}{\hat{P}^2} \frac{\langle 2, \hat{l}_1\rangle^3}{\langle \hat{P}, 2\rangle\langle \hat{l}_1, \hat{P}\rangle} \times \frac{\langle 3, \hat{l}_2\rangle^4}{\langle 3, 4\rangle\langle 4, l_2\rangle\langle l_2, \hat{l}_1\rangle\langle \hat{l}_1, 3\rangle} \tag{2.46}$$

where $\hat{P} = p_2 + \hat{l}_1$. We call this the "hard" term. The other $2n-5$ terms we call the "easy" terms. For simplicity, we first strip off the corresponding tree factor, i.e. the $i = 4$ term in the second line of Eq. (2.43),

$$T_{2,3} = \frac{\langle 1, P_{2,3}\rangle^3}{\langle P_{2,3}, 4\rangle\langle 4, 5\rangle ... \langle n, 1\rangle} \frac{1}{P_{2,3}^2} \frac{\langle 2, 3\rangle^3}{\langle P_{2,3}, 2\rangle\langle 3, P_{2,3}\rangle} \tag{2.47}$$

from the terminal dLIPS integrand in Eq. (2.46). This yields:

$$\text{diag}(c) = T_{2,3}\left(\frac{\langle 1, \hat{P}\rangle^3\langle P_{2,3}, 4\rangle\langle 4, 5\rangle}{\langle 1, P_{2,3}\rangle^3\langle \hat{P}, l_2\rangle\langle l_2, 5\rangle} \frac{P_{2,3}^2}{\hat{P}^2} \frac{\langle 2, \hat{l}_1\rangle^3}{\langle \hat{P}, 2\rangle\langle \hat{l}_1, \hat{P}\rangle} \frac{\langle P_{2,3}, 2\rangle\langle 3, P_{2,3}\rangle\langle 3, l_2\rangle^4}{\langle 2, 3\rangle^3\langle 3, 4\rangle\langle 4, l_2\rangle\langle l_2, \hat{l}_1\rangle\langle \hat{l}_1, 3\rangle}\right) \tag{2.48}$$

We would like to put this into a form where we can readily take the large-z pole followed by the dLIPS integral about the double forward pole. To simplify the analysis, we work out the explicit form of the spinor-inner products:

$$
\begin{aligned}
|\hat{P}\rangle = (p_2 + \hat{l}_1)|\tilde{\eta}] \Rightarrow & \Big\{ \langle \hat{l}_1, \hat{P}\rangle = \langle \hat{l}_1, 2\rangle[2,\tilde{\eta}], \qquad \langle l_2, \hat{P}\rangle = \langle l_2, 2\rangle[2,\tilde{\eta}] + \langle l_2, \hat{l}_1\rangle[l_1,\tilde{\eta}], \\
& \qquad \langle 2, \hat{P}\rangle = \langle 2, \hat{l}_1\rangle[l_1,\tilde{\eta}], \qquad \langle 1, \hat{P}\rangle = \langle 1, \hat{l}_1\rangle[l_1,\tilde{\eta}] + \langle 1, 2\rangle[2,\tilde{\eta}] \Big\} \\
|P_{2,3}\rangle = (p_2 + p_3)|\tilde{\eta}] \Rightarrow & \Big\{ \langle 3, P_{2,3}\rangle = \langle 3, 2\rangle[2,\tilde{\eta}], \qquad \langle l_2, P_{2,3}\rangle = \langle l_2, 2\rangle[2,\tilde{\eta}] + \langle l_2, 3\rangle[3,\tilde{\eta}], \\
& \qquad \langle 2, P_{2,3}\rangle = \langle 2, 3\rangle[3,\tilde{\eta}], \qquad \langle 1, P_{2,3}\rangle = \langle 1, 3\rangle[3,\tilde{\eta}] + \langle 1, 2\rangle[2,\tilde{\eta}] \Big\} .
\end{aligned}
\tag{2.49}
$$

Applying Eq. (2.49) to Eq. (2.48), cancelling all common factors, and setting $|\tilde{\eta}] = |2]$,[14] we find:

$$
\begin{aligned}
\text{diag}(c) &= T_{2,3} \frac{\langle 1\hat{l}_1\rangle^3 \langle 4,5\rangle \langle 3, l_2\rangle^4}{\langle 13\rangle^3 \langle \hat{l}_1 l_2\rangle^2 \langle l_2, 5\rangle \langle 4, l_2\rangle \langle \hat{l}_1, 3\rangle} \\
&= T_{2,3}\left[\left(\frac{\langle 3,4\rangle}{\langle l_2, 4\rangle} - \frac{\langle 3,5\rangle}{\langle l_2,5\rangle} \right) \frac{\langle \hat{l}_1, l_2\rangle}{\langle \hat{l}_1, 3\rangle} \left(-1 + \frac{\langle 1, l_2\rangle \langle \hat{l}_1, 3\rangle}{\langle 1,3\rangle \langle \hat{l}_1, l_2\rangle} \right)^3 \right] .
\end{aligned}
\tag{2.50}
$$

Expanding around $z \to \infty$, one finds

$$
\text{diag}(c)|_{z\to\infty} = T_{2,3}\left[\left(\frac{\langle 3,4\rangle}{\langle l_2,4\rangle} - \frac{\langle 3,5\rangle}{\langle l_2,5\rangle} \right) \left(3\frac{\langle 1, l_2\rangle}{\langle 1,3\rangle} - 3\frac{\langle 1,l_2\rangle^2 \langle l_1, 3\rangle}{\langle 1,3\rangle^2 \langle l_1, l_2\rangle} + \frac{\langle 1,l_2\rangle^3 \langle l_1,3\rangle^2}{\langle 1,3\rangle^3 \langle l_1,l_2\rangle^2} \right) \right] .
$$

The double forward pole corresponds to the $\langle l_2, 4\rangle$ pole. A straightforward evaluation of the residue gives:

$$
\begin{aligned}
\text{diag}(c)|_{df} &= T_{2,3} \int d\text{LIPS}\, \frac{\langle 3,4\rangle}{\langle l_2,4\rangle} \left(3\frac{\langle 1,l_2\rangle}{\langle 1,3\rangle} - 3\frac{\langle 1,l_2\rangle^2 \langle l_1,3\rangle}{\langle 1,3\rangle^2 \langle l_1,l_2\rangle} + \frac{\langle 1,l_2\rangle^3 \langle l_1,3\rangle^2}{\langle 1,3\rangle^3 \langle l_1,l_2\rangle^2} \right) \\
&= \left(3\times 1 - 3\times\frac{1}{2} + 1\times\frac{1}{3} \right) T_{2,3} = \frac{11}{6} T_{2,3} .
\end{aligned}
$$

[14] This ordering of steps is important: this cancels an apparent factor of $[2,\tilde{\eta}]$ in the denominator of Eq. (2.48).

In summary, we find that the double forward pole of the terminal cuts sum to give:

$$\begin{aligned}
[\,\text{diag}(a) + \quad \text{diag}(b) \quad + \text{diag}(c)\,]\Big|_{df} &= \frac{11}{6}\left(\sum_{i=4}^{n} \frac{\langle 1,2\rangle^3}{\langle P_{3,i}, i+1\rangle \cdots \langle P,2\rangle} \frac{1}{P_{3,i}^2} \frac{\langle P_{3,i},3\rangle^3}{\langle 3,4\rangle \cdots \langle i, P_{3,i}\rangle}\right) \\
+ \quad & \frac{11}{6}\left(\sum_{i=5}^{n} \frac{\langle 1, P_{2,i-1}\rangle^3}{\langle P_{2,i-1}, i\rangle \cdots \langle n,1\rangle} \frac{1}{P_{2,i-1}^2} \frac{\langle 2,3\rangle^3}{\langle P_{2,i-1},2\rangle \cdots \langle i-1, P_{2,i-1}\rangle}\right) \\
+ \quad & \frac{11}{6}\left(\frac{\langle 1, P_{2,3}\rangle^3}{\langle P_{2,3},4\rangle\langle 4,5\rangle ... \langle n,1\rangle} \frac{1}{P_{2,3}^2} \frac{\langle 2,3\rangle^3}{\langle P_{2,3},2\rangle\langle 3,P_{2,3}\rangle}\right) \\
= \quad & \frac{11}{6} A(1^-, 2^-, 3^-, 4^+, ..., n^+)
\end{aligned} \tag{2.51}$$

Thus indeed the double forward limit of cut (a), of Fig. 2.14, gives the expected proportionality factor from the sum of bubble coefficients, suggesting complete cancellation of contributions from all the other channels and poles.

We see that using the CSW representation for NMHV tree amplitudes reveals the following structure: the double forward poles in the terminal cut contributes a factor 11/6 for *each* CSW tree diagram. Note that all but one term, diagram (c), goes through trivially as the dLIPS only sees only one MHV vertex in the NMHV tree amplitude. There is a more straightforward way of understanding the factor of 11/6 in diagram (c), which goes as follows.

Consider the same calculation for the split-helicity NMHV five-point amplitude. Since this is secretly a five-point $\overline{\text{MHV}}$ amplitude we know that the forward poles for a given internal helicity configuration evaluate to $11/6\times$ $\overline{\text{MHV}}$, as proven previously. Now consider the same evaluation using the CSW representation. Since diagrams (a) and (b) automatically yield 11/6 times the corresponding CSW tree diagram, diagram (c) must give 11/6 times its corresponding CSW tree diagram, $T_{2,3}$.

Now, for higher point NMHV amplitudes, diagram (c) is modified by additional positive helicity legs on one of the MHV vertices, as indicated in Fig. 2.16. From the point of view of expanding around the pole at $z \to \infty$ and the dLIPS integral, these additional plus helicity legs are simply spectators and do not participate. Thus the evaluation of diagram (c) “must” be 11/6 $T_{2,3}$ as explicitly shown above. Note that this way of understanding the result of the double forward poles allows us to generalize to N^{k}MHV.

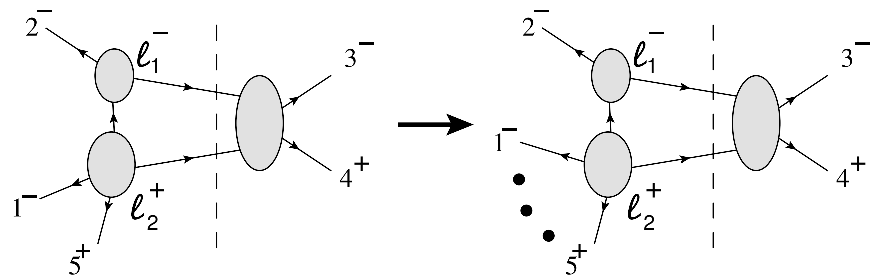


Figure 2.16: Diagram (c) in Fig. 2.15 for the five-point amplitude. Going from five-point to arbitrary n-point simply corresponds to adding additional plus helicity legs on the bottom MHV vertex. Since this modification affects neither extraction of constant term for $z \to \infty$ nor evaluation of the dLIPS integral, the 11/6 factor obtained at five-points holds for arbitrary n.

2.6.2 Recursive generalization to N^{k}MHV bubble coefficients

We are now ready to give an inductive proof that the residue at the double-forward poles gives the entire bubble coefficient, $11/3 \times A_n^{\text{tree}}$, for general split-helicity N^{k}MHV amplitudes. The proof is as follows:

1. Using the CSW representation of N^{k}MHV amplitude, the diagrams that appear inside the terminal cut can be categorized by the number of z dependent CSW propagators. For a given k there will be at most k propagators that have non-trivial z dependence. Diagrams that have $p < k$, z-dependent CSW propagators will be diagrams that have already appeared in the analysis for N^{p}MHV amplitudes, hence are known to give 11/6 times the corresponding CSW tree diagram.

2. There will be a unique diagram that has k, z-dependent, CSW propagators. To evaluate this diagram, we note that the $k+4$-point split-helicity N^{k}MHV amplitude is the same as a $k+4$-point adjacent $\overline{\text{MHV}}$ amplitude, for which we know that the forward limit poles gives $11/6A_n^{\text{tree}}$. In the CSW representation, since all other diagrams already evaluate to 11/6 times the corresponding tree diagram, as discussed in the previous step, this final diagram must as well.

3. For arbitrary n, one simply adds additional positive-helicity legs to MHV vertices. These extra states do not participate in the expansion around the pole at $z \to \infty$ or in the dLIPS integral. The modification only appears as an overall factor, thus proves that for general n this last diagram also evaluates to 11/6 times the corresponding CSW tree diagram.

4. Summing all the CSW diagrams in the double-cut, we obtain $11/6 \times A_n^{\text{tree}}$ for the forward pole

contribution to the bubble coefficient from the terminal cut.

5. The other internal helicity configuration evaluates in the same way, on the other helicity-preserving double-forward terminal pole. Combining with the above result yields $\mathcal{C}_2 = 11/3 \times A_n^{\text{tree}}$

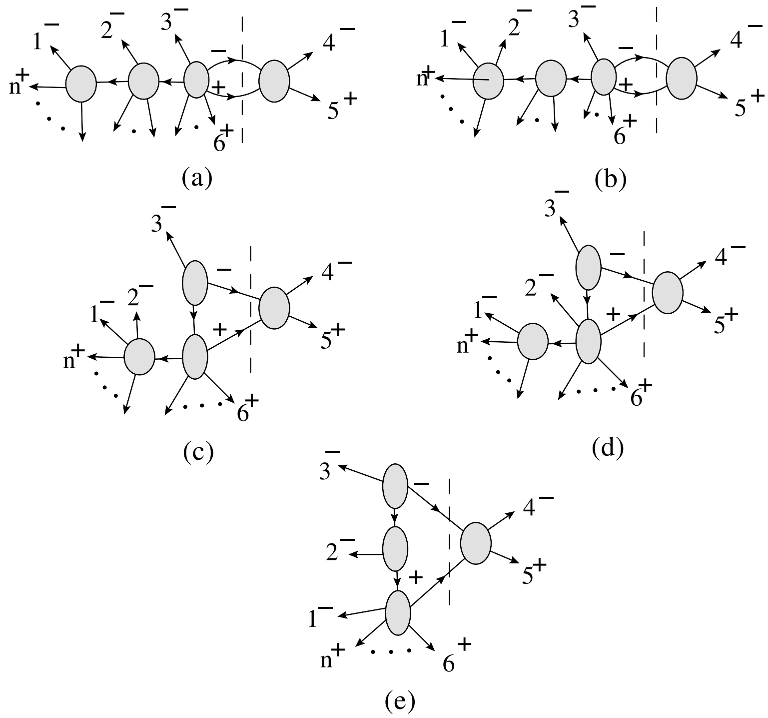


Figure 2.17: The CSW representation of the terminal cut for N^{2}MHV amplitude. The evaluation of diagrams (a) and (b) are identical to that of adjacent MHV amplitudes, while the evaluation of diagrams (c) and (d) are identical to split helicity NMHV amplitude. Diagram (e) requires the more nuanced argument of section 2.6.2.

We use the N^{2}MHV amplitude to illustrate the above steps. The CSW representation for the N^{2}MHV terminal cut is given in Fig. 2.17. Diagrams (a) and (b) have no z-dependent CSW propagators, and hence from the point of view of extracting the constant piece at $z \to \infty$ and integrating over dLIPS, the two left most MHV vertex are just spectators and the evaluation is on the right most MHV vertex on the left hand side of the cut. Thus evaluation of diagrams (a) and (b), is identical to evaluation of adjacent MHV amplitudes. For diagrams (c) and (d), there is one z-dependent CSW propagator. The MHV vertex where 1^- sits is again a spectator and the evaluation is identical to that of diagram (c) in Fig. 2.15 for the NMHV amplitude, and hence evaluates to 11/6 times the corresponding CSW tree diagram. Finally, for the unique diagram (e), we use the argument that for $n = 6$, this is simply the $\overline{\text{MHV}}$ amplitude, from which we deduce that this term must also evaluate

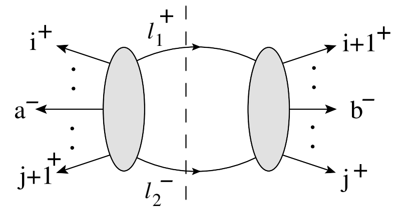


Figure 2.18: The (l_1^+, l_2^-) helicity configuration for the two-particle cut $(j+1,..,a,..,i|i+1,..,b,..,j)$ of $A_n^{\rm MHV}(a^-,b^-)$.

to 11/6 times the corresponding tree diagram. This result will not be modified for $n>6$, and hence completes the proof.

2.7 Sum of MHV bubble coefficients for pure Yang-Mills

As mentioned in subsection 2.4.3, the observed structure of cancellations for $\mathcal{N}=1,2$ super Yang-Mills theory is present in pure Yang-Mills as well. However, it is more involved to derive this since the $\mathcal{O}(z^0)$ part of the BCFW-shifted two-particle cut contains higher-order collinear poles. Nevertheless, adjacent channels again share these higher-order CCP, and their contribution to the sum of bubble coefficient also cancels. The cancellation of CCP renders the summation down to the terminal poles, which evaluate to $11/6A_n^{\rm tree}$. Here, we explicitly deal with the details of this computation.

We begin with a generic BCFW-shifted two-particle cut of $(j+1,..,a,..,i|i+1,..,b,..,j)$ for non-adjacent MHV amplitude $A_n^{\rm MHV}(a^-,b^-)$. Choosing the helicity configuration for the internal lines to be (l_1^+, l_2^-) on the LHS of the cut, as shown in Fig. 2.18, one has:

$$S_{a,b}^{(i,j)} \quad = \quad A_n^{\rm tree}\left\{\frac{\langle i,i+1\rangle\langle b,\hat{l}_1\rangle}{\langle i,\hat{l}_1\rangle\langle \hat{l}_1,i+1\rangle}\right\}\left\{\frac{\langle j,j+1\rangle\langle a,l_2\rangle}{\langle j,l_2\rangle\langle l_2,j+1\rangle}\right\}\frac{\langle l_2,l_1\rangle}{\langle a,b\rangle}\left(\frac{\langle a,l_2\rangle\langle b,\hat{l}_1\rangle}{\langle a,b\rangle\langle l_1,l_2\rangle}\right)^3 . \tag{2.52}$$

Let us extract the bubble coefficient by shifting the loop legs as $|\hat{l}_1\rangle \to |l_1\rangle + z|l_2\rangle$. Note that for our choice of shift, it will be convenient to take $|l_2\rangle = |\lambda\rangle$ as the dLIPS integration spinor. Under the dLIPS integration, there are three kinds of poles that would contribute to the holomorphic anomaly: (1) the $1/\langle\lambda\alpha\rangle$ poles that arise from writing the dLIPS integral as a total derivative, (2) the collinear poles of the form $1/\langle l_2 i\rangle$ and (3) the poles that come from expanding $1/\langle\hat{l}_1 i\rangle$ in $1/z$ to obtain the $\mathcal{O}(z^0)$ piece at $z\to\infty$.

We can remove the poles of type (1) by choosing $|\alpha\rangle = |a\rangle$, since the factor of $\langle l_2, a\rangle$ in the numerator of Eq. (2.52) will cancel this pole. Thus the only contributions remaining are of type (2) and (3). We rewrite Eq. (2.52) such that each type of pole is separated:

$$S_{a,b}^{(i,j)} = A_n^{\rm tree}\frac{\langle b,\hat{l}_1\rangle^3}{\langle l_1,l_2\rangle^2}\left\{\frac{\langle i,b\rangle}{\langle \hat{l}_1,i\rangle}-\frac{\langle i+1,b\rangle}{\langle \hat{l}_1,i+1\rangle}\right\}\left\{\frac{\langle j,a\rangle}{\langle l_2,j\rangle}-\frac{\langle j+1,a\rangle}{\langle l_2,j+1\rangle}\right\}\left(\frac{-\langle a,l_2\rangle^3}{\langle a,b\rangle^4}\right).$$

Next, we expand around $z\to\infty$ obtaining,

$$\begin{aligned}\mathcal{S}_{a,b}^{(i,j)}(l)\equiv S_{a,b}^{(i,j)}\big|_{\mathcal{O}(z^0)} &= A_n^{\rm tree}\left\{\mathcal{G}_{a,b,i}^{(i,j)}(l)-\mathcal{G}_{a,b,i+1}^{(i,j)}(l)\right\}\left\{\frac{\langle j,a\rangle}{\langle \lambda,j\rangle}-\frac{\langle j+1,a\rangle}{\langle \lambda,j+1\rangle}\right\}\left(\frac{-\langle a,\lambda\rangle^3}{\langle a,b\rangle^4}\right),\\ &= A_n^{\rm tree}\big\{I\big\}\big\{J\big\}\left(\frac{-\langle a,\lambda\rangle^3}{\langle a,b\rangle^4}\right)\end{aligned} \tag{2.53}$$

where we've used $\big\{I\big\}$ and $\big\{J\big\}$ as a short hand notation for the terms in the curly bracket. Later we will see manifest cancellation of CCP for the terms in $\big\{I\big\}$ under the summation over the i-indices, and similarly for $\big\{J\big\}$ under the summation over the j-indices. The new functional $\mathcal{G}_{a,b,i}^{(i,j)}$ is defined as:

$$\begin{aligned}\mathcal{G}_{a,b,i}^{(i,j)} &\equiv \frac{\langle i,b\rangle}{\langle l_1,l_2\rangle^2}\frac{\langle b,\hat{l}_1\rangle^3}{\langle \hat{l}_1,i\rangle}\Bigg|_{\mathcal{O}(1)} = \frac{\langle i,b\rangle}{\langle l_1,l_2\rangle^2}\left(3\frac{\langle b,l_1\rangle^2\langle b,l_2\rangle}{\langle l_2,i\rangle}-3\frac{\langle b,l_1\rangle\langle b,l_2\rangle^2\langle l_1,i\rangle}{\langle l_2,i\rangle^2}+\frac{\langle b,l_2\rangle^3\langle l_1,i\rangle^2}{\langle l_2,i\rangle^3}\right)\\ &= \frac{\langle i,b\rangle}{\langle l|P_{i,j}|l]^2}\left(3\frac{\langle b|P_{i,j}|l]^2\langle b,\lambda\rangle}{\langle \lambda,i\rangle}+3\frac{\langle b|P_{i,j}|l]\langle b,\lambda\rangle^2\langle i|P_{i,j}|l]}{\langle \lambda,i\rangle^2}+\frac{\langle b,\lambda\rangle^3\langle i|P_{i,j}|l]^2}{\langle \lambda,i\rangle^3}\right)\end{aligned} \tag{2.54}$$

and similarly,

$$\mathcal{G}_{a,b,i+1}^{(i,j)} \equiv \frac{\langle i+1,b\rangle}{\langle l|P_{i,j}|l]^2}\left(3\frac{\langle b|P_{i,j}|l]^2\langle b,\lambda\rangle}{\langle \lambda,i+1\rangle}+3\frac{\langle b|P_{i,j}|l]\langle b,\lambda\rangle^2\langle i+1|P_{i,j}|l]}{\langle \lambda,i+1\rangle^2}+\frac{\langle b,\lambda\rangle^3\langle i+1|P_{i,j}|l]^2}{\langle \lambda,i+1\rangle^3}\right). \tag{2.55}$$

Note the function $\mathcal{G}_{a,b,i}^{i,j}(l)$ has higher-order (aka not simple) collinear poles in $\langle\lambda,i\rangle$, which will require extra care in using the holomorphic anomaly as we later discuss.

The dLIPS integral of (2.53) is localized by the four poles appearing in the curly brackets, and it will be convenient to separate the contributions from the first and second curly brackets. Writing,

$$\sum_{i,j}\frac{-1}{2\pi i}\int d\text{LIPS}\ \ \mathcal{S}_{a,b}^{(i,j)}(l)=\sum_{i,j}\frac{-1}{2\pi i}\int d\text{LIPS}\ \ \left[\mathcal{S}_{a,b}^{(i,j)}(l)\Big|_{\{J\}}+\mathcal{S}_{a,b}^{(i,j)}(l)\Big|_{\{I\}}\right], \tag{2.56}$$

where $|_{\{I\}}$ indicates the contributions that arises from the presence of poles in $\{I\}$. We first consider the sum of residues of the simple poles $1/\langle\lambda, j\rangle$ and $1/\langle\lambda, j+1\rangle$ in the second curly bracket:

$$\begin{aligned}
\sum_{i,j} \mathcal{S}_{a,b}^{(i,j)}(l)\Big|_{\{J\}} &= \sum_{i,j} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i}^{(i,j)}(l) - \mathcal{G}_{a,b,i+1}^{(i,j)}(l) \Big\} \Big\{ \frac{\langle j, a\rangle}{\langle\lambda, j\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle l,j\rangle} \\
&\quad - \sum_{i,j} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i}^{(i,j)}(l) - \mathcal{G}_{a,b,i+1}^{(i,j)}(l) \Big\} \Big\{ \frac{\langle j+1, a\rangle}{\langle\lambda, j+1\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,j+1\rangle} \\
&= \sum_{i,j} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i}^{(i,j)}(l) - \mathcal{G}_{a,b,i+1}^{(i,j)}(l) \Big\} \Big\{ \frac{\langle j, a\rangle}{\langle\lambda, j\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,j\rangle} \\
&\quad - \sum_{i,j'} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i}^{(i,j'-1)}(l) - \mathcal{G}_{a,b,i+1}^{(i,j'-1)}(l) \Big\} \Big\{ \frac{\langle j', a\rangle}{\langle\lambda, j'\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,j'\rangle} ,
\end{aligned} \tag{2.57}$$

where we've used $|_{\langle\lambda,j\rangle}$ to indicate the collinear pole on which the integrand will be localized. From (2.54), we see that $\mathcal{G}_{a,b,i}^{(i,j)}(l) = \mathcal{G}_{a,b,i}^{(i,j-1)}(l)$ when localized at $l \to j$.[15] Treating j as and j' as dummy variables, the two summations simply cancel with each other and one is left with zero! Of course this is the wrong result and the subtlety lies in the summation limits. We will discuss the limits in detail in the next subsection. For now, we will show the same cancellation occurs for the higher-order poles in the first curly bracket.

The contributions from the poles in the first curly bracket in (2.53) can be written as:

$$\begin{aligned}
\sum_{i,j} \mathcal{S}_{a,b}^{(i,j)}(l)\Big|_{\{I\}} &= \sum_{j,i} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i}^{(i,j)}(l) \Big\} \Big\{ \frac{\langle j, a\rangle}{\langle\lambda, j\rangle} - \frac{\langle j+1, a\rangle}{\langle\lambda, j+1\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,i\rangle} \\
&\quad - \sum_{j,i} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i+1}^{(i,j)}(l) \Big\} \Big\{ \frac{\langle j, a\rangle}{\langle\lambda, j\rangle} - \frac{\langle j+1, a\rangle}{\langle\lambda, j+1\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,i+1\rangle} \\
&= \sum_{j,i} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i}^{(i,j)}(l) \Big\} \Big\{ \frac{\langle j, a\rangle}{\langle\lambda, j\rangle} - \frac{\langle j+1, a\rangle}{\langle\lambda, j+1\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,i\rangle} \\
&\quad - \sum_{j,i'} A_n^{\text{tree}} \Big\{ \mathcal{G}_{a,b,i'}^{(i'-1,j)}(l) \Big\} \Big\{ \frac{\langle j, a\rangle}{\langle\lambda, j\rangle} - \frac{\langle j+1, a\rangle}{\langle\lambda, j+1\rangle} \Big\} \Big(\frac{-\langle a, \lambda\rangle^3}{\langle a, b\rangle^4} \Big) \Big|_{\langle\lambda,i'\rangle} .
\end{aligned} \tag{2.58}$$

Here, the integral will be localized by the poles that are present in $\mathcal{G}_{a,b,i}^{(i,j)}(l)$ which, in addition to simple poles, has higher-order poles at the same kinematic point. Repeated use of the Schouten identity allows one to extract the coefficients of the higher-order poles from those of the simple poles [30]. These simple poles contribute to the dLIPS-integral; which we review in appendix 2.3.

[15] This again can be seen from the fact that on the pole, $P_{i,j}|j] = P_{i,j-1}|j]$.

Denoting the resulting expression as $\mathcal{H}^{(i,j)}_{a,b,i}(l)$, we have

$$\sum_{i,j}\mathcal{S}^{(i,j)}_{a,b}(l)\bigg|_{\{I\}} \quad = \quad \sum_{j,i}\frac{A_n^{\text{tree}}}{\langle a,b\rangle^4}\left\{\mathcal{H}^{(i,j)}_{a,b,i}(l)-\mathcal{H}^{(i,j+1)}_{a,b,i}(l)-\mathcal{H}^{(i-1,j)}_{a,b,i}(l)+\mathcal{H}^{(i-1,j+1)}_{a,b,i}(l)\right\}\bigg|_{\langle\lambda,i\rangle} . \tag{2.59}$$

The explicit form of $\mathcal{H}^{(i,j)}_{a,b,i}(l)$ is given in Eq. (2.65). The key fact of $\mathcal{H}^{(i,j)}_{a,b,i}(l)$ and $\mathcal{H}^{(i-1,j)}_{a,b,i}(l)$ is that they become identical when the integrand is evaluated on the pole $1/\langle\lambda,i\rangle$ and integrated on the real contour $\tilde{\lambda}=\bar{\lambda}$. Therefore Eq. (2.59) again gives zero!

While the limits of the summation requires careful treatment, our analysis shows that indeed for non-adjacent MHV, the cancellation of CCP again reduces the sum of bubble coefficients to a few terminal terms which we will now identify.

2.7.1 dLIPS integrals of higher-order poles

As we have seen, generic pole terms in non-adjacent MHV bubble coefficients' $g(l)$s generically have higher-order poles. We evaluate the integrands in a manner following that in section 2.3 of [30]. Specifically, the dLIPS integrands are rational functions of l of degree negative 2, i.e. degree -2. We can recursively reduce the degree of l in the numerator and denominator by one unit each, through repeated application of the following Schouten identity [30]:

$$\frac{\langle a,\lambda\rangle}{\langle ,\lambda\rangle\langle\gamma,\lambda\rangle}=\frac{\langle a,\beta\rangle}{\langle\gamma,\beta\rangle}\frac{1}{\langle\beta,\lambda\rangle}+\frac{\langle a,\gamma\rangle}{\langle\beta,\gamma\rangle}\frac{1}{\langle\gamma,\lambda\rangle} . \tag{2.60}$$

Repeated application reduces integrands with higher-order poles to sums of integrands with either simple poles, or to multiple poles, such as $1/\langle a,\lambda\rangle^2$, $\langle x,\lambda\rangle/\langle a,\lambda\rangle^3$ etc., with trivial residues as $|l\rangle\to|a\rangle$. The generic form for the residues at second- and third-order poles are:

$$\frac{1}{\langle ,\lambda\rangle^2}\prod_{i=1}^{n}\frac{\langle a_i,\lambda\rangle}{\langle b_i,\lambda\rangle}\bigg|_{\langle\lambda,\beta\rangle} \quad = \quad \prod_{i=1}^{n}\frac{\langle a_i,\beta\rangle}{\langle b_i,\beta\rangle}\sum_{1\le i\le n}\frac{\langle a_i,b_i\rangle}{\langle a_i,\beta\rangle\langle b_i,\beta\rangle} , \tag{2.61}$$

$$\begin{aligned}\frac{\langle\xi,\lambda\rangle}{\langle\beta,\lambda\rangle^3}\prod_{i=1}^{n}\frac{\langle a_i,\lambda\rangle}{\langle b_i,\lambda\rangle}\bigg|_{\langle\lambda,\beta\rangle} \quad = \quad & \langle\xi,\beta\rangle\prod_{i=1}^{n}\frac{\langle a_i,\beta\rangle}{\langle b_i,\beta\rangle}\Bigg\{\sum_{1\le i\le j\le n}\frac{\langle a_i,b_i\rangle}{\langle a_i,\beta\rangle\langle b_i,\beta\rangle}\frac{\langle a_j,b_j\rangle}{\langle a_j,\beta\rangle\langle b_j,\beta\rangle} \\ & +\sum_{1\le k\le n}\frac{\langle a_k,b_k\rangle}{\langle a_k,\beta\rangle\langle b_k,\beta\rangle}\frac{\langle a_k,\xi\rangle}{\langle\xi,\beta\rangle\langle a_k,\beta\rangle}\Bigg\} .\end{aligned} \tag{2.62}$$

With this, we can factor out the irrelevant multiple poles in any expression, for example we have the following rewriting:

$$\frac{\langle b,\lambda\rangle^2\langle\lambda,a\rangle^3}{\langle\lambda,i\rangle^2\langle\lambda,j\rangle\langle l|P_{i,j}|l]^4}\Bigg|_{\langle\lambda,i\rangle} = -\frac{\langle i,a\rangle^3\langle i,b\rangle^2\left(\frac{2\langle a|P_{i,j}|i]}{\langle i|P_{i,j}|i]\langle i,a\rangle}+\frac{2\langle b|P_{i,j}|i]}{\langle i|P_{i,j}|i]\langle i,b\rangle}+\frac{\langle a,j\rangle}{\langle i,a\rangle\langle i,j\rangle}\right)}{\langle i|P_{i,j}|i]^4\langle i,j\rangle}, \tag{2.63}$$

$$\begin{aligned}\frac{\langle\lambda,a\rangle^3\langle b,\lambda\rangle^3}{\langle\lambda,j\rangle\langle\lambda,i\rangle^3\langle l|P_{i,j}|l]^4}\Bigg|_{\langle\lambda,i\rangle} &= \frac{\langle a,i\rangle^3\langle b,i\rangle^3}{\langle i|P_{i,j}|i]^4\langle i,j\rangle}\left(\frac{3\langle a|P_{i,j}|i]^2}{\langle i|P_{i,j}|i]^2\langle a,i\rangle^2}+\frac{3\langle b|P_{i,j}|i]^2}{\langle i|P_{i,j}|i]^2\langle b,i\rangle^2}+\frac{2\langle a|P_{i,j}|i]\langle a,b\rangle}{\langle i|P_{i,j}|i]\langle a,i\rangle^2\langle b,i\rangle}\right.\\ &+\frac{4\langle a|P_{i,j}|i]\langle b|P_{i,j}|i]}{\langle i|P_{i,j}|i]^2\langle a,i\rangle\langle b,i\rangle}+\frac{\langle a,j\rangle^2}{\langle a,i\rangle^2\langle i,j\rangle^2}+\frac{2\langle a|P_{i,j}|i]\langle a,j\rangle}{\langle i|P_{i,j}|i]\langle a,i\rangle^2\langle i,j\rangle}\\ &\left.+\frac{\langle a,b\rangle\langle a,j\rangle}{\langle a,i\rangle^2\langle b,i\rangle\langle i,j\rangle}+\frac{2\langle b|P_{i,j}|i]\langle a,j\rangle}{\langle i|P_{i,j}|i]\langle a,i\rangle\langle b,i\rangle\langle i,j\rangle}\right). \end{aligned}\tag{2.64}$$

Combining these results, we can rewrite $\mathcal{G}^{(i,j)}_{a,b,i}(l)\frac{\langle\lambda,a\rangle^3\langle j,a\rangle}{\langle l|P_{i,j}|l]^2\langle\lambda,j\rangle}$ to $\mathcal{H}^{(i,j)}_{a,b,i}$ as in (2.59):

$$\begin{aligned}\mathcal{G}^{(i,j)}_{a,b,i}(l)\frac{\langle\lambda,a\rangle^3\langle j,a\rangle}{\langle l|P_{i,j}|l]^2\langle\lambda,j\rangle}\Bigg|_{\langle\lambda,i\rangle} &= \frac{\langle i,a\rangle^3\langle j,a\rangle}{\langle i,j\rangle}\langle i,b\rangle\left(3\frac{\langle b|P_{i,j}|i]^2\langle b,i\rangle}{\langle l|P_{i,j}|l]^2}\right.\\ &+3\langle j,a\rangle\langle b|P_{i,j}|i]\langle i|P_{i,j}|i]\times(2.63)\\ &\left.+\langle j,a\rangle\langle i|P_{i,j}|i]^2\times(2.64)\right). \end{aligned}\tag{2.65}$$

The upshot is that (2.65) has vanishing residue at the poles $l=a$ and $l=b$. Note that if one considers $\mathcal{H}^{(i-1,j)}_{a,b,i}$, the only difference is substituting $P_{i,j}$ in $\mathcal{H}^{(i,j)}_{a,b,i}$ with $P_{i-1,j}$. It can be easily seen that on the pole $1/\langle\lambda,i\rangle$, the two are equivalent.

2.7.2 Terminal poles, terminal cuts and their evaluation

In determining the limits of the summation, one has to avoid configurations where there is a three point amplitude one side of the cut, as these produce massless bubbles that are set to zero in dimensional regularization. This implies that in the summation of i,j in Eq. (2.57) and Eq. (2.59), the summation limit of one index will depend on the value of the other.

For Eq. (2.57) one sums over the index j first, and the limit is given as:

$$\sum_{i,j}=\sum_{i=a+1}^{b-2}\sum_{j=b}^{a-1}+\sum_{j=b+1}^{a-1}\Big|_{i=b-1}+\sum_{j=b}^{a-2}\Big|_{i=a}, \tag{2.66}$$

where $|_{i=b-1}$ indicates the index i is held fixed to be $b-1$. Using $j'=j+1$, the summation limit for j' is given as:

$$\sum_{i,j'=j+1}=\sum_{i=a+1}^{b-2}\sum_{j'=b+1}^{a}+\sum_{j'=b+2}^{a}\Big|_{i=b-1}+\sum_{j'=b+1}^{a-1}\Big|_{i=a}. \tag{2.67}$$

Looking back at Eq. (2.57) we see that there are mismatches in the limits between to two sums, and hence the cancellation is not complete, leaving behind:

$$\sum_{i=a+1}^{b-2} X_{a,b}^{(i,j)}\Big|_{j=b} - \sum_{i=a+1}^{b-2} X_{a,b}^{(i,j')}\Big|_{j'=a} + X_{a,b}^{(i,j)}\Big|_{\substack{i=b-1\\ j=b+1}} - X_{a,b}^{(i,j')}\Big|_{\substack{i=b-1\\ j'=a}} + X_{a,b}^{(i,j)}\Big|_{\substack{i=a\\ j=b}} - X_{a,b}^{(i,j')}\Big|_{\substack{i=a\\ j'=a-1}} \tag{2.68}$$

where $X_{a,b}^{(i,j)} = -A_n^{\text{tree}}\{\mathcal{G}_{a,b,i}^{(i,j)}(l) - \mathcal{G}_{a,b,i+1}^{(i,j)}(l)\}\frac{\langle j,a\rangle}{\langle\lambda,j\rangle}\frac{\langle a,\lambda\rangle^3}{\langle a,b\rangle^4}|_{\langle\lambda,j\rangle}$. The first two terms in Eq. (2.68) evaluates to zero. To see this note that these two sums are evaluated on the pole $1/\langle\lambda,b\rangle$ and $1/\langle\lambda,a\rangle$ respectively. Looking at the summand in Eq. (2.57) there is a factor $\langle\lambda,a\rangle$ in the numerator while $\mathcal{G}_{a,b,i}^{(i,j)}$ has at least one $\langle\lambda,b\rangle$ in the numerator, as can be seen from Eq. (2.54). For the same reason, the fourth and fifth term vanishes as well. The remaining terms are given by:

$$\begin{aligned}\sum_{i,j}\mathcal{S}_{a,b}^{(i,j)}(l)\Big|_{\{J\}} \quad = \quad & A_n^{\text{tree}}\left\{\mathcal{G}_{a,b,b-1}^{(b-1,b+1)}(l) - \mathcal{G}_{a,b,b}^{(b-1,b+1)}(l)\right\}\left\{\frac{\langle b+1,a\rangle}{\langle\lambda,b+1\rangle}\right\}\left(\frac{-\langle a,\lambda\rangle^3}{\langle a,b\rangle^4}\right)\Big|_{\langle\lambda,b+1\rangle}\\ & -A_n^{\text{tree}}\left\{\mathcal{G}_{a,b,a}^{(a,a-2)}(l) - \mathcal{G}_{a,b,a+1}^{(a,a-2)}(l)\right\}\left\{\frac{\langle a-1,a\rangle}{\langle\lambda,a-1\rangle}\right\}\left(\frac{-\langle a,\lambda\rangle^3}{\langle a,b\rangle^4}\right)\Big|_{\langle\lambda,a-1\rangle}\end{aligned} \tag{2.69}$$

Therefore, we see the complicated summation (2.57) reduces to only two terms: the residue at the pole $\langle\lambda,a-1\rangle = 0$ in channel $(i=a, j=a-2)$, and the residue of the pole $\langle\lambda,b+1\rangle = 0$ in channel $(i=b-1, j=b+1)$.

We now look at Eq. (2.59), where the index i was summed first. The summation limit is given by:

$$\sum_{j,i} = \sum_{j=b+1}^{a-2}\sum_{i=a}^{b-1} + \sum_{i=a+1}^{b-1}\Big|_{j=a-1} + \sum_{i=a}^{b-2}\Big|_{j=b}. \tag{2.70}$$

Recalling that $i' = i+1$, the summation limit for i' is given by:

$$\sum_{j,i'} = \sum_{j=b+1}^{a-2}\sum_{i'=a+1}^{b} + \sum_{i'=a+2}^{b}\Big|_{j=a-1} + \sum_{i'=a+1}^{b-1}\Big|_{j=b}. \tag{2.71}$$

Again the mismatch of the summation limits for i and i' leads to uncancelled terms in Eq. (2.59), given by:

$$\sum_{j=b+1}^{a-2} Y_{a,b}^{(i,j)}\Big|_{i=a} - \sum_{j=b+1}^{a-2} Y_{a,b}^{(i',j)}\Big|_{i'=b} + Y_{a,b}^{(i,j)}\Big|_{\substack{i=a+1\\ j=a-1}} - Y_{a,b}^{(i',j)}\Big|_{\substack{i'=b\\ j=a-1}} + Y_{a,b}^{(i,j)}\Big|_{\substack{i=a\\ j=b}} - Y_{a,b}^{(i',j)}\Big|_{\substack{i'=b-1\\ j=b}} \tag{2.72}$$

where $Y_{a,b}^{(i,j)} = A_n^{\text{tree}}\{\mathcal{H}_{a,b,i}^{(i,j)}(l) - \mathcal{H}_{a,b,i}^{(i,j+1)}(l) - \mathcal{H}_{a,b,i}^{(i-1,j)}(l) + \mathcal{H}_{a,b,i}^{(i-1,j+1)}(l)\}/\langle a,b\rangle^4|_{\langle\lambda,i\rangle}$. The second and fourth term in Eq. (2.72) evaluates to zero since it has vanishing residue on the pole $1/\langle\lambda, b\rangle$, as can be seen from the presence of $\langle\lambda, b\rangle$ in the numerator of Eq. (2.54) and Eq. (2.55). The first and fifth term also vanishes due to the $\langle\lambda, a\rangle^3$ in the numerator of Eq. (2.58). As a result, the sum in Eq. (2.59) reduces to

$$\sum_{i,j}\mathcal{S}^{(i,j)}_{a,b}(l)\Big|_{\{I\}} = \frac{A^{\text{tree}}_n}{\langle a,b\rangle^4}\Bigg\{\left(\mathcal{H}^{(a+1,a-1)}_{a,b,a+1}(l)-\mathcal{H}^{(a+1,a)}_{a,b,a+1}(l)\right)\Big|_{\langle\lambda,a+1\rangle} - \left(\mathcal{H}^{(b-2,b)}_{a,b,b-1}(l)-\mathcal{H}^{(b-2,b+1)}_{a,b,b-1}(l)\right)\Big|_{\langle\lambda,b-1\rangle}\Bigg\} \tag{2.73}$$

Thus the sum localizes to the pole $\langle\lambda, a+1\rangle = 0$ in channel $(i = a+1, j = a-1)$ and $\langle\lambda, b-1\rangle = 0$ in channel $(i = i'-1 = b-2, j = b)$.

Collecting all the pieces we now have:

$$\begin{aligned}\mathcal{C}_2(l_1^+,l_2^-) \;=\; & \frac{-1}{2\pi i}\int_{\bar\lambda=\tilde\lambda} d\text{LIPS}\;\frac{A^{\text{tree}}_n}{\langle a,b\rangle^4}\Bigg\{-\left(\mathcal{G}^{(b-1,b+1)}_{a,b,b-1}(l)-\mathcal{G}^{(b-1,b+1)}_{a,b,b}(l)\right)\frac{\langle b+1,a\rangle}{\langle\lambda,b+1\rangle}\langle a,\lambda\rangle^3\Big|_{\langle\lambda,b+1\rangle}\\ & +\left(\mathcal{G}^{(a,a-1)}_{a,b,a}(l)-\mathcal{G}^{(a,a-1)}_{a,b,a+1}(l)\right)\frac{\langle a-1,a\rangle}{\langle\lambda,a-1\rangle}\langle a,\lambda\rangle^3\Big|_{\langle\lambda,a-1\rangle}\\ & \left(\mathcal{H}^{(a+1,a-1)}_{a,b,a+1}(l)-\mathcal{H}^{(a+1,a)}_{a,b,a+1}(l)\right)\Big|_{\langle\lambda,a+1\rangle} - \left(\mathcal{H}^{(b-2,b)}_{a,b,b-1}(l)-\mathcal{H}^{(b-2,b+1)}_{a,b,b-1}(l)\right)\Big|_{\langle\lambda,b-1\rangle}\Bigg\}\end{aligned} \tag{2.74}$$

We now evaluate the integral. As discussed in appendix 2.3, writing the above integrand as a total derivative will always introduce a factor of $[l|P|\alpha\rangle$. With the choice of $|\alpha\rangle = |a\rangle$, terms that are evaluated on the pole $1/\langle la \pm 1\rangle$ vanishes since $[a\pm1|P_{a,a\pm1}|a\rangle = 0$. Thus the cancellation of CCP and the judicious choice or reference spinor reduces the sum of the bubble coefficient for n-point MHV amplitude to simply:

$$\begin{aligned}\mathcal{C}_2(l_1^+,l_2^-) \;=\; & \frac{A^{\text{tree}}_n}{\langle a,b\rangle^4}\frac{1}{2\pi i}\int_{\bar\lambda=\tilde\lambda} d\text{LIPS}\;\Bigg\{\left(\mathcal{G}^{(b-1,b+1)}_{a,b,b-1}(l)-\mathcal{G}^{(b-1,b+1)}_{a,b,b}(l)\right)\frac{\langle b+1,a\rangle}{\langle\lambda,b+1\rangle}\langle a,\lambda\rangle^3\Big|_{\langle\lambda,b+1\rangle}\\ & +\left(\mathcal{H}^{(b-2,b)}_{a,b,b-1}(l)-\mathcal{H}^{(b-2,b+1)}_{a,b,b-1}(l)\right)\Big|_{\langle\lambda,b-1\rangle}\Bigg\}.\end{aligned} \tag{2.75}$$

Expanding the parenthesis, there are four different terms to be evaluated. Explicit evaluation shows the first two terms sum to cancel the last term. Thus we have:

$$\mathcal{C}_2(l_1^+,l_2^-) = \frac{A^{\text{tree}}_n}{\langle a,b\rangle^4}\frac{1}{2\pi i}\int_{\bar\lambda=\tilde\lambda} d\text{LIPS}\;\left(\mathcal{H}^{(b-2,b)}_{a,b,b-1}(l)\right)\Big|_{l\to b-1} = \frac{11}{6}A^{\text{tree}}_n \tag{2.76}$$

Adding this with the same calculation for the other helicity configuration, (l_1^-, l_2^+), one obtains the desired result, $\mathcal{C}_2 = \frac{11}{3}A^{\text{tree}} = -\beta_0 A^{\text{tree}}$

2.8 Conclusion and future directions

In this chapter, we have studied the proportionality between the sum of bubble coefficients and the tree amplitude, which is required for renormalizability. For theories where Feynman diagram analysis is tractable, such as scalar theory and pure scalar amplitudes of Yukawa theory, we find that the bubble coefficient only receives contributions from a small class of one-loop diagrams. The contribution of each diagram is proportional to a tree-diagram, and hence summing over all one-loop diagrams that give non-trivial contributions, is equivalent to summing over tree-diagrams. Crucially, these *new* tree diagrams are not necessarily present in the original tree amplitude of the theory. Through restricting our attention to theories where these new tree structures match those in the original tree amplitude, we accurately reproduce the known renormalization conditions derived from power counting analysis.

For (super)Yang-Mills theory, we show that the bubble coefficient for MHV amplitudes can be organized in terms of their origin as collinear poles, which are responsible for the nontrivial contribution to the dLIPS integration, in the two-particle cuts. This representation reveals the existence of systematic cancellation in the sum of bubble coefficient. In particular, the residues of common collinear poles (CCP) cancels, and the sum telescopes down to unique terminal poles. These are poles that arise from cuts that have at least a 4-point tree amplitude on one side of the cut, and the helicity configuration of the internal legs must match that of the two external legs on the four-point tree amplitude, as shown in Fig. 2.2. We conjecture that these double forward poles are the only non-trivial contribution to the sum of bubble coefficients for any helicity configuration. As further evidence, we explicitly proved that for split helicity n-point N^{k}MHV amplitudes, the contribution of each terminal pole indeed give $11/6$ times the tree amplitude.

For more generic external helicity configurations, it will be interesting to see how the contributions from the multi-particle poles cancel with each other. An even more interesting example would be gravity. It is well-known that pure gravity is one-loop finite [31]. The bubble coefficient is non-vanishing for generic two-particle cuts, and hence massive cancellation must occur. The lack of color ordering for gravity amplitudes indicate the pole structure that gives rise to the non-trivial contributions for the dLIPS integral is more complicated than Yang-Mills: presumably new cancellation mechanisms are required even for MHV amplitudes.

We have demonstrated that the UV divergence of the one-loop gauge theory amplitude is completely captured by the residues of a set of unique collinear poles, i.e. it is controlled by a residue at finite loop momentum value. If the same holds for gravity, then through KLT relations [32] the residue of gravity is intimately tied to gauge theories, and it will be interesting to see how the relationship allows cancellation among terminal residues, leading to the known finiteness result for gravity and it's relationship to BCJ duality [33, 34]. Note that the study of tensor bubbles has previously revealed improved UV behavior for gravity amplitudes compared to naive power counting from Einstein-Hilbert action [35]. Even though it is well known that gravity is finite at one-loop, a careful analysis of how finiteness is achieved for generic amplitudes may shed light on additional structure, as we have successfully achieved for (super) Yang-Mills amplitudes.

2.9 Acknowledgements

We thank Nima Arkani-Hamed, Simon Caron-Huot and Michael Kiermaier for many enlightening discussions. We are especially grateful to Zvi Bern, Lance Dixon and Henriette Elvang for careful reading of our draft and giving us helpful suggestions. YH is grateful for the Institute of Advanced Study for invitation as visiting member. YH would also like to thank Isaac Newton Institute for Mathematical Sciences for organizing the workshop "Recent Advances in Scattering Amplitudes", during which part of this work was completed. YH was supported by the US Department of Energy under contract DE–FG03–91ER40662. DAM is supported NSF GRFP grants DGE-1148900 and PHY-0756966. CP is supported by NSF Grant PHY-0953232, and in part by the DOE Grant DE-FG02-95ER 40899.

Chapter 3

Gauge-anomalies on-shell (short)

An intriguing difference between the traditional Lagrangian definition of perturbative quantum field theory (QFT) and the modern analytic S-matrix program, is the role of gauge symmetry. Where as gauge invariance is crucial in determining the Lagrangian and ensures unitarity of the perturbative S-matrix, such notions are completely absent in the modern on-shell approach. In the latter approach, given the free-spectrum of the theory, which is defined as representations of the little group, the lowest-multiplicity non-trivial S-matrix can be determined completely from the global symmetries of the theory. Using factorization [20, 36] as well as unitarity constraints [13, 12, 37], the entire perturbative S-matrix can then be iteratively constructed from that of the lowest order. Such an approach has led to tremendous progress in the computation of high loop-order corrections in four-dimensional super Yang-Mills [38], supergravity [39], higher-dimensional super Yang-Mills [40], as well as the determination of all-loop planar integrand of maximal super Yang-Mills [41]. Since the building blocks that enter the iterative process are completely on-shell, gauge-invariance is at all times manifest.

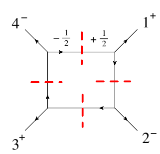
The fact that the physical observables of a QFT can be constructed without the utterance of gauge symmetry leads us to ask how consistency constraints, traditionally *imposed* by the requirement of gauge anomaly cancellation, arise in such on-shell constructions. Establishment of such constraints without knowledge of the interaction Lagrangian becomes crucial in light of the large class of supersymmetric Chern-Simons matter theories [42, 43] whose Lagrangian has been constructed only in the past five years, although their S-matrix elements can be determined independently [44, 45].

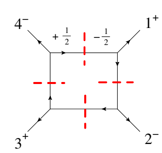
In this chapter, we address the following question: starting with a theory of chiral fermions, as we construct loop-amplitudes through the on-shell program, how do we see that the theory is sick? Tree-level amplitudes of chiral fermions are perfectly well defined. Through general unitarity methods, one necessarily obtains a unitary S-matrix. Superficially, chiral gauge theories should have perfectly sensible loop amplitudes. However, while the S-matrix is manifestly unitary, it contains spurious non-local poles. To ensure that the final result is both unitary and local, one is forced to introduce non cut-constructible rational terms to cancel the spurious poles. We will demonstrate that for chiral fermion loops, cancellation of these spurious singularities induces new factorization channels. In four-dimensions, such factorization channels are inconsistent and thus must cancel. The constraint imposed by such cancellation is precisely the vanishing of the cubic Casimir of the gauge group. In six-dimensions, if the symmetric trace of the four generators does not vanish, the new induced factorization channel reveals the presence of a new particle in the theory: the two-form in the Green-Schwarz (GS) mechanism [46].

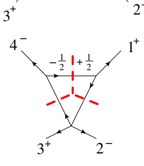
Note that here, we do not assume the existence of symmetry preserving regulators for the theory, which is the source of anomalies in the conventional Feynman diagram approach. We use scalar integrals to form a basis for reproducing all branch cuts. As we demonstrate, the final integrated amplitude is both IR- and UV-finite, and is thus independent of the regulator employed.

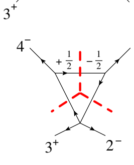
3.1 A prelude in four-dimensions

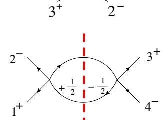
Unitarity methods naturally cast one-loop amplitudes into a basis of scalar integrals whose coefficients depend on the theory at hand. Here, we consider the fermion-loop contribution to the single trace one-loop four-gluon amplitude. For later convenience we give the scalar-integral coefficients originating from two distinct fermion helicities separately:

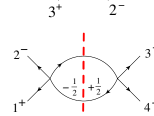
$-\dfrac{t^4s^2}{u^4}$ 

$-\dfrac{s^4t^2}{u^4}$ 

$\dfrac{t^4s}{u^4}$ 

$\dfrac{t^2s^3}{u^4}$ 





$$\frac{t(su-6st-2ut)}{6u^3} \qquad \frac{t(4s^2+2t^2-7su)}{6u^3}\,. \tag{3.1}$$

In the above, we've indicated the helicities of the fermions crossing the unitarity cut, denoted by the (red) dashed lines, and $s = (k_1 + k_2)^2$, $t = (k_2 + k_3)^2$, $u = (k_1 + k_3)^2$. Note that the triangle- and box-integral coefficients are such that the IR-divergence cancels, which is necessary due to the absence of tree-level processes for a fermion in a background gauge field.

The the parity-even part of the fermion-loop amplitude simply corresponds to the sum of the two distinct helicity configurations:

$$
\begin{aligned}
\frac{A^{\text{even}}(1^+2^-3^+4^-)}{A^{\text{tree}}} &= -\frac{st(s^2+t^2)}{2u^4}\left(\log\left(\frac{t}{s}\right)^2 + \pi^2\right) \\
&+ \left[\left(\frac{s-t}{3u} - \frac{st(s-t)}{u^3}\right)\right]\log\left(\frac{s}{t}\right) - \frac{(-s)^{-\epsilon}+(-t)^{-\epsilon}}{3\epsilon} \\
&+ R^{\text{even}}_{(1,2,3,4)}\,,
\end{aligned} \tag{3.2}
$$

where we've included a term R^{even} representing possible rational terms that are undetectable from unitarity cuts, and A^{tree} is the tree-level amplitude. The rational term can be determined from imposing locality. To see this, note that poles in the u-channel are ubiquitous throughout Eq. (3.2). These poles cannot have a local interpretation due to the color-ordering. As $u \to 0$, Eq. (3.2) behaves as:

$$
\left(Eq.\ (4.3) - R^{\text{even}}_{(1,2,3,4)}\right)\Big|_{u\to 0} = -\frac{s^2}{u^2} - \frac{s}{u} + \mathcal{O}(u^0)\,. \tag{3.3}
$$

Locality requires $R^{\text{even}}_{(1,2,3,4)}$ to cancel these spurious poles . Dimension-counting and cyclic invariance uniquely fixes it to be,

$$
R^{\text{even}}_{(1,2,3,4)} = -\frac{st}{u^2}\,. \tag{3.4}
$$

Substituting Eq. (3.4) into Eq. (3.2) reproduces known results in QCD [47]. Note that since the amplitude has an A^{tree} pre factor, the presence of a rational term can potentially introduce new residues on the physical poles of the tree-amplitude. However, due to the st factor in the numerator of Eq. (3.4), the residue vanishes.

We now turn to the parity-odd part of the fermion-loop, which is only present for chiral fermions. It is simply given by the difference of the two helicity configurations:

$$
\begin{aligned}
\frac{A^{\text{odd}}(1^+2^-3^+4^-)}{A^{\text{tree}}} &= -\frac{st(s-t)}{2u^3}\left(\log\left(\frac{t}{s}\right)^2 + \pi^2\right) \\
&- \left(\frac{2st}{u^2}\right)\log\left(\frac{-s}{-t}\right) + R^{\text{odd}}_{(1,2,3,4)}\,.
\end{aligned} \tag{3.5}
$$

First, note that the amplitude is cyclic invariant up to a sign, which is due to the use of helicity basis. As with the parity-even combination, there are spurious u-channel poles in Eq. (3.11):

$$\left(Eq.\ (4.8) - R^{\text{odd}}_{(1,2,3,4)}\right)|_{u\to 0} = -\frac{s}{u} + \mathcal{O}(u^0)\,. \tag{3.6}$$

Locality again requires such spurious poles to be canceled by $R^{\text{odd}}_{(1,2,3,4)}$. Taking into account the fact that the amplitude attains a minus sign under cyclic shift, the requisite parity-odd rational term is:

$$A^{\text{tree}} R^{\text{odd}}_{(1,2,3,4)} = A^{\text{tree}} \frac{s-t}{2u}\,. \tag{3.7}$$

However, as is plain from Eq. (3.7), this new parity-odd rational term has non-trivial contributions to the s- and t-channel residues due to its tree amplitude prefactor. This contrasts sharply with the $R^{\text{even}}_{(1,2,3,4)}$ where the factor of st ensures that the residue is zero. Herein lay the seeds of inconsistencies in parity-violating gauge theories: the rational terms that are required for locality in the parity-odd amplitude, introduce *new* corrections to residues on the s- and t- poles. This is inconsistent, since the residue for the s-channel pole, given by $\frac{1}{2}A_3^{\text{tree}}A_3^{\text{tree}}$, differs by a sign compared to that of the t-channel. This sign difference is precisely due to the fact that the amplitude is parity-odd.

Requiring these inconsistent factorization channels to be absent from the amplitude constrains the theory. To see how, note that there are 6 single-trace color structures at four-points and one-loop. Four of these contain such excess residue in the physical s-channel. Their rational terms sum to,

$$A^{\text{tree}} \frac{s-t}{2u}\left(tr[T^1T^4T^3T^2] - tr[T^1T^2T^3T^4] + (1\leftrightarrow 2)\right) \tag{3.8}$$

Thus we see that the problematic residues from the rational terms exactly cancel if the group-theory factor vanishes:

$$\begin{aligned} & tr[T^1T^2T^3T^4] - tr[T^1T^4T^3T^2] + (1\leftrightarrow 2) \\ = \quad & d^{1a4}f^{23}{}_a + d^{13a}f^{24}{}_a + d^{1a2}f^{34}{}_a + (1\leftrightarrow 2) = 0\,. \end{aligned} \tag{3.9}$$

Since the symmetry property of each term is distinct, the constraint is satisfied only if each term is individually zero. Thus imposing unitarity and locality, one arrives at the following constraint on the group-theory factor:

$$d^{abc}f^{de}{}_a = 0\,. \tag{3.10}$$

This is nothing but the anomaly cancellation condition of the non-abelian box anomaly! In summary, in using unitarity methods to construct one-loop scattering amplitudes, one encounters obstacles in implementing locality if there are chiral fermion loops. Such obstruction ceases to exist if the theory has vanishing d_{abc}.

3.2 The 6D rational term and the GS two-form

We now consider the one-loop four-point amplitude in 6D chiral gauge-theory. The little-group in six-dimensions is SO(4)=SU(2)×SU(2), and for the parity-odd contribution, we again take the difference between $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ fermion in the loop. To obtain the scalar integral coefficients, we utilize six-dimensional spinor-helicity formalism [48] as well as generalized unitarity-methods [49]. Explicit computation gives the following coefficients for the scalar box, triangle and bubble integrals respectively:

$$\begin{aligned} C_4 &= \frac{(s-t)}{6u^2}F^{(4)}, \;\; C_{3s} = -\frac{(s-t)}{6tu^2}F^{(4)}, \\ C_{3t} &= -\frac{(s-t)}{6su^2}F^{(4)}, \;\; C_{2s} = \frac{F^{(4)}}{stu}, \;\; C_{2t} = -\frac{F^{(4)}}{stu}\,, \end{aligned}$$

The function $F^{(4)}$ is explicitly given as:[1]

$$F^{(4)} \equiv \langle 4_d|p_2p_3|4_{\dot{d}}]F^3_{(123)} + (\sigma_i)\text{cyclic}\,,$$

where +cyclic indicates the sum over remaining three cyclic permutations, σ_i is the signature of each permutation, and $F^3_{(ijk)} \equiv F_i \wedge F_j \wedge F_k$. Explicitly evaluating the scalar integrals yield the parity-odd portion of the chiral fermion contribution to the four-gluon amplitude:

$$\begin{aligned} \frac{A^{\text{odd}}(1,2,3,4)}{F^{(4)}} &= \frac{(t-s)\left(\pi^2+\log[s/t]^2\right)}{12u^3} + \frac{\log[t/s]}{3u^2} \\ &\quad + \frac{s-t}{18stu} + R^{\text{odd}}_{(1,2,3,4)}\,, \end{aligned} \tag{3.11}$$

where again $R^{\text{odd}}_{(1,2,3,4)}$ represents the possible cut-free rational term. Firstly, note that the ultraviolet (UV) divergences explicitly canceled, just as the infrared divergence cancelled in four-dimensions.

[1]The on-shell form of the wedge product of three field strengths is simply

$$\begin{aligned} &F_1\wedge F_2\wedge F_3 \equiv \epsilon^{abcdef}F_{1ab}F_{2cd}F_{3ef} \\ &= \left(\langle 1_a|2_{\dot{b}}]\langle 2_b|3_{\dot{c}}]\langle 3_c|1_{\dot{a}}] + \langle 2_b|1_{\dot{a}}]\langle 1_a|3_{\dot{c}}]\langle 3_c|2_{\dot{b}}]\right)\,. \end{aligned}$$

The absence of UV-divergences in the parity-odd amplitude must hold, as there are no local operators available as viable counter-terms. Secondly, rational terms are already present in the cut-constructible anwser. This is a subtle difference from the previous $D = 4$ analysis, and only appears in higher-dimensions. The origin of this is due to the fact that while only scalar bubbles are UV-divergent in $D = 4$, for higher dimensions all $n \leq D/2$-gon scalar integrals are UV-divergent. The cancellation of UV-divergence then invariably leaves behind a rational term. For example in $D = 6$ the bubble- and triangle-integrals, in dimensional regularization, are given as,

$$\begin{aligned} I_3[K^2] &= \frac{1}{2\epsilon} + \frac{1}{2}\left(3 - \gamma_E - \log[K^2]\right), \\ I_2[K^2] &= -\frac{K^2}{6\epsilon} + \frac{K^2}{18}\left(-8 + 3\gamma_E + 3\log[K^2]\right), \end{aligned} \tag{3.12}$$

where K^2 is the unique kinematic invariant of the integral. As one can see, the cancellation of UV-divergences inevitably lead to a nontrivial rational term. Note that while the scalar integrals contain divergences and require regularization, the amplitude is finite and any result derived from the analysis of the amplitude will be scheme independent.

Again the ubiquitous presence of u-channel poles requires us to ensure that the residue of this pole, which is *spurious* for this ordering, must vanish to ensure locality. As $u \to 0$ one finds:

$$\left(Eq.\ (3.11) - R^{\rm odd}_{(1,2,3,4)}\right)\Big|_{u\to 0} = -\frac{1}{18tu} + \mathcal{O}(u^0)\,. \tag{3.13}$$

Note that although the u-channel is spurious in two of the six orderings, it is also present in the remaining four, due to the presence of rational terms arising from the cut-constructible part. This is the non-trivial consequence that was previously alluded to. It is straightforward to check that the leading $u \to 0$ behavior of all orderings are identical to Eq. (3.13), and thus for the full color-dressed amplitude, with $R = 0$, the leading $u \to 0$ behavior is given as:

$$\mathcal{A}|_{u\to 0} = -\frac{1}{18ut} sTr(1234) + \mathcal{O}(u^0) \tag{3.14}$$

where $sTr(1234)$ is the symmetric trace of the four generators. Thus for the absence of factorization poles one must have:

$$sTr(1234) = 0\,. \tag{3.15}$$

This reproduces the standard anomaly cancellation condition in six-dimensions. If Eq. (3.15) is not satisfied, then one must give a physical interpretation for this new factorization pole. Here, un-like D=4, the mass dimension of this residue is 4, implying that it can factorize into three-point functions that have mass-dimension 2. Again possible three-point amplitudes are highly constrained by Lorentz invariance, and it can be shown that the only possible dimension 2 amplitudes involving two vector fields is the three-point coupling of a graviton, a scalar, or a two-form to the vectors. Only the latter allows for parity odd-coupling (for details, see following chapter). In other words, in the event that Eq. (3.15) is not satisfied, the factorization pole implies the presence of a new particle in the spectrum: the two-form for the GS mechanism [46].

However this is not the end of the story, since in order for Eq. (3.14) to truly correspond to the singularity associated with the exchange of a two-form, the symmetric trace must factorize. This is one of the well known conditions for GS mechanism to apply. However, when the $sTr(1234)$ factorizes, it factorizes into three distinct double trace structure:

$$sTr(1234) \rightarrow tr(t_2t_4)tr(t_1t_3) + \text{cyclic}(123) \tag{3.16}$$

where t_i are the generators in the fundamental representation. Note that only the first term in Eq. (3.16) is consistent with an u-channel exchange, and the latter still represents inconsistent residues. Thus locality again demands us to add additional rational terms to cancel the inconsistent residues. Again, symmetry properties of the trace structure uniquely fixes the color dressed rational term to be:

$$\mathcal{R}^{\text{odd}} = F^{(4)}\left[tr(t_2t_4)tr(t_1t_3)\frac{t-s}{18stu} + \text{cyclic}(123)\right] . \tag{3.17}$$

Remarkably, $\mathcal{R}^{\text{odd}}$ is precisely the combination of the anomalous rational term of the Feynman-diagram calculation and the tree-diagram from GS mechanism. Using integral reduction on the Feynman-diagram loop-integral, one obtains the following parity-odd rational term:[2]

$$R^{anom}_{6\text{D}} = R^{anom}_{234} + \text{cyclic}(1234)\,, \tag{3.18}$$

where:

$$R^{anom}_{234} \quad = \quad -\frac{1}{18}\left(\frac{(\epsilon_1\cdot k_2)}{s}+\frac{(\epsilon_1\cdot k_3)}{u}+\frac{(\epsilon_1\cdot k_4)}{t}\right)F^3_{(234)}\,.$$

We did not present the expression for distinct orderings since Eq. (3.18) is manifestly permutation invariant, as expected. One can easily check that under a gauge transformation $\epsilon_i \rightarrow \epsilon_i + k_i$, Eq. (3.18)

[2]We use normalization such that the overall factor $1/(4\pi)^3$ is 1.

is anomalous. Working in the fundamental representation and combining with the contribution from the GS mechanism for $tr(t_1t_3)tr(t_2t_4)$, one finds the following gauge invariant combination for this color-factor :

$$-\frac{\left[F^3_{(234)}(tu(\epsilon_1\cdot k_2)+su(\epsilon_1\cdot k_4-2st(\epsilon_1\cdot k_3)+\text{cylic}\right]}{18stu}$$

Converting Eq. (3.19) to on-shell form one finds exactly that of Eq. (3.17)! epsilon

In conclusion, in applying unitarity methods to construct one-loop amplitudes, enforcing locality on chiral fermion loops gives rise to new factorization channels. In $D = 4$, such factorization channels lead to inconsistent residues, whose cancellation reproduces the anomaly cancellation conditions. In $D = 6$, the absence of spurious singularities requires either constraints which are precisely the well known anomaly cancellation conditions, or the introduction of rational terms to cancel the spurious singularities, leaving behind the physical factorization channels. The new channels then reflect the presence of a new particle in the spectrum: the GS two-form. Furthermore, this unique rational term is precisely the gauge invariant rational term that arrises form the combination of the parity-odd one-loop anomalous rational term and the contribution from the GS-mechanism, computed from Feynman rules. Thus starting with a chiral-gauge theory, imposing locality on the one-loop amplitude directly gives us the complete GS-contribution. In a sense there are no "gauge anomalies" per se. There are only rational terms in amplitudes, needed to enforce locality in amplitudes built from unitarity methods. When present, these rational terms make it impossible to have massless vectors in $D = 4$, whereas they force the existence of new degrees of freedom in higher-dimensions.[3] From this point of view, as rational terms only appear in even-dimensions at one-loop, and can appear in parity odd-amplitudes beginning at $n = D/2 - 1$-points, these are the places where such inconsistencies can arise in general, in agreement with the usual gauge-anomaly analysis. Finally, just as the lowest-multiplicity S-matrix can be uniquely determined, so can the parity-odd rational term as we have demonstrated. It will be interesting to see what kind of recursion one can set up to obtain all higher-multiplicity counterparts.

3.3 Acknowledgements

We are extremely grateful to Nima Arkani-Hamed for important discussions regarding the Green-Schwarz mechanism, Lance Dixon for the useful discussions on the role of rational terms, H. Johans-

[3] Note that rational terms also play a role in global anomalies, which has been recently discussed in the context of supergravity amplitudes [50].

son and Z. Bern for many fruitful discussions. This research was supported by the US DoE grant DE-SC0007859, and DE–FG03–91ER40662.

Chapter 4

Gauge anomalies on-shell (long)

4.1 Introduction and summary of results

Analyticity has long been known to impose stringent constraints on the S-matrix, and, from this, non-trivial predictions can be drawn for the underlying effective field theory. Classic examples include Coleman and Grossman's proof of 't Hoofts anomaly matching conditions [51], as well as Adams' et al. [52] proof of positivity for signs of the leading irrelevant operator. For the past decade, it has been understood that for a large class of gauge and gravity theories, systematic use of such constraints in the form of generalized unitarity methods [13, 12, 37] and recursion relations [20], one retains sufficient information to fully determine the amplitude iteratively from the fundamental building block: the three-point amplitude. For massless theories, with a given mass-dimension of the coupling constant the latter is in fact unique.

The complete irrelevance of the interaction Lagrangian in these developments has an appealing feature for the following reason: there are a plethora of interesting interacting theories for which no action is known. That this is so has been attributed to various reasons such as the absence of small expansion parameter, the inability to non-abelianize the underlying gauge symmetry of the asymptotic states, et cetera. However, it serves to remind ourselves that not too long ago such statements were widely used for the world-volume theory of multiple M2 branes, prior to the ascent of BLG [42] and ABJM [43] theories. Thus the removal of the "off-shell" yoke is welcoming. Now instead of asking if one can write down an interacting Lagrangian, we ask if one can construct a consistent S-matrix for the asymptotic states of the theory. Any difficulty encountered in the process will be independent of any off-shell formulation. Indeed it has been shown that for both BLG and

ABJM theories, global symmetries as well as factorization constraints is sufficient to completely determine the tree-level amplitude [44, 45].

However, gauge and gravitational anomalies also impose constraints on the quantum consistency of chirally coupled gauge and gravitational theories, and one might doubt whether the on-shell approach, being blind to any gauge symmetry, fails to capture such inconsistencies. The on-shell manifestation of such inconsistencies is usually stated as the non-decoupling of the longitudinal degrees of freedom. Unfortunately such discussion utilizes formal polarization (tensors) vectors and Feynman rules, for which an action is a prerequisite. Such discussion also appears to be incompatible with the modern on-shell program, where only the physical degrees of freedom are present throughout the construction. At this point, one might be tempted to conclude that unitarity methods are ill-fitted with chiral theories, and a completely on-shell construction is simply too much to ask for these theories. However, this cannot be the case, since if one were to blindly construct the quantum S-matrix using generalized unitarity methods, in absence of any sickness, one could simply declare that a consistent S-matrix is obtained, as the construction manifestly respects unitarity. One would then move on to conclude that there is no perturbative inconsistency for chiral gauge theories. Thus there must be a more physical manifestation of such sickness that does not rely on any notion of gauge symmetry or unphysical degrees of freedom.

Generalized unitarity utilizes the fact that the discontinuities of branch cuts are given by the product of lower loop-order amplitudes. Using sufficiently many cuts, one can fully determine the integrand for the cut-constructible part of the amplitude. Since by construction this method correctly reproduces all branch cuts, the amplitude manifestly respects unitarity. However, this does not guarantee locality. Here locality is defined by the property that the only singularity in the amplitude is associated with propagator-like singularities, $1/(p_i + p_j + \cdots + p_l)^2$. The unitarity methods determine the amplitude up to rational terms which are free of branch cuts.[1] However, such terms may have singularities, and its role is precisely to ensure that the full amplitude respects locality [54].

Not surprisingly, as discussed in the previous chapter which focused on chiral-gauge theories, the result obtained from unitarity methods violates locality, and is unrepairable unless certain conditions are met. More precisely, by enforcing locality on the cut-constructible result, the requisite rational term inevitably introduces new factorization channels. In four-dimensional Yang-Mills theory coupled to chiral fermions, the residue of this factorization channel cannot be interpreted as a product

[1] Although one can always embed the theory in higher dimensions and use the higher-dimensional unitarity cuts to fix such rational terms [30, 53], this cannot be done for chiral theories which have no higher-dimensional parent.

of tree-amplitudes: the factorization channel is therefore unphysical. Requiring the absence of such factorization-channels exactly leads to the well known anomaly cancellation condition, $d^{abc} = 0$! In six-dimensions, the same applies, except that now the residue of the new factorization channel can be interpreted as the product of two three-point amplitudes, each corresponding to a two-form coupling to two gauge fields. Thus one recovers the full anomaly cancellation conditions in six-dimensions: the requirement of locality requires either $str(T_1T_2T_3T_4) = 0$, or that it factorizes. For the latter, the factorization channel automatically introduces a new state to the spectrum, the two-form of the Green-Schwarz mechanism [46, 55]! Thus, in short, inconsistency of chiral-theories arises as the incompatibility of unitarity constraints and that of locality.

In this chapter, we expand on the analysis for six-dimensional chiral QCD, and extend it to chiral QED and gravity. For the gauge theories, while locality mandates a unique rational term which introduces a new factorization channel, the exchanged particle is identified to be a two-form only via dimensional and little-group analysis. Here we explicitly show that by gluing three-point amplitudes of a two-form coupled to two gauge-fields, one reproduces the exact residue that appeared in the factorization channel of the complete chiral-fermion loop-amplitude. This completes the proof that on-shell methods automatically include the requisite spectrum for a local and unitary quantum field theory, even if the initial spectrum is incomplete. Conversely, we also demonstrate that standing on its own, parity-odd rational terms cannot have consistent factorization in all channels, if any one of them involve the exchange of a two-form between vector fields. This is an on-shell manifestation of the well known fact that if the two-form couples to the gauge fields via $H^2 = (dB + AdA)^2$ and $B \wedge F \wedge F$, one cannot simultaneously maintain gauge invariance for both operators.

Remarkably, direct Feynman diagram calculation reveals that the rational term obtained from on-shell methods exactly matches that coming from the combination of the anomalous rational term from the loop-amplitude and that from the Green-Schwarz mechanism. It is interesting that whilst the anomaly is completely canceled, a remnant remains in the form of a parity-odd gauge-invariant rational term, whose presence is perhaps an afterthought in the traditional point of view. However, from the amplitude point of view, such rational terms are of utmost importance as they are mandated by locality, whilst anomaly cancellation is merely an "off-shell" mechanism to reproduce the requisite rational terms. As a side result, we also present the anomalous rational terms for chiral gauge theories in arbitrary even dimensions.

Chirally coupled gravity presents an interesting test case for our approach. In particular, in six-dimensions, no covariant action exists which couples (a generic number of) self-dual two-forms to gravity. In fact, in the original work of Alvarez-Gaume and Witten [56], this has been presented as an argument for the possibility of a gravitational anomaly. Here we demonstrate that generalized unitarity methods do define a consistent quantum S-matrix. To begin, we explicitly construct the parity-odd amplitude for the four-point one-loop amplitude, associated with chiral-matter. Consistency is demonstrated by showing that the conditions imposed by locality of the quantum S-matrix precisely match the gravitational anomaly cancellation conditions. Once these conditions are satisfied, we obtain the unitary and local gravity amplitude associated with chiral matter.

This chapter is organized as follows. In section 2, we review generalized unitarity, with an emphasis on the role of rational terms in enforcing locality. In section 3, we show that parity-odd loop-amplitudes constructed from generalized unitarity, generically requires rational terms that introduces new factorization channels, and thus potential violation of locality. Enforcing locality, in this setting, requires the same group theory constraints as the traditional anomaly cancellation. In section 4, we explicitly show, that the "new" poles which appear in *local* parity-odd loop amplitude in chiral gauge-theories must be identified with an exchange of a two-form with parity-violating couplings to gauge-fields. In section 5, we revisit anomaly cancellation from the point of view of Feynman diagrams, and demonstrate that the rational term obtained from the unitarity methods are in fact a combination of the anomalous rational terms from one-loop chiral fermion amplitude and the tree-level Green-Schwarz mechanism. Finally, in section 6, we make contact with the existence of, and recover cancellation conditions for, perturbative gravitational anomalies in six-dimensions.

4.2 Cut constructibility and the rational terms

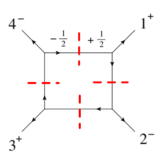
Generalized unitarity [13, 12, 37], by construction, furnishes a representation for the loop-integrand which reproduces the correct unitarity-cuts. Within the context of an integral basis, the perturbative structure of a particular S-matrix element is encoded in the coefficients of each integral. Depending on the nature of the theory, different integral bases manifest different properties. At one-loop, for general purposes it is convenient to use a basis of scalar integrals. Here, one-loop amplitudes in a generic D-dimensional massless theory—in this chapter D is always taken to be integer—decompose into:

$$A^{1-\text{Loop}} = \sum_{i_D} c_D^{i_D} I_D^i + \sum_{i_{D-1}} c_{D-1}^{i_{D-1}} I_{D-1} + \cdots + \sum_{i_{D-1}} c_2^{i_2} I_2 + R + \mathcal{O}(\epsilon)\,, \tag{4.1}$$

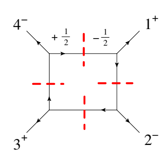
where $I_a^{i_a}$ represent scalar integrals with a propagators, diagrammatically an a-gon, and i_a labels the different distinct a-gons. The coefficient in front of each integral is to be fixed by unitarity cuts. For a review of why such an integral basis is sufficient for all one-loop amplitudes, we refer the reader to [27] for a brief discussion. The last term R indicates a possible rational function that is cut-free.

An important subtlety is whether the unitarity cuts are defined in D or $D-2\epsilon$-dimensions. Since the scalar loop-integrals in Eq. (4.1) do not contain any loop momentum dependence in the numerator, their coefficients are blind to whether or not the unitarity cut is defined in D or $D-2\epsilon$-dimensions. On the other hand, the rational terms R, can be obtained via $D-2\epsilon$-dimensional unitarity cuts [30, 53], which requires tree-amplitudes within the unitarity-cut to be analytically continued into higher dimensions. While this approach for obtaining the rational terms is widely applied in QCD, it becomes problematic for chiral theories or theories with Chern-Simons terms. Finally, to properly extend the theory to higher dimensions, one requires the knowledge of the explicit action, which is antonymous to the program of the on-shell S-matrix.

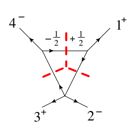
It is then useful to take a step back and ask: what physical principle mandates the existence of these rational terms? Specifically, given that the result in Eq. (4.1), absent the rational terms R, is manifestly unitary in all channels, in what sense are the cut-constructed terms without their rational counterparts incorrect? To answer this question, it is useful to consider an explicit example. Let us consider a fundamental fermion contribution to the four-point one-loop gluon amplitude in four-dimensions. Applying standard unitary cut methods [19], the integral coefficients for $A^{1-\text{Loop}}(1^+2^-3^+4^-)$ from this fermion loop:



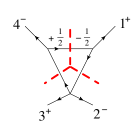
$$c_4 = -\frac{t^4 s^2}{u^4}$$



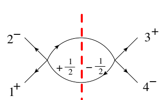
$$c_4 = -\frac{s^4 t^2}{u^4}$$



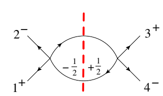
$$c_3^{(23)} = \frac{t^4 s}{u^4}$$



$$c_3^{(23)} = \frac{t^2 s^3}{u^4} \tag{4.2}$$



$$c_2^{(12)} = \frac{t(2t^2 - s^2 - 5st)}{6u^3}$$



$$c_2^{(12)} = \frac{t(2t^2 + 11s^2 + 7st)}{6u^3}$$

where the superscript in $c^{(ij)}$ denotes the legs on the massive corners. Note: for future convenience, the contributions to the integral coefficients from each of the two different fermion helicities circling in the loop have been separated. This separation facilitates finding the even- and odd-parity components of the loop-amplitude. The other orderings can be obtained by cyclic rotation. Combined with the integrated scalar integrals, we find the following fermion contribution to the $A^{1-\text{L}}(1^+2^-3^+4^-)$

amplitude :

$$A^{1-\mathrm{L,even}}_{1/2,\,Cut}(1^+2^-3^+4^-) \quad = \quad A^{\mathrm{tree}}(1^+2^-3^+4^-)\Bigg[\frac{-(-s)^{-\epsilon}-(-t)^{-\epsilon}}{3\epsilon} - \frac{st(s^2+t^2)}{2u^4}(\log^2 x+\pi^2)$$
$$-\frac{(s-t)(s^2-st+t^2)}{3u^3}\log x\Bigg]\,, \tag{4.3}$$

where $x \equiv t/s$, the subscript in $A_{1/2}$ indicates it is a fermion loop contribution, Cut indicates it is the result derived from purely four-dimensional unitarity cuts, and "even" indicates we have taken the parity-even combination of the various unitarity-cuts.[2]

We would now like to ask, what precisely is wrong with Eq. (4.3)? The branch-cut structure is guaranteed to be correct by construction, and thus unitarity is not violated. However, the amplitude not only has to satisfy unitarity, but locality as well. Locality requires that the loop amplitude can only have propagator singularities, and that the residue on these propagator-poles must be given by the product of lower-order amplitudes (lower-order means reduced number of legs and reduced loop-order):

$$A^{\ell}_n|_{K^2\to 0} = \sum_{\ell_1,\ell_2} A^{\ell_1}_{n+1-k}\frac{1}{K^2}A^{\ell_2}_{k+1} + \mathcal{O}((K^2)^0)\,, \tag{4.4}$$

where ℓ, ℓ_1, and ℓ_2 denote loop-order and are subject to the constraint $\ell = \ell_1+\ell_2$.

Going back to Eq. (4.3) we see that $1/u^n$ factors are ubiquitous in each term, and thus provide potential violations of locality. Expanding Eq. (4.3) around $u=0$ one finds:

$$A^{1-\mathrm{L,even}}_{1/2,\,Cut}(1^+2^-3^+4^-)\Big|_{u\to 0} = A^{\mathrm{tree}}(1^+2^-3^+4^-)\left(-\frac{s^2}{u^2}-\frac{s}{u}+\mathcal{O}(u^0)\right)\,. \tag{4.5}$$

Since the spurious poles have rational residues, the result in Eq. (4.3) violates locality! Given that the part of the amplitude containing cuts is completely determined, the only way this may be remedied is through addition of a cut-free term to the amplitude to cancel this singularity: a rational term. Dimensional analysis and cyclic symmetry entirely fix the rational term needed to correct this $1/u^2$-singularity to be,

$$R^{\mathrm{even}}(1^+2^-3^+4^-) = -A^{\mathrm{tree}}(1^+2^-3^+4^-)\frac{st}{u^2}\,, \tag{4.6}$$

Note that, since this is a color-ordered amplitude, a $1/u$ pole with is unphysical. Luckily R^{even}

[2]Notice that the $-2/3$ in front of the $1/\epsilon$ UV-divergence is precisely the minus of the one-loop beta function, as expected. See chapter 2 and references therein for more details.

automatically removes the $1/u$ pole as well. Thus the unique unitary and local answer is:

$$A_{1/2}^{1-\mathrm{L,even}} = A_{1/2,\,Cut}^{1-\mathrm{Loop,even}} + R^{\mathrm{even}}\,, \tag{4.7}$$

reproducing the correct answer obtained, long ago, in [47] .

From the above discussion we see that the role of rational terms is to correct any non-locality introduced into the results from D-dimensional unitarity cuts. Note that this implies that the rational terms must have certain factorization properties, which allow it to be obtained through recursion relations, as demonstrated in [57, 58, 59]. One might wonder if it is possible to add a term that is simply a constant times the tree-amplitude. This is not an acceptable modification to the loop amplitude: the tree-amplitude contains s- and t-factorization channels, and the presence of such a term would lead to a one-loop modification to the three-point on shell amplitudes, which are not supposed to be corrected in perturbation theory. Note that R^{even} avoids this problem as its numerator is proportional to st, which vanishes on either of these two factorization channels. This subtlety will rear its head when we consider chiral fermions, and is the avatar of four-dimensional anomalies as we discuss in the next section.

4.3 Anomalies as spurious poles

In the section, we will repeat the process discussed above for chiral gauge theories. That is, we will construct parity-odd contributions to unitarity cuts of gauge-theory amplitudes due to chiral fermions, in four- and six-dimensions. We will study the pole structure of the unitarity cut result, and determine whether rational terms are needed to insure locality. As we will see, there is a subtlety in the prerequisite rational terms that plants the seed for the inconsistency of chirally coupled gauge and, in a later section, gravity theories coupled to chiral matter.

4.3.1 D=4 QCD

Let us now revisit the same scattering process fermion-loop, and now focus on the parity-odd configuration, where we take the *the difference* of the two helicity configurations in each unitarity cut. Note that this part of the amplitude is only present in chiral theories. The result is:

$$A_{1/2,\,Cut}^{1-\mathrm{L,odd}}(1^+2^-3^+4^-) = A^{\mathrm{tree}}(1^+2^-3^+4^-)\left[-\frac{st(s-t)}{2u^3}(\log^2 x+\pi^2) - \left(\frac{2st}{u^2}\right)\log x\right].$$

Note that the above result is only cyclic invariant up to a sign. This is because we are using helicity basis in combination with the Levi-Cevita tensor. Indeed one can identify:

$$\begin{aligned}
A^{\text{tree}}(1^+2^-3^+4^-)(s-t) &= -2u\left[(F_1\wedge F_3)\left(\frac{\epsilon_4\cdot k_2}{u}-\frac{\epsilon_4\cdot k_3}{s}\right)\left(\frac{\epsilon_2\cdot k_4}{u}-\frac{\epsilon_2\cdot k_1}{s}\right)\right.\\
&+ \left.(F_2\wedge F_4)\left(\frac{\epsilon_1\cdot k_3}{u}-\frac{\epsilon_1\cdot k_4}{t}\right)\left(\frac{\epsilon_3\cdot k_1}{u}-\frac{\epsilon_3\cdot k_2}{t}\right)\right]+cyclic\,,
\end{aligned}$$

which reveals the cyclic structure outside of a particular helicity assignment. To see if rational terms are needed, we look at the residue of the apparent spurious u-pole:

$$A^{\text{1-L,odd}}_{1/2,\,Cut}(1^+2^-3^+4^-)|_{u\to 0} = A^{\text{tree}}(1^+2^-3^+4^-)\left(-\frac{s}{u}+\mathcal{O}(u^0)\right)\,. \tag{4.8}$$

Locality again requires the absence of such spurious poles through addition of a rational term. Taking into account the fact that the amplitude attains a minus sign under cyclic shift, the requisite parity-odd rational term is:

$$R^{\,\text{odd}}(1^+2^-3^+4^-) = A^{\text{tree}}(1^+2^-3^+4^-)\frac{s-t}{2u} = \langle 24\rangle^2[13]^2\frac{s-t}{2stu}\,. \tag{4.9}$$

Naively the following sum is fully unitary and local:

$$A^{\text{1-L,odd}}_{1/2,\,Cut}+R^{\,\text{odd}}\,. \tag{4.10}$$

However, as is plain from Eq. (4.9), this new parity-odd rational term introduces non-trivial contributions to the s- and t-channel residues. This contrasts sharply with the $R^{\,\text{even}}$. Herein lay the seeds of inconsistencies in parity-violating gauge theories: the rational terms, required for locality in the parity-odd amplitude, introduce *new* corrections to residues on the s- and t- poles. Since the dimension of the residue is 2, for Eq. (4.4) to be respected the residue must be given by either (a) the product of two mass-dimension 1 three-point amplitudes, or (b) a product of mass-dimension zero and a mass-dimension two tree amplitude. Let us consider each case in detail:

- (1,1): We consider possible mass-dimension one three-point amplitudes with two vectors of opposite helicity and one unknown particle specie $A_3(g^+g^-*)$. A simple analysis of helicity constraint tells us that the unknown particle can only be a vector. Thus the only three-point amplitude allowed is the MHV, or $\overline{\text{MHV}}$, amplitude which does not have one-loop corrections.

- (2,0): Note that there is a mass-dimension two amplitude involving two gluons and one scalar, generated by the operator ϕF^2. However there are no mass-dimension zero amplitudes involving two gluons, as is discussed in the following chapter.

Thus from the above analysis we conclude that Eq. (4.10) does not respect locality, as incarnated in Eq. (4.4): theories which chirally couple fermions to gauge bosons have an inherent tension between locality and unitarity!

However, there is a potential way out. Since the fully color-dressed amplitude is given by a sum of single-trace amplitudes (we will assume fundamental fermions for now), it is possible that the new factorization poles introduced by the rational terms actually cancel in the fully dressed amplitude. There are six distinct single trace structures, and the rational term in the fully color-dressed amplitude is given by:

$$\begin{aligned}\mathcal{R} \quad = \quad & \frac{\langle 24\rangle^2[13]^2}{2stu}[(s-t)Tr(1234)+(u-s)Tr(1342)+(t-u)Tr(1423) \\ & +(s-u)Tr(1243)+(u-t)Tr(1324)+(t-s)Tr(1432)]\,. \end{aligned} \tag{4.11}$$

Let us consider the residue for the $s\to 0$. The contribution from $Tr(1234)$, $Tr(1342)$, $Tr(1432)$ and $Tr(1243)$ sum to:

$$\frac{\langle 24\rangle^2[13]^2}{2su}[-Tr(1234)-Tr(1342)+Tr(1243)+Tr(1432)]\,. \tag{4.12}$$

Note that we did not include the contribution from $Tr(1423)$ and $Tr(1324)$, since the s-channel pole in these two trace structures by construction cancels against the s-pole from the cut-constructed part of the amplitude. Thus if the group theory structure in Eq. (4.12) vanishes, one obtains a fully consistent amplitude. This constraint on the group theory factor can be rewritten as:

$$Tr(1432)-Tr(1234)+(1\leftrightarrow 2)=d^{1a4}f^{23}{}_{a}+d^{13a}f^{24}{}_{a}+d^{1a2}f^{34}{}_{a}+(1\leftrightarrow 2)=0\,. \tag{4.13}$$

Since the symmetry property of each term is distinct, the constraint is satisfied only if each term is individually zero. Thus imposing unitarity and locality, enforces the group theory factor constraint,

$$d^{abc}f^{de}{}_{a}=0\,, \tag{4.14}$$

which is nothing but the anomaly cancellation condition for the non-abelian box anomaly!

Let us pause and see what we have achieved. Starting with tree-amplitudes, generalized unitarity constructs putative loop-amplitudes that are manifestly unitary. However, they fail to be local. Imposing locality forces inclusion of rational terms to cancel the spurious singularities within these putative loop-amplitudes. Here the chiral theory differs from the non-chiral theories: the requisite rational terms introduce "new" factorization poles with non-trivial residues. In non-chiral theories, such factorization channels are simply absent. Because there are no suitable tree-amplitudes which it can factorize into, the presence of these residues are simply inconsistent, and therefore must vanish. The only way a chiral theory can achieve this vanishing is with the aid of the group theory factors. This leads precisely to the non-abelian anomaly cancellation condition. Nowhere in the discussion did we invoke or mention of gauge symmetry. Furthermore, as $A^{1-\mathrm{L,odd}}_{1/2,\,Cut}$ is both IR and UV-finite, it does not have any regulator dependence. Thus the amplitude presents a *manifestly* regulator independent representation of the inconsistencies of chiral theories.

We now move on to higher dimensional chiral gauge and gravity theories, where the story becomes much more interesting.

4.3.2 D=6 QED

We now move on to six-dimensions, and consider the simplest chirally coupled system, chiral QED. Using the six-dimensional spinor helicity formalism introduced by Cheung and O'Connel [48], the four-point tree-amplitude of two photons and two chiral-fermions is,

$$A_4(\gamma_{a_1\dot{a}_1}\gamma_{a_2\dot{a}_2}e_{a_3}e_{a_4}) = \frac{\langle 1_{a_1}2_{a_2}3_{a_3}4_{a_4}\rangle[1_{\dot{a}_1}|3|2_{\dot{a}_2}]}{tu}\,, \tag{4.15}$$

where $(a_i,\dot{a}_i)$ are the SU(2)$\times$ SU(2) little group indices for each external leg. To extract the parity-odd piece contribution from the chiral fermions to the loop, we again take the difference between the chiral and anti-chiral fermion contributions to the loop. For this purpose we will need the amplitude for chiral fermions as well, which is given by:

$$A_4(\gamma_{a_1\dot{a}_1}\gamma_{a_2\dot{a}_2}e_{\dot{a}_3}e_{\dot{a}_4}) = \frac{[1_{\dot{a}_1}2_{\dot{a}_2}3_{\dot{a}_3}4_{\dot{a}_4}]\langle 1_{a_1}|3|2_{a_2}\rangle}{tu}\,. \tag{4.16}$$

We will follow [49] (see also [60]) to apply generalized unitarity method in six-dimensions. The chiral integral coefficients, defined as the difference between chiral and anti-chiral fermion loop, are:

$$C_4(1,2,3,4) = \frac{(s-t)}{3u^2}F^{(4)} \quad C_4(1,3,4,2) = \frac{(u-s)}{3t^2}F^{(4)}, \quad C_4(1,4,2,3) = \frac{(t-u)}{3s^2}F^{(4)}\,, \tag{4.17}$$

$$C_{3s} = \frac{4}{3}\left[\frac{1}{u^2}-\frac{1}{t^2}\right]F^{(4)}\,, \quad C_{3t} = \frac{4}{3}\left[\frac{1}{s^2}-\frac{1}{u^2}\right]F^{(4)}\,, \quad C_{3u} = \frac{4}{3}\left[\frac{1}{t^2}-\frac{1}{s^2}\right]F^{(4)}\,, \tag{4.18}$$

where $C_4(1,2,3,4)$ represent the coefficient for the box integral with ordered legs 1,2,3,4, and $C_{3s}(C_{3t})$ represent one-mass s-channel (t-channel) triangle coefficients. The function $F^{(4)}$ is,

$$F^{(4)} \equiv \frac{1}{2}\big(\langle 4_d|p_2 p_3|4_{\dot{d}}]F_1 \wedge F_2 \wedge F_3 + (\sigma_i)\text{cyclic}\big)\,, \tag{4.19}$$

where $\sigma_i = -1$ for odd cyclic permutations and $+1$ for even. The on-shell form of the wedge product of three field strengths is simply[3]

$$F_1 \wedge F_2 \wedge F_3 = \big(\langle 1_a|2_{\dot{b}}]\langle 2_b|3_{\dot{c}}]\langle 3_c|1_{\dot{a}}] + \langle 2_b|1_{\dot{a}}]\langle 1_a|3_{\dot{c}}]\langle 3_c|2_{\dot{b}}]\big)\ . \tag{4.20}$$

Thus the cut-constructed part of the amplitude is given by

$$A^{1-\text{L,odd}}_{1/2,\,Cut} \;=\; \sum_{i,j,k,l\in S_4} C_4(i,j,k,l)I_4(i,j,k,l) + C_{3s}I_{3s} + C_{3t}I_{3t} + C_{3u}I_{3u}\,. \tag{4.21}$$

Explicitly, the massless-box, one-mass triangle, and (massive) bubble scalar integrals, in $(6-2\epsilon)$-dimnesions, evaluate to:

$$\begin{aligned}
I_3[K^2] &= \frac{1}{2\epsilon} + \frac{1}{2}\left(3 - \gamma_E - \log[K^2]\right)\,,\\
I_2[K^2] &= -\frac{K^2}{6\epsilon} + \frac{K^2}{18}\left(-8 + 3\gamma_E + 3\log[K^2]\right)\,,\\
I_4(1,2,3,4) &= -\frac{\log^2 x + \pi^2}{2u}\,.
\end{aligned} \tag{4.22}$$

Combining Eq. (4.17), Eq. (4.21), and Eq. (4.22), it is easy to see that again there are no UV-divergences in $A^{1-\text{L,odd}}_{1/2,\,Cut}$, as expected. Note in chiral QED, there is no color-ordering and all propagator poles are physical. But to ensure locality, again the residue of any factorization poles must correspond to the product of lower-point amplitudes. In the vicinity of the $t \to 0$ pole, the cut-constructed terms approach

$$A^{1-\text{L,odd}}_{\frac{1}{2},\,Cut}|_{t\to 0} = \frac{F^{(4)}}{tu} + \mathcal{O}(t^0)\,. \tag{4.23}$$

We see the presence of a factorization pole with mass-dimension 4 residue.

We now analyze all possible three-point amplitudes involving at least two vectors with mass-dimension 1,2 and 3.[4] As usual, three-point kinematics are degenerate, and one needs to take special

[3]The sign between the two terms can be determined from the fact that the parity-odd configuration is symmetric under the exchange of any two labels, while the parity-even configuration is anti-symmetric.

[4]We don't consider mass-dimension 0 or 4, since it is easy to see a vector cannot couple with a mass-dimension 0 coupling. The absence of mass-dimension 0 three-point amplitude makes any possible mass-dimension 4 amplitude irrelevant for the purpose of constructing an acceptable residue.

care in defining the on-shell elements. In [48], special unconstrained variables were introduced to parameterize the kinematics, which carry redundant degrees of freedom. The "gauge invariance" requirement is then the manifestation of Lorentz invariance for the three-point on-shell amplitudes. A systematic study of all such possible three-point amplitudes involving massless fields were presented in [61], facilitates the following analysis:

- (3,1): As with four-dimensions, mass-dimension one three-point amplitudes involving two vectors necessarily imply the third particle is also a vector, and the three-vector amplitude is unique, coming from F^2. This then restricts the mass-dimension three-amplitude to be of three vectors as well, which is also unique, coming form an $F_\mu{}^\nu F_\nu{}^\rho F_\rho{}^\mu$ operator. Since the two operators are parity even, they cannot contribute to a parity odd-amplitude.

- (2,2): There are three possible mass-dimension two three-point amplitudes involving two vectors: coupling to a scalar, a graviton and a two-form. The former two correspond to the operator ϕF^2 and $\sqrt{g}F^2$, which also does not introduce any parity odd contribution. The third possibility is a two-form $B_{\mu\nu}$, which has two possible couplings, H^2 and $B\wedge F\wedge F$, where $H_{\mu\nu\rho}\equiv\partial_{[\mu}B_{\nu\rho]}+A_{[\mu}\partial_\nu A_{\rho]}$ is the three-form field strength. Note that the BF^2 operator indeed introduces a parity-odd contribution.

Thus we see that the only way for the factorization pole in Eq. (4.23) to have a reasonable residue is to understand it as the propagation of a two-form. In other words, by forcing the unitarity result to also satisfy locality, the scattering amplitude "responds" by telling us there is a new particle in the spectrum, the two-form of the Green-Schwarz mechanism [46]! Note that we again see $A^{1-\mathrm{L,odd}}_{\frac{1}{2},Cut}$ is completely finite, and thus this discussion is manifestly independent of any regulator.

Later, we will study the residue more closely to see that it is indeed given by the gluing of the H^2 and $B\wedge F\wedge F$ interactions. But first we consider chiral-fermions coupled to Yang-Mills, i.e. $6D$ chiral QCD.

4.3.3 D=6 QCD

We now consider a chiral fermion coupled to Yang-Mills theory. The relevant four-point amplitudes for two gluon and two chiral (anti-chiral) fermions are:

$$\text{Chiral}:\quad A_4(g_{a_1\dot a_1}g_{a_2\dot a_2}q_{a_3}q_{a_4})=\frac{\langle 1_{a_1}2_{a_2}3_{a_3}4_{a_4}\rangle[1_{\dot a_1}|3|2_{\dot a_2}]}{ts}\,, \tag{4.24}$$

$$\text{Anti}-\text{chiral}:\quad A_4(g_{a_1\dot a_1}g_{a_2\dot a_2}q_{\dot a_3}q_{\dot a_4})=\frac{[1_{\dot a_1}2_{\dot a_2}3_{\dot a_3}4_{\dot a_4}]\langle 1_{a_1}|3|2_{a_2}\rangle}{ts}\,. \tag{4.25}$$

The integral coefficients now become

$$
\begin{aligned}
C_4 &= \frac{(s-t)}{6u^2}F^{(4)}\,,\quad C_{3s}=-\frac{(s-t)}{6tu^2}F^{(4)}\,,\\
C_{3t} &= -\frac{(s-t)}{6su^2}F^{(4)}\,,\quad C_{2s}=\frac{F^{(4)}}{stu}\,,\quad C_{2t}=-\frac{F^{(4)}}{stu}\,,
\end{aligned}
\tag{4.26}
$$

where $F^{(4)}$ is defined as before. Note that the bubble and triangle coefficients are such that there is no overall UV-divergence. Indeed one funds that the integrated result for the cut-constructible piece is given by:

$$
A^{1-\mathrm{L,odd}}_{\frac{1}{2},Cut}=F^{(4)}\left[\frac{(t-s)(\pi^2+\log^2[x])}{12u^3}+\frac{\log[x]}{3u^2}+\frac{s-t}{18stu}\right]\,. \tag{4.27}
$$

Again the ubiquitous u-poles in the planar amplitude requires us to inquire if it has a rational residue. One finds:

$$
A^{1-\mathrm{L,odd}}_{\frac{1}{2},Cut}\Big|_{u\to 0}=-\frac{F^{(4)}}{18tu}+\mathcal{O}(u^0)\,. \tag{4.28}
$$

Note that, un-like in four-dimensions where we can immediately conclude that one needs to include rational term, here we need to consider the fact that the same pole will appear in other color-orderings. In other words the presence of this pole may cancel between different color-orderings, just like the physical poles introduced by the rational term in 4D chiral QCD can cancel among different color-orderings, as discussed in sec. 4.3.1. Indeed the same pole also appears in other color-ordering. Analyzing the behavior of the cut constructed color-dressed amplitude, one finds the following collinear behavior in the s-,t- and u-channels:

$$
\begin{aligned}
\lim_{u=0}\mathcal{A}^{1-\mathrm{L,odd}}_{\frac{1}{2},Cut} &\to -\frac{F^{(4)}}{18ut}str(T^4)+\mathcal{O}(u^0)\,,\quad \lim_{s=0}\mathcal{A}^{1-\mathrm{L,odd}}_{\frac{1}{2},Cut}\to -\frac{F^{(4)}}{18su}str(T^4)+\mathcal{O}(s^0)\,,\\
\lim_{t=0}\mathcal{A}^{1-\mathrm{L,odd}}_{\frac{1}{2},Cut} &\to -\frac{F^{(4)}}{18ts}str(T^4)+\mathcal{O}(t^0)\,.
\end{aligned}
\tag{4.29}
$$

Here $str(T^4)$ represents the symmetric trace of four generators. Thus we see that the collinear factorization pole will not contribute, if the group theory structure satisfies:

$$
str(T^4)=0\,. \tag{4.30}
$$

This is nothing but the anomaly free condition in six-dimensions. Note that the reason why spurious poles of the cut-constructed answer can cancel between different ordering can attributed to the fact that in six-dimensions, the cut-constructed answer actually already contains a rational term, whereas such terms are not present in four-dimensions. The distinction is that in six-dimensions there are two scalar integrals in Eq. (4.22) that contain UV-divergences. The relative coefficients between the $1/\epsilon$-divergent pieces and finite pieces are different between the two integrals. Thus if the final answer is UV-finite, then when combined with cyclic symmetry one can conclude there must be a left-over rational piece already present in the cut-constructible terms. In $D = 4$ there is only one UV-divergent scalar integral, the bubble. Thus if the UV-divergence cancels in $D = 4$, so will the rational term in the scalar bubble.

On the other hand, if $str(T^4) \neq 0$, in $D = 6$ chiral QCD, the cut-constructible part has non-trivial rational residues on each factorization channel, and locality then requires each such residue to correspond to the exchange of a particle. From the analysis of QED, we know that this pole can only be interpreted as an exchange of a GS two-form. For this interpretation to be true the group theory factor must factorize, i.e. we must have $str(T^4) \sim tr(t^2)tr(t^2)$ for some representation t. This is the well known story that when the generators are rewritten in fundamental representation, for a general gauge group one has:

$$str(T^4) = \alpha\, str(t^4) + \beta\, tr(t^2)tr(t^2)\,. \qquad (4.31)$$

For the poles to have a physical interpretation we must have $\alpha = 0$.

However even with $\alpha = 0$, and the symmetric trace factorizes, there is still a problem. Eq. (4.29) becomes,

$$\lim_{u=0} \mathcal{A}^{1-\mathrm{L,O}}_{\frac{1}{2},Cut} \;\rightarrow\; -\frac{F^{(4)}}{18ut}\left[tr(t_1t_2)(t_3t_4) + tr(t_1t_3)(t_2t_4) + tr(t_1t_4)(t_3t_2)) + \mathcal{O}(u^0)\right]. \qquad (4.32)$$

Clearly, only the group theory factor $tr(t_1t_3)(t_2t_4)$ makes any sense as a factorization channel for the u- channel pole. In other words, this u-pole reside should not have any term proportional to $tr(t_1t_2)(t_3t_4)$ or $tr(t_1t_4)(t_2t_3)$. Similar conclusions can be reached from the s- and t-pole analysis. Thus for the poles to have an interpretation as a proper factorization channel, extra rational terms must be added to the cut-constructible terms. From symmetry properties required from the trace structure, we can deduce that the requisite rational term is:

$$\mathcal{R} = F^{(4)}\left[tr(t_1t_2)(t_3t_4)\frac{u-t}{18stu} + tr(t_1t_3)(t_2t_4)\frac{t-s}{18stu} + tr(t_1t_4)(t_3t_2)\frac{s-u}{18stu}\right]. \qquad (4.33)$$

Indeed we now find that with the above rational term, we now have:

$$\lim_{u=0} \mathcal{A}^{1-\mathrm{L,O}}_{\frac{1}{2},\,Cut} + \mathcal{R} \quad \rightarrow \quad -\frac{F^{(4)}}{6ut} tr(t_1 t_3)(t_2 t_4) + \mathcal{O}(u^0) \tag{4.34}$$

$$\lim_{s=0} \mathcal{A}^{1-\mathrm{L,O}}_{\frac{1}{2},\,Cut} + \mathcal{R} \quad \rightarrow \quad -\frac{F^{(4)}}{6su} tr(t_1 t_2)(t_3 t_4) + \mathcal{O}(s^0) \tag{4.35}$$

$$\lim_{t=0} \mathcal{A}^{1-\mathrm{L,O}}_{\frac{1}{2},\,Cut} + \mathcal{R} \quad \rightarrow \quad -\frac{F^{(4)}}{6ts} tr(t_1 t_4)(t_3 t_2) + \mathcal{O}(t^0)\,. \tag{4.36}$$

In summary, we have found that for chiral fermions coupled to Yang-Mills in six-dimensions, the answer obtained from unitarity methods can be local only if $str(T^4) = 0$ or $str(T^4) = \beta\, tr(t^2)tr(t^2)$, for some representation t. Furthermore, for the case that $str(T^4) \neq 0$, locality requires one to add a rational term $\mathcal{R}$ given in Eq. (4.33). With this additional term one obtains a unitary and local answer. One might wonder what is this magical rational term $\mathcal{R}$? Since it is constructed from on-shell methods, it is gauge invariant. In sec 4.5, we will show that, in terms of Feynman diagrams, this is nothing but the gauge invariant combination of the anomalous rational term plus the tree-diagrams in the GS mechanism! But, before moving on to Feynman diagrams, we analyze whether or not the residue is indeed given by the gluing of two 3-point amplitudes involving the exchange of a two-form.

4.4 Factorization properties of the rational term

In this section we study the residue of the single pole that appeared both in the chiral QED, in Eq. (4.23), and chiral QCD, in Eq. (4.34). They are all proportional to $F^{(4)}$ defined in 4.19, divided by the Mandlestam variable of the non-factorizing channel. Dimensional and little group analysis tell us that the residues on these poles can only correspond to propagation of a two-form. However we have not explicitly provided the three-point amplitudes whose gluing should in principle give the this residue. Three-point elements are highly constrained, and not surprisingly, the three-point amplitude for two vector fields and a two-form is unique [61]. In this section, we will show that the parity odd part of the two-form exchange indeed gives the residue in Eq. (4.23) and Eq. (4.34).

4.4.1 Review on six-dimensional On-shell variables for three-point kinematics

The possible three-point amplitude in six-dimensions have been studied extensively and is severely constrained using the following SU(2) spinors [48]:

$$\begin{aligned}
\langle 1_a|2_{\dot a}] &\equiv u_{1a}\tilde u_{2\dot a}\,, \quad \langle 2_a|1_{\dot a}] \equiv -u_{2a}\tilde u_{1\dot a}\,,\\
\langle 2_a|3_{\dot a}] &\equiv u_{2a}\tilde u_{3\dot a}\,, \quad \langle 3_a|2_{\dot a}] \equiv -u_{3a}\tilde u_{2\dot a}\,,\\
\langle 3_a|1_{\dot a}] &\equiv u_{3a}\tilde u_{1\dot a}, \quad \langle 1_a|3_{\dot a}] \equiv -u_{1a}\tilde u_{3\dot a}\,,
\end{aligned} \tag{4.37}$$

as well as their pseudo inverse:

$$u_{ia}w_{ib} - u_{ib}w_{ia} = \epsilon_{ab}, \quad \tilde u_{i\dot a}\tilde w_{i\dot b} - \tilde u_{i\dot b}\tilde w_{i\dot a} = \epsilon_{\dot a\dot b}\,. \tag{4.38}$$

Note that the above definitions are invariant under the following rescaling and shift symmetries:

$$(u_i, w_i) \to (\alpha u_i, \alpha^{-1} w_i), \qquad (\tilde u_i, \tilde w_i) \to (\alpha^{-1}\tilde u_i, \alpha\tilde w_i), \tag{4.39}$$

$$w_i \to w_i + b_i u_i, \qquad \tilde w_i \to \tilde w_i + \tilde b_i \tilde u_i\,. \tag{4.40}$$

Repeated indices are not summed. Since these symmetries arise from their definition with respect to Lorentz invariants, preservation of these symmetries is equivalent to Lorentz invariance. Thus all possible Lorentz invariant three-point amplitudes must be a polynomial of these objects satisfying the rescaling and shift symmetry. Momentum conservation further implies

$$u_{ia}|i^a\rangle = u_{ja}|j^a\rangle, \quad \tilde u_{i\dot a}|i^{\dot a}] = \tilde u_{j\dot a}|j^{\dot a}], \quad \sum_i w_{ia}|i^a\rangle = \sum_j \tilde w_{j\dot a}|j^{\dot a}] = 0\,. \tag{4.41}$$

Again repeated indices are not summed, and when combined with Eq. (4.39) implies that $b_1+b_2+b_3 = 0$, $\tilde b_1 + \tilde b_2 + \tilde b_3 = 0$.

The three-point amplitudes involving one (anti-) chiral two-form and two vector fields are uniquely fixed by Lorentz invariance and little group constraint [61]:

$$A^B_{(ad),b\dot b,c\dot c} = \Big(\Delta_{(a|bc}u_{1|d)}\Big)\tilde u_{2\dot b}\tilde u_{3\dot c}\,, \;\; A^{\bar B}_{(\dot a\dot d),b\dot b,c\dot c} = u_{2b}u_{3c}\Big(\tilde\Delta_{(\dot a|\dot b\dot c}\tilde u_{1|\dot d)}\Big)\,,$$

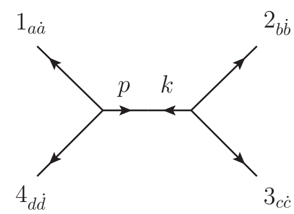
where

$$\begin{aligned}\Delta_{abc} &= (u_{1a}u_{2b}w_{3c} + u_{2b}u_{3c}w_{1a} + u_{3c}u_{1a}w_{2b})\,,\\ \tilde{\Delta}_{\dot a\dot b\dot c} &= (\tilde{u}_{1\dot a}\tilde{u}_{2\dot b}\tilde{w}_{3\dot c} + \tilde{u}_{2\dot b}\tilde{u}_{3\dot c}\tilde{w}_{1\dot a} + \tilde{u}_{3\dot c}\tilde{u}_{1\dot a}\tilde{w}_{2\dot b})\,.\end{aligned}\tag{4.42}$$

The little group indices tells us that the two vector fields sit on legs 2 and 3, whilst the field on leg 1 transform as a (3,0) and (0,3) in A^B and $A^{\bar{B}}$ respectively. Because the six on-shell components of a two-form can be split into self-dual and anti-self dual sectors, A^B and $A^{\bar{B}}$ represents the couplings of self-dual anti-self-dual two-forms, respectively.

4.4.2 Gluing (A^B, A^B) and $(A^{\bar{B}}, A^{\bar{B}})$

We now consider the gluings of two A^Bs, as well as of $A^{\bar{B}}$s. Taking the difference of these two products yields the parity-odd contribution from the two-form exchange.[5] Let us consider the product of three-point amplitudes in a t-channel exchange. The kinematic labels are as follows:



.

The sewing of two self-dual tensors and anti-self-dual tensors are given by:[6]

$$\begin{aligned}A^B_{3L}A^B_{3R} &= \frac{1}{2}\left(\Delta_{(e|da}u_{p|f)}\Delta^e{}_{bc}u^f_k\right)\left(\tilde{u}_{1\dot a}\tilde{u}_{4\dot d}\tilde{u}_{2\dot b}\tilde{u}_{3\dot c}\right)\\ &= \frac{1}{2}\Big[u_{1a}u_{4d}u_{2b}u_{3c} - 2s(u_{1a}w_{4d}+u_{4d}w_{1a})(u_{2b}w_{3c}+u_{3c}w_{2b})\Big]\tilde{u}_{1\dot a}\tilde{u}_{4\dot d}\tilde{u}_{2\dot b}\tilde{u}_{3\dot c}\\ A^{\bar{B}}_{3L}A^{\bar{B}}_{3R} &= \frac{1}{2}u_{1a}u_{4d}u_{2b}u_{3c}\Big[\tilde{u}_{1\dot a}\tilde{u}_{4\dot d}\tilde{u}_{2\dot b}\tilde{u}_{3\dot c} - 2s(\tilde{u}_{1\dot a}\tilde{w}_{4\dot d}+\tilde{u}_{4\dot d}\tilde{w}_{1\dot a})(\tilde{u}_{2\dot b}\tilde{w}_{3\dot c}+\tilde{u}_{3\dot c}\tilde{w}_{2\dot b})\Big]\,,\end{aligned}$$

[5]Recall that the coupling of two-forms to gauge field is done via the three-form field strength H^2 as well as $B\wedge F\wedge F$. Denoting the resulting three-point coupling as V_1 and V_2 respectively, we are interested in the parity-odd contribution from the gluing of V_1 and V_2. We will discuss this in more detail in subsection 4.4.3

[6]For the variables related to legs k and p, one can show that $(u_p\cdot u_k)(\tilde{u}_k\cdot\tilde{u}_p) = -s$. Using the b_i shift invariance, one can "gauge fix" w_p and w_k such that:

$$(w_p\cdot u_p) = (w_k\cdot u_k) = 0$$

and similar result for $(\tilde{w}_p\cdot\tilde{u}_p)$ and $(\tilde{w}_k\cdot\tilde{u}_k)$. Furthermore, we also have $(w_p\cdot w_k) = 1/(u_p\cdot u_k)$. The leftover rescaling symmetries α_L can α_R can be further fixed by requiring $(u_k\cdot u_p) = (\tilde{u}_k\cdot\tilde{u}_p) = \sqrt{-s}$.

where k and p refers to the momentum in the internal leg with $k = -p$. Thus the parity-odd contribution to the rational term coming from a two-form exchange is given by:

$$\begin{aligned} Res_t &= A^B_{3L}A^B_{3R} - A^{\bar{B}}_{3L}A^{\bar{B}}_{3R} \\ &= s(u_{1a}w_{4d} + u_{4d}w_{1a})(u_{2b}w_{3c} + u_{3c}w_{2b})(\tilde{u}_{1\dot{a}}\tilde{u}_{4\dot{d}}\tilde{u}_{2\dot{b}}\tilde{u}_{3\dot{c}}) \\ &\quad -s(\tilde{u}_{1\dot{a}}\tilde{w}_{4\dot{d}} + \tilde{u}_{4\dot{d}}\tilde{w}_{1\dot{a}})(\tilde{u}_{2\dot{b}}\tilde{w}_{3\dot{c}} + \tilde{u}_{3\dot{c}}\tilde{w}_{2\dot{b}})(u_{1a}u_{4d}u_{2b}u_{3c}) \,. \end{aligned} \tag{4.43}$$

Following [48] one can straightforwardly rewrite this in terms of the full six-dimensional spinors as:

$$Res_t = -\frac{\langle 1_a 2_b 3_c 4_d\rangle[1_{\dot{a}}|2|3_{\dot{c}}][4_{\dot{d}}|1|2_{\dot{b}}] - \langle 1_a|2|3_c\rangle\langle 4_d|1|2_b\rangle[1_{\dot{a}}2_{\dot{b}}3_{\dot{c}}4_{\dot{d}}]}{s} \,. \tag{4.44}$$

We will now show that Eq. (4.44) is equivalent to $F^{(4)}/s$, the residue in Eq. (4.23) and Eq. (4.34). First we list the following useful identities:

$$\begin{aligned} \langle ijml\rangle[i|j|n] &= \langle i|j|m\rangle\big\{\langle j|n]\langle l|i] - \langle j|i]\langle l|n]\big\} && (4.45)\\ &\quad +\langle i|j|l\rangle\big\{\langle j|i]\langle m|n] - \langle j|n]\langle m|i]\big\} + \langle l|j|m\rangle\langle j|i]\langle i|n] \,, && (4.46)\\ [ijml]\,\langle i|j|n\rangle &= [i|j|m]\big\{[\langle n|j]\langle i|l] - \langle i|j]\langle n|l]\big\} && (4.47)\\ &\quad +[i|j|l]\big\{[\langle i|j]\langle n|m] - \langle n|j]\langle i|m]\big\} + [l|j|m]\langle i|j]\langle n|i] \,. && (4.48) \end{aligned}$$

We have suppressed the little group labels. These identities can be derived by noting that each product on the LHS involves the product of two SU(4) Levi-Cevitas, the latter of which can be rewritten as a product of Kronecker deltas. Defining

$$f_{ijml} \equiv \langle ijml\rangle[i|j|m][l|i|j\rangle - [ijml]\langle i|j|m\rangle\langle l|i|j] \,, \tag{4.49}$$

and using momentum conservation, $k_i + k_j + k_m + k_l = 0$, it is trivial to show that $f_{ijml} = f_{jilm} = f_{mlij} = f_{lmji}$. Substituting Eq. (4.45) into Eq. (4.49), one finds

$$\begin{aligned} f_{ijml} &= s_{ij}\big\{\langle i|m]\langle m|j]\langle j|l]\langle l|i] + \langle i|j]\langle j|l]\langle l|m]\langle m|i] && (4.50)\\ &\qquad - \langle i|l]\langle l|j]\langle j|m]\langle m|i] - \langle i|m]\langle m|l]\langle l|j]\langle j|i]\big\} && (4.51)\\ &\quad +\big\{\langle i|j]\langle j|i]\langle l|m] - \langle l|i]\langle i|j]\langle j|m] - \langle l|j]\langle j|i]\langle i|m]\big\}\langle m|p_ip_j|l] && (4.52)\\ &\quad -\big\{\langle i|j]\langle j|i]\langle m|l] - \langle m|j]\langle j|i]\langle i|l] - \langle m|i]\langle i|j]\langle j|l]\big\}\langle l|p_jp_i|m] && (4.53)\\ &\quad +2\big\{\langle i|j]\langle j|m]\langle m|i] + \langle i|m]\langle m|j]\langle j|i]\big\}\langle l|p_ip_j|l] && (4.54) \end{aligned}$$

so that one can show that

$$8F^{(4)} = f_{1234} + f_{2143} + f_{3412} + f_{4321} - f_{2134} - f_{1243} - f_{3421} - f_{4312}\,. \tag{4.55}$$

Thus we conclude that

$$2F^{(4)} = f_{1234} - f_{2134}\,. \tag{4.56}$$

Finally, one can show that $f_{ijml} = -f_{jiml}$ by using Schouten identity,

$$\begin{aligned} [i|j|m][l|i|j] &= 2s_{ij}[mlij] - [l|j|i][m|i|j]\,, \\ \langle i|j|m\rangle\langle l|i|j\rangle &= 2s_{ij}\langle mlij\rangle - \langle l|j|i\rangle\langle m|i|j\rangle\,. \end{aligned} \tag{4.57}$$

Thus in conclusion, we indeed find that

$$Res_t = -\frac{f_{1234}}{s} = -\frac{F^{(4)}}{s}\,, \tag{4.58}$$

which is precisely the residue observed in the t-channel factorization limit in chiral QED and QCD, i.e. Eq. (4.23) and Eq. (4.34). This confirms that the final parity-odd amplitude that satisfies unitarity and locality has a factorization channel which is consistent with the tree-level exchange of a two-form.

4.4.3 Comments on tree-level gauge invariance

Let us consider the following question: is it possible to construct a parity-odd four-point gluon amplitude that is purely rational and whose factorization channels include the exchange of a self-dual two-form? From Eq. (4.58), we see that the t-channel exchange of a two-form has a residue that also involves an s (u) channel pole. This implies that any amplitude that has a t-channel B-field exchange must also include an s- or an u-channel exchange.[7] Without loss of generality, let us assume that the other channel is an s-channel pole. Since the residue on the s-channel pole is simply a cyclic rotation of Eq. (4.58), it is easy to see that the only solution is given by:[8]

$$\frac{F^{(4)}(s-t)}{stu}\,. \tag{4.59}$$

[7] For $t = 0$, $s = -u$. So on the t-channel pole one cannot distinguish an s-channel from an u-channel pole.

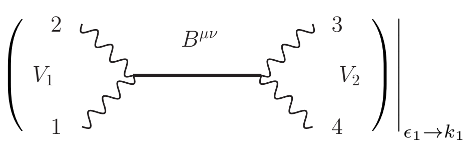
[8] It is impossible to have terms that do not have poles. This is simplest to see using polarization vectors. The amplitude has mass-dimensions 2 and involves a Levi-Cevita. The only pole-free structure that satisfies this criteria is that of the form $\epsilon(F_1F_2e_3e_4)$, which cannot be made gauge invariant.

Thus, naively, one would conclude that the answer to the question at the beginning of this section is yes. Note however, that the result in Eq. (4.59) has a non-vanishing u-pole, whose residue is *twice* that of the would-be two-form exchange residues in the s- or t-channels:

$$Res_u\left[\frac{F^{(4)}(s-t)}{stu}\right]=\frac{-2F^{(4)}}{s},\ Res_s\left[\frac{F^{(4)}(s-t)}{stu}\right]=\frac{-F^{(4)}}{u},\ Res_t\left[\frac{F^{(4)}(s-t)}{stu}\right]=\frac{F^{(4)}}{u}\,. \quad (4.60)$$

This factor of 2 makes it impossible for the collinear limit of all channels in Eq. (4.59) to consistently factorize into the same three-point amplitudes. In other words, one cannot obtain a purely rational four-point parity-odd amplitude whose poles can be consistently interpreted as the exchange of a two-form.

The difficulty above is precisely an on-shell way of saying that one cannot have consistent tree-level coupling of a two-form with vector fields, if the coupling allows for parity-odd amplitudes. This is a reflection of the fact that one cannot maintain gauge invariance if the vectors couple to the two-form via both the three-form field strength, and a parity-odd coupling. Recall that in the standard Green-Schwarz mechanism, one introduces an irrelevant operator $B\wedge F\wedge F$, which when combined with the cross terms in $H^2\equiv(dB+\omega_3)^2$ with $\omega_3=tr(FA-A^3/3)$, generates the requisite gauge anomaly. The anomaly arises from the fact that one cannot define the gauge variation of the two-form $B_{\mu\nu}$ such that *both* $B\wedge F\wedge F$ and H^2 are simultaneously gauge invariant. As a consequence, the tree-diagram that involves a vertex from $B\wedge F\wedge F$ and another from H^2 is always anomalous under the gauge variation:



$$= 2F_2\wedge F_3\wedge F_4 \quad (4.61)$$

where $V_1=\partial_{[\mu}B_{\nu\rho]}tr(A^\mu\partial^\nu A^\rho)$ and $V_2=B\wedge tr(F\wedge F)$. Note that the combination of V_1 and V_2 is precisely what produces a parity odd-amplitude.

The only way out of this predicament, from an on-shell view point, is if something else absorbed the excess residues. This is precisely what the cut-constructible part of the one-loop chiral fermion accomplished in sec. 4.3.3! Thus we have arrived at the following conclusion: A tree-level exchange of two-forms coupled to external gauge fields is consistent with locality and unitarity *only if* it is accompanied by the cut-constructible part of a chiral fermion loop. Thus instead of gauge anomaly cancellation, from the on-shell view point it is the cancellation of excess residues on factorization poles. Not surprisingly, Eq. (4.59) appears as the rational term required by locality in Eq. (4.33).

Note that for chiral chiral QED, the same analysis applies, and the only difference is that the rational term cancels in the end, as can be seen from Eq. (4.33) by identifying all trace factors.

4.5 Gauge invariant rational terms from Feynman rules

In this section, we make contact between the traditional discussion of perturbative anomalies, from analysis of Feynman diagrams, and the previous on-shell presentation. Straightforward analysis of one-loop Feynman diagrams shows that, in $D = 2(n-1)$-dimensions, the first non-zero parity-odd terms appear at n-points. As such, only the parity-odd terms in the n-gon Feynman diagram are relevant to our discussion.

After integral reduction, the scalar integral coefficients exactly match those obtained from generalized unitarity. However, rational terms obtained from this integral reduction are not gauge-invariant. This is the source of the anomaly. Crucially, the Green-Schwarz mechanism introduces new gauge-*variant* interactions directly into the action which cancel the gauge-variation of the rational term from the chiral fermion loop. Combining the Green-Schwarz trees with the anomalous rational term remarkably, though unsurprisingly, produces precisely the gauge-invariant rational term, in section 4.3.3, required by locality from generalized unitarity.

The calculation is organized as follows. First, we extract the gauge-variant rational terms from the $n = D/2+1$-point one-loop Feynman diagram with internal chiral fermions and external gluons. The rational term, for a given cyclic ordering of gluons, is given by the product of a permutation-invariant function of the external data multiplied by a color-trace. Second, guided by the color-structure of the full one-loop amplitude, we extract the unique, gauge-invariant, combination of one-loop chiral QCD amplitudes and tree-level, parity-violating, gravitational amplitudes. Finally, we present explicit forms for the n-gon integral coefficients, in both six- and eight-dimensions. The structure of these integral coefficients makes it clear that the structure of the highest-order integral coefficient in D-dimensions contains all the information concerning the Green-Schwarz mechanism that is conventionally accessed only through an action point of view.

4.5.1 The anomalous rational term

To cleanly extract the parity-odd rational terms in $D = 2(n-1)$-dimensions, it is convenient to separate the loop-momenta, ℓ^μ, into D-dimensional and -2ϵ dimensional pieces, respectively $\bar{\ell}^\mu$ and $\mu^{\hat{\mu}}$:

$$\ell^\mu \to \bar{\ell}^\mu + \mu^{\hat{\mu}}\,. \tag{4.62}$$

In this notation, the external momenta is purely D-dimensional, and one has,

$$\frac{1}{(\ell+k)^2} = \frac{1}{(\bar{\ell}+k_1)^2+\mu^2}\,. \tag{4.63}$$

Similar separation is required for the definition of the corresponding Gamma matrices. We follow the conventions of 't Hooft and Veltman:

$$\gamma^\mu \to \bar{\gamma}^\mu + \hat{\gamma}^{\hat{\mu}}\,. \tag{4.64}$$

The extra -2ϵ-dimensional piece, $\hat{\gamma}^{\hat{\mu}}$, satisfies the following commutation relations,

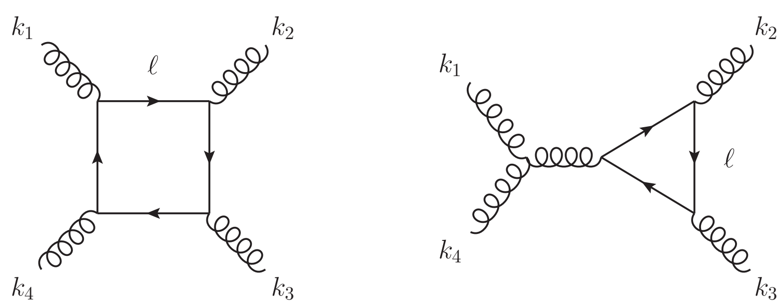
$$\{\hat{\gamma}^{\hat{\mu}},\hat{\gamma}^{\hat{\nu}}\} = \eta^{\hat{\mu}\hat{\nu}},\ \{\hat{\gamma}^{\hat{\mu}},\bar{\gamma}^{\nu}\} = 0,\ \ [\gamma_{-1},\hat{\gamma}^{\hat{\nu}}] = 0\,, \tag{4.65}$$

where γ_{-1} denotes the D-dimensional chiral matrix ($\gamma_{-1}=\gamma_5$ in four-dimensional notation). Note that $\hat{\gamma}^{\hat{\mu}}$ may only be contracted with $\mu^{\hat{\mu}}$; all other vectors in the problem are purely D-dimensional.

First note that due to the following identity,

$$(1+\gamma_{-1})(\not{K}_{(D)})(1-\gamma_{-1}) = 2(1+\gamma_{-1})(\not{K}_{(D)}),\ \ (1+\gamma_{-1})(\not{\mu})(1-\gamma_{-1}) = 0\,, \tag{4.66}$$

where K_D is some D-dimensional vector, it is straightforward to see that the spinor traces of the $n = D/2+1$-point amplitude will project out all -2ϵ-dimensional components of the loop-momenta. For example, in $D=6$ the numerator of the box-diagram (a), derived by stringing together four chiral-fermion-gluon vertex $V_3 = \bar{\psi}\not{A}(1+\gamma_{-1})\psi$, is given as:



(a) (b)

$$\begin{aligned} n_{(a)} &= tr[\not{\epsilon}_1(1+\gamma_{-1})\not{\ell}\not{\epsilon}_2(1+\gamma_{-1})(\not{\ell}-\not{k}_2)\not{\epsilon}_3(1+\gamma_{-1})(\not{\ell}-\not{k}_2-\not{k}_3)\not{\epsilon}_4(1+\gamma_{-1})(\not{\ell}+\not{k}_1)] \\ &= 8tr[\gamma_{-1}\not{\epsilon}_1\bar{\not{\ell}}_{(6)}\not{\epsilon}_2(1+\gamma_{-1})(\bar{\not{\ell}}-\not{k}_2)\not{\epsilon}_3(\bar{\not{\ell}}-\not{k}_2-\not{k}_3)\not{\epsilon}_4(\bar{\not{\ell}}+\not{k}_1)]\,. \end{aligned} \tag{4.67}$$

To reiterate, by virtue of the commutation-relations for $\hat{\gamma}^{\hat{\mu}}$, the loop-momenta in the numerator are purely D-dimensional. Similarly, the numerator of the triangle diagram is purely D-dimensional:

$$\begin{aligned} n_{(b)} &= tr[\gamma_{-1} \not{J}_{1,4} (\bar{\not{\ell}} + \not{k}_2) \not{e}_2 \bar{\not{\ell}} \not{e}_3 (\bar{\not{\ell}} - \not{k}_3)] \\ &= \epsilon(J_{14} k_2 e_2 \bar{\ell} e_3 k_3)\,, \end{aligned} \tag{4.68}$$

where J^{μ}_{14} is the current at the three-point gluon vertex. This diagram integrates to zero due to its linear dependence in loop-momentum: through integral reduction, $\bar{\ell}$ is expanded on the vector k_2 and k_3, and therefore vanishes in the six-dimensional Levi-Cevita tensor.

Equipped with this specific six-dimensional example, it is straightforward to generalize: the parity-odd amplitudes in D-dimensions first appear at $n = D/2 + 1$-points, and all non-trivial contributions come from the numerator of the n-gon chiral fermion loop. Standard gamma-matrix trace identities dictate that the gamma-trace tensor, schematically, reduces in the following way:

$$tr[\gamma_{-1} \gamma^{a_1} \cdots \gamma^{a_{2n}}] = \eta^{a_1 a_2} \epsilon^{a_3 \cdots a_{2n}} + \cdots\,. \tag{4.69}$$

Again, by the antisymmetry of the D-dimensional $\epsilon^{a_1 \cdots a_D}$-tensor, no term in the numerator may contain more than three powers of $\bar{\ell}$. Now, integral reduction of the n-gon numerator casts the parity-odd amplitude onto a scalar integral basis. With at most three powers of loop-momenta in the numerator, the amplitude is a combination of n, $(n-1)$ and $(n-2)$-gon scalar integrals—and the attendant rational terms.[9] Of these three scalar integrals, only the $(n-1)$-gon and the $(n-2)$-gon integrals diverge in the ultraviolet. Their integral coefficients, which are gauge invariant, must ensure cancellation of ultraviolet divergences.

Since these scalar integral coefficients may be uniquely determined from unitarity cuts, which are simply products of gauge-invariant quantitaties, they must be gauge invariant. However, the same cannot be said of rational terms. Indeed, integral reduction produces rational terms that are not gauge invariant. This is most simply exhibited by terms in the n-gon integral with tensor numerators of degree-two in the D-dimensional loop-momenta:

$$\frac{\bar{\ell}^{\mu} \bar{\ell}^{\nu}}{\ell^2 (\ell - k_1)^2 \cdots (\ell + k_n)^2}\,. \tag{4.70}$$

[9]Note the distinction between $m < n$-gons in the chiral gauge theory and those, obtained by integral reduction *from* the defining n-gon, that multiply the *scalar* $m < n$-gon loop-integrals.

In integral reduction, one expands

$$\bar{\ell}^\mu \bar{\ell}^\nu = a_0 \eta^{\mu\nu} + a_{ij} k_i^\mu k_j^\mu \ , \tag{4.71}$$

where i, j runs over the independent momenta of the external states. Contraction of both sides of the equation with $\eta^{\mu\nu}$ partially fixes the coefficients a_0, and a_{ij}. Under this operation, the LHS simply becomes $\bar{\ell}^2 = \ell^2 - \mu^2$, i.e. an inverse propagator and a $-\mu^2$-term. Insertion of the expansion in Eq. (4.71) back into Eq. (4.70), thus yields a collection of n-, $(n-1)$-, and $(n-2)$-gon scalar integrals *and* an n-gon scalar integral with a μ^2 numerator. Explicit integration of the latter either vanishes at $\mathcal{O}(\epsilon)$, or gives a rational term. For example, one finds:

$$\int \frac{d\ell^{(D)} d\mu^{-2\epsilon}}{(2\pi)^{D-2\epsilon}} \ I_{D/2+1}[\mu^2] = -\frac{1}{(4\pi)^{D/2}} \frac{1}{(D/2)!} + O(\epsilon) \, . \tag{4.72}$$

Note that since the numerator of this integral depends on μ^2, it is not directly detectable in D-dimensional unitarity cuts. Furthermore, such terms only arise from terms in the numerator with two or more powers of loop-momenta, which is indeed the case for $n_{(a)}$ in Eq. (4.67). Note that since rational terms only appear at one-loop in even-dimensions, this automatically leads to the conclusion that anomalies can only appear at one-loop in even-dimensions.

Straightforwardly carrying-out integral reduction in this manner, and isolating the rational terms, yields the anomalous, parity-odd, rational terms for $D = 6, 8$, and 10:

$$\begin{aligned} R^{anom}_{D=6} &= -\frac{\left[G_{(123)} + G_{(124)} + G_{(134)}\right] |_{k_1 \to e_1} F_2 \wedge F_3 \wedge F_4}{3 * 4! * Gram_{(1,2,3)}} + (\text{cyclic}) \\ R^{anom}_{D=8} &= -\frac{\left[G_{(1234)} + G_{(1345)} + G_{(1235)} + G_{(1245)}\right] |_{k_1 \to e_1} F_2 \wedge F_3 \wedge F_4 \wedge F_5}{4 * 5! * Gram_{(1,2,3,4)}} + (\text{cyclic}) \\ R^{anom}_{D=10} &= -\frac{\left[G_{(12345)} + G_{(13456)} + G_{(12356)} + G_{(12456)} + G_{(12346)}\right] |_{k_1 \to e_1} F_2 \wedge F_3 \wedge \cdots \wedge F_6}{5 * 6! * Gram_{(1,2,3,4,5)}} \\ &\qquad + (\text{cyclic}) \, , \end{aligned} \tag{4.73}$$

where $G_{(i_1, \cdots, i_l)}$ is the determinant of the Gram-matrix of $k_{i_1}, \cdots, k_{i_l}$, and the notation $G_{(12345)}|_{k_1 \to e_1}$ indicates that we linearly replace k_1 with $k_1 + a e_1$ in $G_{(12345)}$, and collect the coefficient in front of a. For example for $D = 6$:

$$\left[G_{(123)} + G_{(124)} + G_{(134)}\right] |_{k_1 \to e_1} = \left[tu(\epsilon_1 \cdot k_2) + st(\epsilon_1 \cdot k_3) + su(\epsilon_1 \cdot k_4)\right] \, . \tag{4.74}$$

The presence of the Gram-determinant in the denominator is simply due to the integral reduction procedures. These rational terms are in fact invariant under permutations of the kinematic labels, and thus the rational term for the full color-dressed amplitude is given by:

$$\mathcal{R}^{anom} = R^{anom} str(T_1 \cdots T_n)\,. \tag{4.75}$$

It is straightforward to verify that these terms precisely give the anomaly: the variation of the polarization vector, $\epsilon_i \to k_i$, automatically produces Gram determinants in the numerator that cancel against those in the denominator, leaving

$$\delta_{e_1 \to k_1} \mathcal{R}^{anom} = F_2 \wedge F_3 \wedge \cdots \wedge F_n \times str(T_1 \cdots T_n)\,. \tag{4.76}$$

Crucially if $str(T_1 \cdots T_n) = 0$, the original condition for anomaly cancellation, then the entire parity-odd rational term vanishes. Note that this is consistent with the previous on-shell analysis, which showed that if $str(T_1 \cdots T_4) = 0$, all spurious poles cancel and no rational terms are needed for locality.

4.5.2 Gauge-invariant rational terms

Anomaly cancellation is more subtle when $str(T_1 \cdots T_n)$ does not vanish. Here, we construct the parity-odd gauge-invariant rational building-blocks seen in gauge-invariant expressions of parity-odd integral coefficients. Rather than construct gauge-invariant rational terms from unitarity cuts, we extract them in the more traditional way, from Feynman diagrams.

As in the previous calculation, we begin with a study of anomaly cancellation in six-dimensions, where we make contact with both the Green-Schwarz mechanism, and with the previous on-shell results. Here, the anomalous rational term can be conveniently written as:

$$\mathcal{R}^{anom} = -\frac{1}{18}\left[\left(\frac{(\epsilon_1 \cdot k_2)}{s} + \frac{(\epsilon_1 \cdot k_3)}{u} + \frac{(\epsilon_1 \cdot k_4)}{t}\right) F_2 \wedge F_3 \wedge F_4 + (\text{cyclic})\right] str(T_a T_b T_c T_d)\,. \tag{4.77}$$

If the symmetrized color-factor does not vanish, then the gauge-variation of the parity-odd rational term may be cancelled via the Green-Schwarz mechanism if and only if there exists a representation t_a such that the trace factorizes $str(T_a T_b T_c T_d) = tr(t_a t_b) tr(t_c t_d)$. As discussed above, the Green-Schwarz mechanism introduces a two-form with both parity-even and parity-odd couplings to gauges field via,

$$\begin{aligned}
\mathcal{L}_{\rm I} &= H_{abc}H^{abc} = [dB + tr(A \wedge dA)]^2 \ , \text{ and} \\
\mathcal{L}_{\rm II} &= B \wedge tr(F^{(1)} \wedge \cdots \wedge F^{(D/2-1)}) \, , \text{ which for } D = 6 \text{ reduces to} \qquad (4.78) \\
&\longrightarrow B \wedge tr(F^{(1)} \wedge F^{(2)}) \, . \qquad (4.79)
\end{aligned}$$

As repeatedly stated above, these two interactions cannot be made mutually gauge-invariant, and their gauge-*variance* cancels against that of the anomalous rational term.

To see how this happens, consider the parity-odd s-channel Feynman trees coupling gluons via two-form exchange. Parity violation mandates that both distinct interactions enter into the tree amplitude; the color-structure of the interactions dictates that these and only these parity-odd trees are proportional $tr(t_1t_2)tr(t_3t_4)$. The detailed structure of the s-channel tree contribution is,

$$GS^{6d\,\mathrm{tree}}_{(12)} = \frac{tr(t_1t_2)tr(t_3t_4)}{6\,s}\Big[(F_1\wedge F_2\wedge F_4)(\epsilon_3\cdot k_4)+(F_1\wedge F_2\wedge F_3)(\epsilon_4\cdot k_3)+\big\{(1,2)\leftrightarrow(3,4)\big\}\Big] \, . \qquad (4.80)$$

Adding $GS^{6d\,\mathrm{tree}}_{(12)}$ and R^{anom}, the term proportional to $tr(t_1t_2)tr(t_3t_4)$ is:

$$\begin{aligned}
\left(R^{anom} + GS^{6d\,\mathrm{tree}}_{(12)}\right) &= \frac{1}{18}\left[\left(2\frac{(\epsilon_1\cdot k_2)}{s} - \frac{(\epsilon_1\cdot k_4)}{t} - \frac{(\epsilon_1\cdot k_3)}{u}\right) F_2\wedge F_3\wedge F_4\right] \\
&+ \frac{1}{18}\left[\left(2\frac{(\epsilon_2\cdot k_1)}{s} - \frac{(\epsilon_2\cdot k_3)}{t} - \frac{(\epsilon_2\cdot k_4)}{u}\right) F_3\wedge F_4\wedge F_1\right] \\
&+ \frac{1}{18}\left[\left(2\frac{(\epsilon_3\cdot k_4)}{s} - \frac{(\epsilon_3\cdot k_2)}{t} - \frac{(\epsilon_3\cdot k_1)}{u}\right) F_4\wedge F_1\wedge F_2\right] \\
&+ \frac{1}{18}\left[\left(2\frac{(\epsilon_4\cdot k_3)}{s} - \frac{(\epsilon_4\cdot k_1)}{t} - \frac{(\epsilon_4\cdot k_2)}{u}\right) F_1\wedge F_2\wedge F_3\right] . \qquad (4.81)
\end{aligned}$$

It is amusing that the Green-Schwarz contribution simply tweaks the coefficient in the anomalous rational term such that the combination is gauge invariant. Converting the above into on-shell form, one finds:

$$R^{anom} + GS^{6d\,\mathrm{tree}}_{(12)} = -\frac{(t-u)}{18stu}F^4 \, . \qquad (4.82)$$

We see that the gauge invariant combination is precisely the rational term in Eq. (4.33)! Imposing locality on the result obtained from unitarity cuts inevitably leads to the rational term that is precisely the remnant combination of the anomalous rational term and the Green-Schwarz tree-amplitude. In other words, the amplitude not only "knows" that there is a new particle in the spectrum, it also knows that there is a remnant from the cancelation and its precise form.

Cancellation between the parity-odd, anomalous, rational term and the Green-Schwarz mechanism is slightly more subtle in eight-dimensions and higher: as noted in the original literature, new terms in the action are needed [46, 62]. Recall that the full $\mathcal{R}^{anom}$ is given by $R^{anom}(1\cdots n)str(T_1\cdots T_n)$. For gauge-theories which admit Green-Schwarz anomaly cancelation, in factorized form, this can be rewritten as,

$$\mathcal{R}^{anom} = R^{anom}(1\cdots n)\sum_{ab} tr(t_a t_b)str(t_{c_3}\cdots t_{c_n})\,, \tag{4.83}$$

where the sum is over all distinct $\binom{n}{2}$ pairs of particle labels. Under a gauge-variation, $\epsilon_1 \to k_1$, the anomalous rational term becomes,

$$\mathcal{R}^{anom}\big|_{\epsilon_1\to k_1} = \big(F_2\wedge\cdots\wedge F_n\big)\times\sum_{ab} tr(t_a t_b)str(t_{c_3}\cdots t_{c_n})\,. \tag{4.84}$$

Thus we see that a gauge-variation of the anomalous rational terms will have *all* possible double-trace structures.

However the gauge variation of any Green-Schwarz tree, say $\epsilon_1 \to k_1$, must be to proportional to a particular subset of double-trace structures: $tr(t_1 t_j)str(t_{c_3}\cdots t_{c_n})$. In other words, the classic Green-Schwarz trees are insufficient to cancel all gauge-anomalies in eight-dimensions and higher! Alternatively, it is not possible for the coefficient of any particular double-trace structure, arising from the conventional Green-Schwarz tree and the anomalous rational term, to be made gauge-invariant for all possible variations.

Not surprisingly, the missing cancellation comes from local counter-terms which are free of poles:

$$\begin{aligned}\mathcal{L}_{\text{III}} &= \omega_3^{\{12\}}\wedge\omega_{D-3}^{\{3\ldots n\}} \\ &= \Big\{\big[tr(A_1\wedge dA_2)\wedge tr(A_3\wedge F_4\wedge\cdots\wedge F_n) + (1\leftrightarrow 2)\big] + \mathcal{P}(3\cdots n)\Big\}\,,\end{aligned} \tag{4.85}$$

Under gauge-variations, these interactions transform in the following way:

$$\begin{aligned}\omega_3^{\{45\}}\wedge\omega_5^{\{123\}}\big|_{\epsilon_1\to k_1} &= -3\big(F_2\wedge F_3\wedge F_4\wedge F_5\big)\times\sum_{b=2}^{5} tr(t_4t_5)str(t_1t_2t_3)\,,\\ \omega_3^{\{45\}}\wedge\omega_5^{\{123\}}\big|_{\epsilon_5\to k_5} &= +2\big(F_1\wedge F_2\wedge F_3\wedge F_4\big)\times\sum_{b=2}^{5} tr(t_4t_5)str(t_1t_2t_3)\,,\end{aligned}$$

where, for concreteness, we have chosen a particular wedge-product in eight-dimensions. Using these new contact terms, we may now construct gauge-invariant rational terms in arbitrary even-dimensional spaces. The gauge-invariant coefficient of, for instance, the $tr(12)str(345)$ double-trace term present in the eight-dimensional five-point one-loop parity-odd rational term is:

$$\mathrm{Inv}^{8d}_{(12)} = R^{\mathrm{anom}} - \left\{ \frac{5}{2}\left(GS^{8d\,\mathrm{tree}}_{(12)}\right) + \frac{1}{2}\left(\omega_3^{\{12\}} \wedge \omega_5^{\{345\}}\right) \right\} \text{ , where} \tag{4.86}$$

$$GS^{8d\,\mathrm{tree}}_{(12)} = \left(\frac{2\epsilon_1 \cdot k_2}{s_{12}} F_2 \wedge F_3 \wedge F_4 \wedge F_5 + \frac{2\epsilon_2 \cdot k_1}{s_{12}} F_3 \wedge F_4 \wedge F_5 \wedge F_1 \right), \tag{4.87}$$

where these expressions come explicitly from the Feynman rules for the interaction terms in Eqs. (4.78) and (4.85). The corresponding structure which appears in ten dimensions is given by,

$$\mathrm{Inv}^{10d}_{(12)} = R^{\mathrm{anom}} - \left\{ \frac{6}{2}\left(GS^{10d\,\mathrm{tree}}_{(12)}\right) + \frac{1}{2}\left(\omega_3^{\{12\}} \wedge \omega_7^{\{3456\}}\right) \right\} \text{ , where} \tag{4.88}$$

$$GS^{10d\,\mathrm{tree}}_{(12)} = \left(\frac{2\epsilon_1 \cdot k_2}{s_{12}} F_2 \wedge F_3 \wedge F_4 \wedge F_5 \wedge F_6 + \frac{2\epsilon_2 \cdot k_1}{s_{12}} F_3 \wedge F_4 \wedge F_5 \wedge F_6 \wedge F_1 \right). \tag{4.89}$$

One may explicitly check that these sums are manifestly gauge-invariant in all lines.

4.5.3 Integral coefficients from invariant rational terms

These gauge-invariant parity-odd rational objects, "remnants", constructed explicitly from interference between the gauge-anomalies between Feynman diagrams, furnish a set of building-blocks for the gauge-invariant integral-coefficients. Intriguingly, the structure of the highest-oder integral coefficient (the box in $D = 6$) in a given cyclic ordering may be neatly rewritten,

$$C_4^{D=6}(1,2,3,4) = \frac{s_{12}s_{23}s_{34}s_{41}}{Gram_{(1,2,3)}} \times (\mathrm{Inv}^{6d}_{(12)} + \mathrm{Inv}^{6d}_{(23)} + \mathrm{Inv}^{6d}_{(34)} + \mathrm{Inv}^{6d}_{(41)}) \tag{4.90}$$

where, as commented above, $Gram_{(1,2,3)} = stu = s_{12}s_{14}s_{13}$, and the six-dimensional invariants may be rewritten in terms of the spinor-helicity variables, as reflected in Eq. (4.82). Motivated by this natural form for the parity-odd box-coefficient in six-dimensions, one would naturally conjecture that the pentagon-coefficient in eight-dimensions, etc., would be related to those invariant building-blocks in a similarly nice way. Indeed, the ansatz,

$$C_5^{D=8}(1,2,3,4,5) = \frac{s_{12}s_{23}s_{34}s_{45}s_{51}}{Gram_{(1,2,3,4)}} \times (\mathrm{Inv}^{8d}_{(12)} + \mathrm{Inv}^{8d}_{(23)} + \mathrm{Inv}^{8d}_{(34)} + \mathrm{Inv}^{8d}_{(45)} + \mathrm{Inv}^{8d}_{(51)}) \tag{4.91}$$

exactly reproduces the parity-odd scalar pentagon-coefficient extracted from the $(1,2,3,4,5)$-pentagon *diagram* in the chiral gauge-theory. In other words, the gauge-invariant integral coefficients contain explicit information about these "remnant" rational terms, leftover after anomaly cancellation between the Green-Schwarz trees and the chiral-fermion loops.

Natural extensions of this ansatz to the (famous) parity-odd hexagon-coefficient [46] fail, due to the complicated singularity structure at six-points, inherent in the scalar hexagon-integral. However, as mentioned in the introduction, recent methods [63] exploit differential equations to capture the behavior of integrals near their singularities without the need to explicitly integrate them. It would be extremely interesting to use these methods to uniquely fix the parity-violating sectors in ten-dimensions. Successful implementation of this program would forcefully assert the conclusion that the highest-order integral-coefficient in the parity-odd sector of a theory entirely dictates the structure of the whole amplitude. We leave such explorations for future work, and move-on to address the pressing issue of how the perturbative gravitational anomalies, in theories with chiral- yet CPT-invariant spectra, manifest themselves in the on-shell S-matrix in, for example, six-dimensions [56].

4.6 $D = 6$ Gravitational anomaly

In this section we will be interested in gravity coupled to various chiral-matter in six-dimensions. We will compute the cut-constructible piece of the parity-odd one-loop four-point graviton amplitude, which will be unique to chiral-matter. In six-dimensions, there are three types of chiral coupling for gravity theories. Chiral-matter includes chiral-fermions as well as (anti) self-dual two-forms, while for supergravity theories there is also the presence of chiral-gravitinos. The presence of self-dual two-forms introduces an interesting question which can be readily addressed from our perspective: for a generic number of (anti) self-dual two-forms, there are no covariant actions and hence Feynman rules are unavailable. However, from an S-matrix point of view, there is no problem of defining consistent tree-level scattering amplitudes involving chiral two-forms, and this is all that is needed for the construction of it's quantum corrections. We will proceed by applying unitarity methods to compute the coefficients in the scalar-integral basis. This again defines the S-matrix up to rational ambiguities, which we will fix by enforcing locality.

It is well known that at the quantum level, chiral two-forms induce an anomaly when coupled to gravity. Thus a non-trivial test is to demonstrate that one recovers both the gravitational anomaly, and their attendant anomaly cancellation conditions, in the process of demanding unitarity and locality. As we now show, this is indeed the case.

4.6.1 $D = 6$ chiral tree-amplitude for $M_4(hhXX)$

We begin by presenting the four-point amplitude for two gravitons coupled to a pair of chiral gravitinos, self-dual two-forms, and chiral spinors respectively:

$$M_4(h_{a_1b_1\dot{a}_1\dot{b}_1}, h_{a_2b_2\dot{a}_2\dot{b}_2}, \psi_{a_3b_3\dot{a}_3}, \psi_{a_4b_4\dot{a}_4}) = \left.\frac{\langle 1_{a_1}2_{a_2}3_{a_3}4_{a_4}\rangle\langle 1_{b_1}2_{b_2}3_{b_3}4_{b_4}\rangle[1_{\dot{a}_1}2_{\dot{a}_2}3_{\dot{a}_3}4_{\dot{a}_4}][1_{\dot{b}_1}|3|2_{\dot{b}_2}]}{s\ t\ u}\right|_{sym}, \tag{4.92}$$

$$M_4(h_{a_1b_1\dot{a}_1\dot{b}_1}, h_{a_2b_2\dot{a}_2\dot{b}_2}, B_{a_3b_3}, B_{a_4b_4}) = \left.\frac{\langle 1_{a_1}2_{a_2}3_{a_3}4_{a_4}\rangle\langle 1_{b_1}2_{b_2}3_{b_3}4_{b_4}\rangle[1_{\dot{a}_1}|3|2_{\dot{a}_2}][1_{\dot{b}_1}|4|2_{\dot{b}_2}]}{s\ t\ u}\right|_{sym}, \tag{4.93}$$

$$M_4(h_{a_1b_1\dot{a}_1\dot{b}_1}, h_{a_2b_2\dot{a}_2\dot{b}_2}, \chi_{a_3}, \chi_{a_4}) = \left.\frac{\langle 1_{a_1}2_{a_2}3_{a_3}4_{a_4}\rangle\langle 1_{b_1}|3|2_{b_2}\rangle[1_{\dot{a}_1}|4|2_{\dot{a}_2}][1_{\dot{b}_1}|4|2_{\dot{b}_2}]}{s\ t\ u}\right|_{sym}, \tag{4.94}$$

where $|_{sym}$ indicates the symmetrization of the SU(2) indices if more than one is present for a given leg. The symmetrization reflects the fact that the graviton, gravitino, self-dual two-form and chiral spinor transforms as $(3,3)$, $(2,3)$, $(0,3)$ and $(0,2)$ under the SU(2)×SU(2) little group.

The validity of the above amplitude can be confirmed by showing that on all three factorization poles, it factorizes correctly into a product of three-point amplitudes involving two chiral states coupled to a graviton. The three-point amplitudes are uniquely determined from little group requirements [61]:

$$M_3(h_{a_1b_1\dot{a}_1\dot{b}_1}, \psi_{a_2b_2\dot{a}_2}, \psi_{a_3b_3\dot{a}_3}) = (\Delta_{a_1a_2a_3}\tilde{\Delta}_{\dot{a}_1\dot{a}_2\dot{a}_3})(\Delta_{b_1b_2b_3}\tilde{u}_{1\dot{b}_1})\Big|_{sym}, \tag{4.95}$$

$$M_3(h_{a_1b_1\dot{a}_1\dot{b}_1}, B_{a_2b_2}, B_{a_3b_3}) = (\Delta_{a_1a_2a_3}\tilde{u}_{1\dot{a}_1})(\Delta_{b_1b_2b_3}\tilde{u}_{1\dot{b}_1})\Big|_{sym}, \tag{4.96}$$

$$M_3(h_{a_1b_1\dot{a}_1\dot{b}_1}, \chi_{a_2}, \chi_{a_3}) = (\Delta_{a_1a_2a_3}\tilde{u}_{1\dot{a}_1})(u_{1b_1}\tilde{u}_{1\dot{b}_1})\Big|_{sym}. \tag{4.97}$$

Alternatively, the above four-point amplitudes can be derived from the fully supersymmetric $\mathcal{N} = (2,2)$ supergravity amplitude. In superspace, for the sake of being explicit, the full multiplet is given by an expansion of $\eta^{ia}, \tilde{\eta}^{\hat{i}\dot{a}}$s, where $i, \hat{i} = 1, 2$. Thus we have in total 8 grassmann variables, and the superfield is expanded as

$$
\begin{aligned}
\Phi(\eta^{ia},\tilde{\eta}^{\hat{i}\dot{a}}) &= \phi+\eta^{ia}\psi_{ia}+\tilde{\eta}^{\hat{i}\dot{a}}\tilde{\psi}_{\hat{i}\dot{a}}+\eta^{2(ij)}\phi_{(ij)}+\tilde{\eta}^{2(\hat{i}\hat{j})}\tilde{\phi}_{(\hat{i}\hat{j})}+\eta^{i(a}\eta_i^{b)}B_{(ab)}+\tilde{\eta}^{\hat{i}(\dot{a}}\tilde{\eta}_{\hat{i}}^{\dot{b})}\tilde{\phi}_{(\dot{a}\dot{b})}+\tilde{\eta}_{\hat{i}}^{\dot{a}}\eta_j^{a}A_{a\dot{a}}^{\hat{i}j}\\
&+ \eta^{2(ij)}\tilde{\eta}^{\dot{a}\hat{i}}\tilde{\psi}_{(ij)\hat{i}\dot{a}}+\tilde{\eta}^{2(\hat{i}\hat{j})}\eta^{ai}\psi_{(\hat{i}\hat{j})ia}+\eta^{ai}\eta_a^{j}\eta_i^{b}\psi_{jb}+\tilde{\eta}^{\dot{a}\hat{i}}\tilde{\eta}_{\dot{a}}^{\hat{j}}\tilde{\eta}_{\hat{i}}^{\dot{b}}\tilde{\psi}_{\hat{j}\dot{b}}+\eta^{2(ab)}\tilde{\eta}^{\hat{j}\dot{a}}\psi_{(ab)\dot{a}\hat{j}}+\tilde{\eta}^{2(\dot{a}\dot{b})}\eta^{ia}\tilde{\psi}_{i(\dot{a}\dot{b})a}\\
&+ \eta^{2(ij)}\tilde{\eta}^{2(\hat{i}\hat{j})}\phi_{(ij)(\hat{i}\hat{j})}+\eta^{2(ab)}\tilde{\eta}^{2(\dot{a}\dot{b})}g_{(ab)(\dot{a}\dot{b})}+\eta^{2(ab)}\tilde{\eta}^{2(\hat{i}\hat{j})}B_{(ab)(\hat{i}\hat{j})}+\eta^{2(ij)}\tilde{\eta}^{2(\dot{a}\dot{b})}B_{(ij)(\dot{a}\dot{b})}\\
&+ \eta^{ia}\eta_a^{j}\eta_i^{b}\tilde{\eta}^{\hat{i}\dot{a}}A_{\hat{i}jb\dot{a}}+\tilde{\eta}^{\hat{i}\dot{a}}\eta_{\dot{a}}^{\hat{j}}\eta_{\hat{i}}^{\dot{b}}\eta^{ia}A_{\hat{j}ia\dot{b}}+\eta^4\phi'+\tilde{\eta}^4\tilde{\phi}'+\cdots\,,
\end{aligned}
$$

where $\cdots$ are degree 5 and higher in grassmann variables. One can see the bosonic states are $2\times 1+2\times(4\times 3+4\times 4)+(4\times 3\times 3+2\times 4\times 4)=126$, as expected for maximal super gravity. The $\mathcal{N}=(2,2)$ supergravity four-point amplitude is given by

$$
\mathcal{A}_4=\frac{\delta^8(Q)\delta^8(\tilde{Q})}{stu},\qquad \delta^8(Q)\equiv\prod_{i=1}^{2}\prod_{A=1}^{4}\left(\sum_{n=1}^{4}\lambda_n^{Aa}\eta_{na}^{i}\right),\quad \delta^8(\tilde{Q})\equiv\prod_{\hat{i}=1}^{2}\prod_{A=1}^{4}\left(\sum_{n=1}^{4}\tilde{\lambda}_{nA}^{\dot{a}}\tilde{\eta}_{n\dot{a}}^{\hat{i}}\right).\tag{4.98}
$$

The component amplitudes can be obtained as:

$$
\begin{aligned}
Eq.\ (4.93) &= \mathcal{A}_4\big|_{\eta_1^{2(ab)}\tilde{\eta}_1^{2(\dot{a}\dot{b})}\eta_2^{2(ab)}\tilde{\eta}_2^{2(\dot{a}\dot{b})}\eta_3^{2(ab)}\tilde{\eta}_3^{\hat{j}\dot{a}}\tilde{\eta}_{3\dot{a}}^{\hat{i}}\tilde{\eta}_{3\hat{j}}^{\dot{b}}\eta_4^{2(ab)}\tilde{\eta}_4^{\hat{j}\dot{a}}}\,,\\
Eq.\ (4.94) &= \mathcal{A}_4\big|_{\eta_1^{2(ab)}\tilde{\eta}_1^{2(\dot{a}\dot{b})}\eta_2^{2(ab)}\tilde{\eta}_2^{2(\dot{a}\dot{b})}\eta_3^{2(ab)}\tilde{\eta}_3^{2(\hat{i}\hat{j})}\eta_4^{2(ab)}\tilde{\eta}_4^{2(\hat{i}\hat{j})}}\,,\\
Eq.\ (4.95) &= \mathcal{A}_4\big|_{\eta_1^{2(ab)}\tilde{\eta}_1^{2(\dot{a}\dot{b})}\eta_2^{2(ab)}\tilde{\eta}_2^{2(\dot{a}\dot{b})}\tilde{\eta}_3^{2(\hat{i}\hat{j})}\eta_3^{ai}\tilde{\eta}_4^{2(\hat{i}\hat{j})}\eta_4^{ai}\eta_{4a}^{j}\eta_{4i}^{b}}\,.
\end{aligned}
$$

We now proceed to use these results to construct the parity-odd four-point one-loop amplitude, where the potential anomalies reside.

4.6.2 Chiral loop-amplitudes

Using the four-point tree-amplitudes with two graviton and two chiral-matter, we can now proceed and compute the one-loop four-graviton amplitude with chiral-matter in the loop. Again we begin by expanding the integrals in the scalar integral basis:

$$
\begin{aligned}
\mathcal{M}_4^{1-\mathrm{L,odd};X} &= C_4^X(1,2,3,4)I_{1,2,3,4}+C_4^X(1,3,4,2)I_{1,3,4,2}+C_4^X(1,4,2,3)I_{1,4,2,3}\\
&+ C_{3s}^X I_{3s}+C_{3u}^X I_{3u}+C_{3t}^X I_{3t}+C_{2s}^X I_{2s}+C_{2u}^X I_{2u}+C_{2t}^X I_{2t}+R\,,
\end{aligned}\tag{4.99}
$$

where $X=(\psi,B,\chi)$ represents the contribution from the chiral gravitino, two-form, and fermion respectively. Again to simplify our task, we single out a four-dimensional sub plane spanned by the momenta of the external legs. This allows us to project the six-dimensional states onto four-dimensional ones: the $3 \times 3 = 9$-degrees of freedom then becomes the graviton, two vectors, and three-scalars respectively:

$$(h_{11,\dot{1}\dot{1}}\,,\; h_{22,\dot{2}\dot{2}})\;,\;(h_{12,\dot{2}\dot{2}}\,,\; h_{12,\dot{1}\dot{1}})\;,\;(h_{11,\dot{1}\dot{2}}\,,\; h_{22,\dot{1}\dot{2}})\;,\;(h_{12,\dot{1}\dot{2}}\,,\; h_{11,\dot{2}\dot{2}}\,,\; h_{22,\dot{1}\dot{1}})\,. \tag{4.100}$$

Since we are considering the parity-odd part of the amplitude, to ensure that we have enough vectors to span all six directions, we will choose all four six-dimensional graviton states to be distinct. We begin with the following helicity configuration:

$$(1_{12\dot{1}\dot{2}}\,, 2_{22\dot{1}\dot{1}}\,, 3_{22\dot{2}\dot{2}}\,, 4_{11\dot{2}\dot{2}})\,.$$

The computation of the integral coefficients is straightforward and similar to that of the QCD and QED computation. The slight distinction is that the highest power of $m\tilde{m}$ appearing in the unitarity cuts is three times that of the gauge theories, which leads to the result that after integral reduction, the power of poles for the scalar integrals will be higher. As an example, the $m\tilde{m}$-dependent integral coefficients for the self-dual two-form are given as:

$$\tilde{C}_4^B(1,2,3,4) = -\frac{4(m\tilde{m})^2(s-t)(st+2um\tilde{m})R^{(4)}}{stu^2}\,, \tag{4.101}$$

$$\tilde{C}_4^B(1,3,4,2) = -\frac{4(m\tilde{m})^2[s^2u+2t(s-t)m\tilde{m}]R^{(4)}}{st^2u}\,, \tag{4.102}$$

$$\tilde{C}_4^B(1,3,2,4) = \frac{4(m\tilde{m})^2[t^2u-2s(s-t)m\tilde{m}]R^{(4)}}{s^2tu}\,, \tag{4.103}$$

$$\tilde{C}_{3s}^B = -\frac{4(m\tilde{m})^2(s^2+3t^2+u^2)R^{(4)}}{t^2u^2}\,, \quad \tilde{C}_{3t}^B = \frac{4(m\tilde{m})^2(3s^2+t^2+u^2)R^{(4)}}{s^2u^2}\,, \tag{4.104}$$

$$\tilde{C}_{3u}^B = -\frac{8(m\tilde{m})^2u(s-t)R^{(4)}}{s^2t^2}\,, \tag{4.105}$$

where $R^{(4)} = \langle 13\rangle^2\langle 23\rangle^2[12]^2$. In this special case, the coefficients of different loop species have a simple relation:

$$5\tilde{C}^B = -20\tilde{C}^\chi = -4\tilde{C}^\psi\,. \tag{4.106}$$

After integral reduction, the $m\tilde{m}$-dependent integral coefficients now becomes:

$$C^B_{2s} = -\frac{2s(12s^4 + 53s^3t + 91s^2t^2 + 75st^3 + 49t^4)R^{(4)}}{105t^3u^3}\,, \tag{4.107}$$

$$C^B_{2t} = \frac{2t(49s^4 + 75s^3t + 91s^2t^2 + 53st^3 + 12t^4)R^{(4)}}{105s^3u^3}\,, \tag{4.108}$$

$$C^B_{2u} = -\frac{2u^2(12s^3 - 11s^2t + 11st^2 - 12t^3)R^{(4)}}{105s^3t^3}\,, \tag{4.109}$$

$$C^B_{3s} = -\frac{4s^2(4s^6 + 23s^5t + 55s^4t^2 + 70s^3t^3 + 50s^2t^4 + 15st^5 + 7t^6)R^{(4)}}{105t^4u^4}\,, \tag{4.110}$$

$$C^B_{3t} = \frac{4t^2(7s^6 + 15s^5t + 50s^4t^2 + 70s^3t^3 + 55s^2t^4 + 23st^5 + 4t^6)R^{(4)}}{105s^4u^4}\,, \tag{4.111}$$

$$C^B_{3u} = -\frac{4u^3(4s^5 - s^4t + s^3t^2 - s^2t^3 + st^4 - 4t^5)R^{(4)}}{105s^4t^4}\,, \tag{4.112}$$

$$C^B_4(1,2,3,4) = -\frac{8s^2t^2(s-t)R^{(4)}}{105u^4}\,, \quad C^B_4(1,3,4,2) = -\frac{2s^2u^2(4s+3t)R^{(4)}}{105t^4}\,, \tag{4.113}$$

$$C^B_4(1,3,2,4) = \frac{2t^2u^2(3s+4t)R^{(4)}}{105s^4}\,, \qquad 5C^B = -20C^\chi = -4C^\psi\,. \tag{4.114}$$

Similarly, we can also computed the configuration $(1_{11\dot{2}\dot{2}}\,, 2_{22\dot{1}\dot{1}}\,, 3_{22\dot{2}\dot{2}}\,, 4_{11\dot{1}\dot{1}})$ and the parity-odd integral coefficients of self-dual 2-form loops are

$$\tilde{C}^B_4(1,2,3,4) = -\frac{R^{(4)}}{t^2u^4}[8u^2(t-u)(m\tilde{m})^3 + 2su(8s^2 + 17su + 8u^2)(m\tilde{m})^2 \tag{4.115}$$

$$+ 4st^3(s-u)m\tilde{m} - s^2t^4]\,, \tag{4.116}$$

$$\tilde{C}^B_4(1,3,4,2) = -\frac{R^{(4)}}{t^4u^2}[8t^2(t-u)(m\tilde{m})^3 - 2st(8s^2 + 17st + 8t^2)(m\tilde{m})^2 \tag{4.117}$$

$$- 4su^3(s-t)m\tilde{m} + s^2u^4]\,, \tag{4.118}$$

$$\tilde{C}^B_4(1,3,2,4) = -\frac{2(t-u)(s+4m\tilde{m})(m\tilde{m})^2R^{(4)}}{t^2u^2}\,, \tag{4.119}$$

$$\tilde{C}^B_{3s} = -\frac{2R^{(4)}}{t^4u^4}[4tu(t-u)(2t^2 + 3tu + 2u^2)(m\tilde{m})^2 \tag{4.120}$$

$$+ 2s(t-u)(2t^4 + 5t^3u + 5t^2u^2 + 5tu^3 + 2u^4)m\tilde{m} + s^2(t^5 - u^5)]\,, \tag{4.121}$$

$$\tilde{C}^B_{3t} = \frac{2[4u(2t-u)(m\tilde{m})^2 + 2t^2(2s-u)m\tilde{m} - st^3]R^{(4)}}{tu^4}\,, \tag{4.122}$$

$$\tilde{C}^B_{3u} = \frac{2[4t(t-2u)(m\tilde{m})^2 - 2u^2(2s-t)m\tilde{m} + su^3]R^{(4)}}{t^4u}\,, \tag{4.123}$$

$$\tilde{C}^B_{2s} = -\frac{(t-u)[s(2t^2 + tu + 2u^2) + 4(2t^2 + 3tu + 2u^2)m\tilde{m}]R^{(4)}}{t^3u^3}\,, \tag{4.124}$$

$$\tilde{C}^B_{2t} = -\frac{(s-t)(4m\tilde{m} - t)R^{(4)}}{tu^3}\,, \quad \tilde{C}^B_{2u} = \frac{(s-u)(4m\tilde{m} - u)R^{(4)}}{t^3u}\,, \tag{4.125}$$

where $R^{(4)} = \langle 13\rangle^4[14]^2$. After the integal reduction of $m\tilde{m}$ terms, the integral coefficients of self-dual 2-form become

$$
\begin{aligned}
C_4^B(1,2,3,4) &= -\frac{s^2(20t^3 - 14t^2u + 8tu^2 + 7u^3)R^{(4)}}{105u^5}\,, && (4.126)\\
C_4^B(1,3,4,2) &= \frac{s^2(7t^3 + 8t^2u - 14tu^2 + 20u^3)R^{(4)}}{105t^5}\,, && (4.127)\\
C_4^B(1,4,2,3) &= -\frac{(7t^3 + t^2u - tu^2 - 7u^3)R^{(4)}}{105s^3}\,, && (4.128)\\
C_{3s}^B &= \frac{2(20t^9 + 26t^8u + 9t^6u^3 + 15t^5u^4 - 15t^4u^5 - 9t^3u^6 - 26tu^8 - 20u^9)R^{(4)}}{105t^5u^5}\,, && (4.129)\\
C_{3t}^B &= \frac{2t(20t^6 + 66t^5u + 72t^4u^2 + 35t^3u^3 + 47t^2u^4 + 61tu^5 + 35u^6)R^{(4)}}{105s^3u^5}\,, && (4.130)\\
C_{3u}^B &= -\frac{2u(35t^6 + 61t^5u + 47t^4u^2 + 35t^3u^3 + 72t^2u^4 + 66tu^5 + 20u^6)R^{(4)}}{105s^3t^5}\,, && (4.131)\\
C_{2s}^B &= -\frac{2(30t^6 - 11t^5u + 11tu^5 - 30u^6)R^{(4)}}{105t^4u^4}\,, && (4.132)\\
C_{2t}^B &= \frac{2(30t^4 + 59t^3u + 21t^2u^2 - 12tu^3 + 14u^4)R^{(4)}}{105s^2u^4}\,, && (4.133)\\
C_{2u}^B &= -\frac{2(14t^4 - 12t^3u + 21t^2u^2 + 59tu^3 + 30u^4)R^{(4)}}{105s^2t^4}\,. && (4.134)
\end{aligned}
$$

The contributions for chiral-fermion and chiral graviton are calculated analogously. Thus we have the parity-odd contribution of the chiral-two from, up to the rational term in Eq. (4.99). We will now obtain it using locality.

4.6.3 The rational term and gravitational anomaly

We will now fix our rational term by enforcing locality on the cut constructed piece of the amplitude. However, before doing so, it is well known that gravity theories in $4k+2$-dimensions can develop gravitational anomalies when coupled to chiral [56]. Recall, once more, that the fingerprint of anomalies in on-shell amplitudes is the tension between enforcing locality and unitarity on the full amplitude. We must see the presence of anomalies in our attempt to construct the rational terms. Let us consider the various factorization limit at four-points (non-singular terms are dropped). The factorization limit of $(1_{12\dot{1}\dot{2}}, 2_{22\dot{1}\dot{1}}, 3_{22\dot{2}\dot{2}}, 4_{11\dot{2}\dot{2}})$ is

$$
\begin{aligned}
M_{4,\,Cut}^{1-\mathrm{L,odd},B}\Big|_{s\to 0} &= -\frac{u^2(5s+4u)R^{(4)}}{315s^3}\,,\\
M_{4,\,Cut}^{1-\mathrm{L,odd},B}\Big|_{u\to 0} &= \frac{2t(4t^2+6tu+u^2)R^{(4)}}{315u^3}\,,\\
M_{4,\,Cut}^{1-\mathrm{L,odd},B}\Big|_{t\to 0} &= \frac{u^2(5t+4u)R^{(4)}}{315t^3}\,. && (4.135)
\end{aligned}
$$

The presence of higher-order poles is a clear violation of locality, and needs to be rectified by additional rational terms. However, with a little thought, it is simple to see that there are no gauge invariant rational terms to cancel the degree-3 poles. To see this, it is worthwhile to note that the parity-odd gauge invariant rational terms for Yang-Mills theory constructed so far are nicely built from the following gauge invariant combination:

$$F_i \wedge F_j \wedge F_k\,, \quad (\epsilon_i \cdot k_j - \epsilon_j \cdot k_i)\,, \quad (s_{ij}\; \epsilon_i \cdot \epsilon_j - 2(k_i \cdot \epsilon_j)(k_j \cdot \epsilon_i))\,, \tag{4.136}$$

where the last term is added for completeness of the basis. At linear level, we can factorize the gravity polarization tensor as $\epsilon_{\mu\nu} \to \epsilon_\mu \epsilon_\nu$, and thus an acceptable parity-odd rational term must be given by a sum of terms with single factor of $F_i \wedge F_j \wedge F_k$ multiplied with other invariants. Finally, assuming that the loop-amplitude is given by a covariant integrand, through integral reduction, one can deduce that the highest power of poles for the rational term must be at most one-less than that of the scalar-integral coefficients. Now since the integral coefficients have at most fourth-order poles, the rational function must have at most third-order poles. Furthermore, since in the four-dimensional representation, we have three-scalars and a graviton, the only terms involving polarization vectors that are non-vanishing when they are projected into scalar states are $F_i \wedge F_j \wedge F_k$ and $\epsilon_j \cdot \epsilon_i$. Since there can only be one factor of the former, for a non-vanishing contribution one must include at least one power of $(s_{ij}\; \epsilon_i \cdot \epsilon_j - 2(k_i \cdot \epsilon_j)(k_j \cdot \epsilon_i))$, with only the first term in the parenthesis contributing if i and j are scalars. However, the presence of an additional factor s_{ij} means that the power of poles are further surpressed. Thus in summary we conclude that for the parity-odd gravity invariant rational term for helicity $(1_{12\dot{1}\dot{2}}\,, 2_{22\dot{1}\dot{1}}\,, 3_{22\dot{2}\dot{2}}\,, 4_{11\dot{2}\dot{2}}):$

Any factorization channel associated with two four-dimensional scalars on one-side can have at most degree-2 poles.

Thus there is no way for us to cancel the degree-3 poles in Eq. (4.135)! In other words, they have to cancel amongst themselves. Not surprisingly, the known anomaly-free combination:

$$21M^\chi - M^\psi + 4M^B \tag{4.137}$$

precisely achieves this! Note that the projection on to four-dimensional space, while simplifies the computation, loses the information of the full six-dimensional covariance. Thus in this simplified computation, we don't expect to obtain the full six-dimensional anomaly conditions, but we should see the solution to the anomaly conditions ensuring locality of the final result, and indeed it does.

In fact for the current case, the anomaly-free combination cancels all poles!

We can consider the helicity configuration of of $(1_{11\dot{2}\dot{2}}\,, 2_{22\dot{1}\dot{1}}\,, 3_{22\dot{2}\dot{2}}\,, 4_{11\dot{1}\dot{1}}):$

$$
\begin{aligned}
M^{1-\mathrm{L,odd}\psi}_{4,\,Cut}\Big|_{s\to 0} &= \frac{s+2t}{84s^2}R^{(4)}\,,\\
M^{1-\mathrm{L,odd}\psi}_{4,\,Cut}\Big|_{u\to 0} &= \frac{(100t^3-96t^2u-181tu^2-471u^3)R^{(4)}}{2520u^4}\,,\\
M^{1-\mathrm{L,odd}\psi}_{4,\,Cut}\Big|_{t\to 0} &= \frac{(471t^3+181t^2u+96tu^2-100u^3)R^{(4)}}{2520t^4}\,,\\
M^{1-\mathrm{L,odd},B}_{4,\,Cut}\Big|_{s\to 0} &= -\frac{s+2t}{105s^2}R^{(4)}\,,\\
M^{1-\mathrm{L,odd},B}_{4,\,Cut}\Big|_{u\to 0} &= \frac{(93t^3-22t^2u+12tu^2+40u^3)R^{(4)}}{1260t^4}\,,\\
M^{1-\mathrm{L,odd},B}_{4,\,Cut}\Big|_{t\to 0} &= -\frac{(40t^3+12t^2u-22tu^2+93u^3)R^{(4)}}{1260u^4}\,,\\
M^{1-\mathrm{L,odd},\chi}_{4,\,Cut}\Big|_{s\to 0} &= \frac{s+2t}{420s^2}R^{(4)}\,,\\
M^{1-\mathrm{L,odd},\chi}_{4,\,Cut}\Big|_{u\to 0} &= \frac{(20t^3+48t^2u+31tu^2+u^3)R^{(4)}}{2520u^4}\,,\\
M^{1-\mathrm{L,odd},\chi}_{4,\,Cut}\Big|_{t\to 0} &= -\frac{(t^3+31t^2u+48tu^2+20u^3)R^{(4)}}{2520t^4}\,.
\end{aligned}
\tag{4.138}
$$

Again the anomaly-free combination in Eq. (4.137) precisely cancels the degree-4 poles. The remaining degree-3 and lower poles can be completely canceled by the following invariant rational term:

$$
R = \frac{-4t^5-4t^4u+t^3u^2-t^2u^3+4tu^4+4u^5}{10t^3u^3}R^{(4)}\,. \tag{4.139}
$$

Thus with the above rational term, we've obtained a unitary and local parity-odd one-loop four-point amplitude. Of crucial importance is the fact that the same cancellation occurred in distinct helicity configuration, thus compensating the fact six-dimensional covariance was not manifest by identifying a particular four-dimensional sub-plane.

4.7 Conclusions

In this chapter, we have constructed the parity-odd one-loop amplitudes for chiral-gauge and gravity theories using unitarity-methods, which constructs the integrand entirely from on-shell building blocks that are ignorant of any notion of gauge symmetry. We show that the constraints of unitarity and locality necessarily introduce new factorization channels in the one-loop amplitude. This only occurs in the parity-odd sector of the amplitude, which would be absent for non-chiral theories.

Consistency of the theory then relies on whether or not this new channel can be interpreted as an exchange of physical fields. For gauge theories, the absence of such new factorization channel imposes constraints on color factors that are exactly that of anomaly cancellation, the vanishing of the symmetric trace with $D/2+1$ generators. For group theory factors that do not satisfy this constraint, the new factorization channel is here to stay. In four-dimensions, we show that the residue of this new factorization pole cannot correspond to the exchange of physical fields, and thus the theory is inconsistent. In six-dimensions, the residue is exactly the exchange of a two-form between gluons (photons in QED). Thus if the symmetrized trace factorizes, then consistency is achieved and the new factorization channel automatically reveal a new particle in the spectrum, the Green-Schwarz two-form.

This tension between unitarity and locality is in fact isolated in a unique part of the amplitude, the cut-free rational terms. Indeed whilst unitarity method automatically gives an S-matrix that respects unitarity, there are potential violations of locality, and it is the job of the rational terms to restore locality. The new feature in chiral theories is that the requisite rational terms come with their own baggage. More precisely, they hold locality ransom: if we want locality back, the price to pay is to enlarge our spectrum. In the traditional Feynman diagram approach, such rational terms can be identified with the combination of the anomalous rational term from the one-loop Feynman diagrams and the tree-contribution of the Green-Schwarz mechanism, leaving behind a gauge invariant remnant. This remnant is an afterthought in the traditional literature, however from the S-matrix point of view, it is of upmost importance and anomaly cancellation is simply an "off-shell" way of producing the requisite rational term. An interesting question is whether Feynman-diagrams are the unique "off-shell" way of obtaining them. We will come back to this shortly.

Finally we also extend our analysis to gravity coupled to chiral-matter. We study the parity-odd one-loop four-graviton amplitude with chiral fermions, two-forms and gravitinos in the loop. The case of chiral two-form is very interesting as no covariant action exists for gravity coupled to arbitrary number of these fields. It is thus interesting to see if unitarity methods can be employed to construct the correct loop-level amplitudes for such theories. A non-trivial test would be if the constraint of locality for the one-loop amplitude, derived though unitarity methods, exactly corresponds to that of gravitational anomaly cancellation. Indeed we demonstrate that the parity-odd four-graviton amplitude, which is again unique for chiral-matter, contains spurious singularities that cannot be canceled by any invariant rational terms. Thus for consistency, such singularities must cancel between different chiral-matter loops. We have demonstrated that the known anomaly free combination indeed achieves this cancellation.

There are many interesting extensions. In the analysis of chiral-gauge theory, the absence of UV-divergences and spurious poles strongly constrain the scalar integral coefficients. In particular in $D = 2k$ ($k > 2$) dimensions, the parity-odd piece of the $(k+1)$-point amplitude can be expanded on the basis of scalar $(k+1)$-, k- and $(k-1)$-gon integrals. The k- and $(k-1)$-gon integral have ultraviolet divergences: their coefficients must be such that the overall divergence cancels. The $(k+1)$-gon integral has spurious singularities, which, when combined with that of the two lower-gons, must be such that it cancels with that of the gauge invariant rational term that is completely determined by Eq. (4.73). Indeed, it can be shown that using such constraint, one can fully determined the six-dimensional parity odd amplitude. Recently, it has been shown that one can utilize differential equations [63] to capture the behavior of integrals near singularities without explicitly integrating the full answer. Thus it is perceivable that the interrelation of the integral coefficients can be simplified using such differential equations.

The Adler-Bardeen theorem states that anomalies are one-loop exact [64]. In our language this is tantamount to the statement that only at one-loop can the rational terms introduce new singularities that violate locality. In our current study this fact is completely not transparent. In $D = 6$ dimensions, we have shown that the first non-trivial gauge invariant rational term factorizes completely into known three-point amplitudes. It should then be possible to construct a recursion relation for the rational terms. For supersymmetric theories, it is also possible to apply recursion relations for the planar loop-amplitude [65, 41, 66]. Through the lens of these two recursion relations, it will be extremely appealing to see how Adler-Bardeen theorem arises for supersymmetric chiral-Yang-Mills theories, i.e. $\mathcal{N} = (1, 0)$ in six-dimensions and maximal SYM in ten-dimensions.

In both unitarity methods and Feynman diagrams, the parity-odd gauge invariant rational term is obtained by adding extra terms to the amplitude. For the former, it is an extra term that is required for locality, while for the latter, it is obtained by adding an extra contribution from the Green-Schwarz mechanism. One might ask if there is any way in which one directly obtains the full gauge invariant amplitude in one go. One possibility is applying color-kinematic duality to chiral gauge and gravity theories. It would be interesting to see if the duality satisfying numerator, satisfying certain constrains, automatically gives the invariant rational term [67].

Finally it is well known that gauge-anomalies are related to global anomalies. The presence of global anomalies in supergravity amplitudes has been recently discussed in detail for $\mathcal{N} = 4$ supergravity [50]. Again here the symmetry violating amplitude appears as rational functions. It has been proposed that such anomaly controls the structure of the UV-divergence of $\mathcal{N} = 4$ supergraivty [68], and it is conceivable that the effect of such global anomalies in multi-loop amplitudes, is similar to the effect of gauge anomalies, which is governed by the Adler-Bardeen theorem. Thus it will be interesting to establish the connection at the level of the amplitude.

4.8 Acknowledgements

It is a pleasure to thank Nima Arkani-Hamed, Lance Dixon, Johannes Henn and Henrik Johansson for discussions. Y H. is supported by the Department of Energy under contract DE-SC0009988. W-M. C. thanks National Science Council and NCTS of Taiwan and NTU CTS for support.

Chapter 5

Consistency conditions on four point amplitudes

5.1 Consistency conditions on massless S-matrices

Pioneering work by Weinberg showed that simultaneously imposing Lorentz-invariance and unitarity, while coupling a hard scattering process to photons, necessitates *both* charge conservation and the Maxwell equations [69]. Similarly, he showed the same holds for gravity: imposing Lorentz-invariance and unitarity on hard scattering processes coupled to gravitons implies both the equivalence principle and the Einstein equations [70]. Weinberg's theorems were then extended to fermions in [71], where it was shown that spin-3/2 particles lead to supersymmetry. In the case of higher spin theories [72], which are closely related to string theory [73], it was shown that unitarity and locality impose severe restrictions, and many no-go theorems were established [74][75], more recently including in conformal field theories (CFTs) [76][77].

The goal of this chapter is to systematize and extend the previous analysis of Refs. [78] and [79] of the leading-order interactions between any set of massless states in four dimensions, within the context of the *on-shell* perturbative S-matrix. In short, our results are: (1) a new classification of three-particle amplitudes in constructible massless S-matrices, (2) ruling out all S-matrices built from three-point amplitudes with $\sum_{i=1}^{3} h_i = 0$ (other than ϕ^3-theory), (3) a new on-shell proof of the uniqueness of interacting gravitons and gluons, (4) development of a new test on four-particle S-matrices, and (5) showing how supersymmetry naturally emerges from consistency constraints on certain four-particle amplitudes which include spin-3/2 particles.

Massless vectors (gluons and photons) and tensors (gravitons) are naturally described via on-shell methods [54][78][16][79][80]. On- and off-shell descriptions of these massless higher-spin states are qualitatively different: on-shell they have only transverse polarization states, while off-shell all polarization states may be accessed. Local field theory descriptions necessarily introduce these unphysical, longitudinal, polarization states. They must be removed through introducing extra constraints which "gauge them away". Understanding consistency conditions on the interactions of massless gravitons and gluons/photons, should therefore stand to benefit from moving more-and-more on-shell, where gauge-invariance is automatic.

"Gauge anomalies", discussed in chapters 3 and 4, provide a recent example [81][46]. What is called a "gauge-anomaly" in off-shell formulations, on-shell is simply a tension between parity-violation and locality. Rational terms in parity-violating loop amplitudes either do not have local descriptions, or require the Green-Schwarz two-form to restore unitarity to the S-matrix.

Along these lines, there are recent, beautiful, papers by Benincasa and Cachazo [78], and by Schuster and Toro [79] putting these consistency conditions more on-shell. Ref. [78] explored the constraints imposed on a four-particle S-matrix through demanding consistent on-shell factorization in various channels on the coupling constants in a given theory [20].[1] Four-particle tests based on BCFW have been used to show the inconsistency of some higher spin interactions in Refs. [83][84], where it was suggested that non-local objects must be included in order to provide consistent theories (see also [85]).

The analysis of Ref. [78] hinged upon the existence of a "valid" BCF-shift,

$$A_n(z) = A_n(\{p_i + qz, p_j - qz\}) \text{ , with } p_i \cdot q = p_j \cdot q = 0 \text{ and } q \cdot q = 0,$$

of the amplitude that does not have a pole at infinity. If $A(z)$ does not have a large-z pole, then its physical value, $A(0)$, is a sum over its residues at the finite-z poles. These finite-z poles are factorization channels; their residues are themselves products of lower-point *on-shell* amplitudes: an on-shell construction of the whole S-matrix [20][41][**?**]. However, then extant existence proofs for such shifts resorted to local field theory methods [20][25][28].

Hence, ref. [79] relaxes this assumption, through imposing a generalized notion of unitarity, which they refer to as "complex factorization". The consequences of these consistency conditions are powerful. For example, they uniquely fix (1) the equivalence of gravitational couplings to all matter, (2) decoupling of multiple species of gravitons within S-matrix elements, and (3) the Lie

[1] This consistency condition is further explored, for instance, in ref. [82].

Algebraic structure of spin-1 interactions.

This chapter is organized as follows. We begin in section 5.2 by developing a useful classification of all on-shell massless three-point amplitudes. We here pause to make contact with standard terminology for "relevant", "marginal" and "irrelevant" operators in off-shell formulations of Field Theory, and to review basic tools of the on-shell S-matrix.

This is applied in section 5.3, first, to constructible four-particle amplitudes in these theories. Locality and unitarity sharply constrain the analytic structure of scattering amplitudes. Specifically, four-point amplitudes cannot have more than three poles. Simple pole-counting, using the classification system in section 5.2, rules out all lower-spin theories, save ϕ^3-theory, (super) Yang-Mills theory (YM and SYM), and General Relativity or Supergravity (GR or SUGRA)–and one pathological example, containing the interaction vertex $A_3(\frac{1}{2}, -\frac{1}{2}, 0)$. In section 5.4, we show that the gluons can only consistently interact via YM, GR and the higher-spin amplitude $A_3(1,1,1)$. Similarly, gravitons can only interact via GR and the higher-spin amplitude $A_3(2,2,2)$. Gravitons and gluons are unique, and cannot couple to higher-spin states. These two sections strongly constrain the list of possible interacting high-spin theories, in accordance with existing no-go theorems.

Utilizing the information in sections 5.3 and 5.4, in section 5.5 we derive and apply a systematic four-particle test, originally discussed in Ref. [86]. This test independently demonstrates classic results known about S-matrices of massless states, such as the equivalence principle, the impossibility of coupling gravitons to massless states with $s > 2$ [87], decoupling of multiple spin-2 species, and the Lie Algebraic structure of vector self-interactions.

Knowing the equivalence principle, in section 5.6 we then study the consistency conditions of S-matrices involving massless spin-3/2 states. From our experience with supersymmetry, we should expect that conserved fermionic currents correspond to massless spin-3/2 states. In other words, we expect that a theory which interacts with massless spin-3/2 particles should be supersymmetric.

Supersymmetry manifests itself through requiring all poles within four-point amplitudes have consistent interpretations. The number of poles is fixed, mandated by locality and unitarity and the mass-dimension of the leading-order interactions. Invariably, for S-matrices involving external massless spin-3/2 particles, at least one of these poles begs for inclusion of a new particle into the spectrum, as a propagating internal state on the associated factorization channel. For these S-matrices to be consistent, they *require* both gravitons to be present in the spectrum [88] and supersymmetry. We close with future directions in section 5.7.

5.2 Basics of on-shell methods in four-dimensions

In this section, we briefly review three major facets of modern treatments of massless S-matrices: the spinor-helicity formalism, kinematic structure of three-point amplitudes in these theories, and the notion of constructibility. The main message is three-fold:

- The kinematic dependence of three-point on-shell amplitudes is uniquely fixed by Poincare invariance (and is best described with the spinor-helicity formalism).
- On-shell construction methods, such as the recursion due to Britto, Cachazo, Feng and Witten (BCFW-recursion), allow one to recursively build up the entire S-matrix from these on-shell three-point building blocks. Amplitudes constructed this way are trivially "gauge-invariant". There are no gauges.
- Any pole in a local and unitary scattering amplitude must both (a) be a simple pole in a kinematical invariant, e.g. $1/K^2$, and (b) have a corresponding residue with a direct interpretation as a factorization channel of the amplitude into two sub-amplitudes.

5.2.1 Massless asymptotic states and the spinor-helicity formalism

In a given theory, scattering amplitudes can only be functions of the asymptotic scattering states. Relatively few pieces of information are needed to fully characterize an asymptotic state: momentum, spin, and charge/species information. Spinor-helicity variables automatically and fully encode both momentum and spin information for massless states in four-dimensions.

Four-dimensional Lorentz vectors map uniquely into bi-spinors, and vice versa (the mapping is bijective): $p_{\alpha\dot{\alpha}} = p_\mu \sigma^\mu_{\alpha\dot{\alpha}}$. Determinants of on-shell momentum bi-spinors are proportional to m^2. Bi-spinors of massless particles thus have rank-1, and must factorize into a product of a left-handed and a right-handed Weyl spinor: $p^2 = 0 \Rightarrow p^{\alpha\dot{\alpha}} = \lambda^\alpha \tilde{\lambda}^{\dot{\alpha}}$.

These two Weyl spinors λ and $\tilde{\lambda}$ are the spinor-helicity variables, and are uniquely fixed by their corresponding null-momentum, p, up to rescalings by the complex parameter z: $(\lambda, \tilde{\lambda}) \to (z\lambda, \tilde{\lambda}/z)$. Further, they transform in the (1/2,0) and (0,1/2) representations of the Lorentz group. Dot products of null momenta have the simple form, $p_i \cdot p_j = \langle ij \rangle [ji]$, where the inner-product of the (complex) left-handed (LH) spinor-helicity variables, is $\langle AB \rangle \equiv \lambda^A_\alpha \lambda^B_\beta \epsilon^{\alpha\beta}$, and the contraction of the right-handed (RH) Weyl spinors is $[AB] \equiv \tilde{\lambda}^A_{\dot{\alpha}} \tilde{\lambda}^B_{\dot{\beta}} \epsilon^{\dot{\alpha}\dot{\beta}}$.

A good deal of the power of the spinor-helicity formalism derives from the dissociation between the left-handed and right-handed degrees of freedom. Real null-momenta are defined by the relation,

$$\tilde{\lambda} = \bar{\lambda} \tag{5.1}$$

between the two Weyl-spinors. Complex momenta are not similarly bound: the left-handed and right-handed Weyl-spinors need not be related for complex momentum. For this reason, they can be independently deformed by complex parameters; this efficiently probes the analytic properties of on-shell amplitudes that depend on these variables. From here on out, we refer to the left-handed Weyl-spinors, i.e. the λs, as *holomorphic* variables; right-handed Weyl-spinors, i.e. the $\tilde{\lambda}$s are referred to as *anti-holomorphic* variables. Similarly, *holomorphic* spinor-brackets and *anti-holomorphic* spinor-brackets refer to $\langle\lambda, \chi\rangle$- and $[\tilde{\lambda}, \tilde{\chi}]$-contractions.

Identifying the ambiguity $(\lambda, \tilde{\lambda}) \to (z\lambda, \tilde{\lambda}/z)$ with little-group (i.e. helicity) rotations, $(\lambda, \bar{\lambda}) \to (e^{-i\theta/2}\lambda, e^{i\theta/2}\bar{\lambda})$, allows one to use the spinor-helicity variables to express not only the momenta of external states in a scattering process, but also their spin (helicity). In other words, the spinor-helicity variables encode all of the data needed to characterize massless asymptotic states, save species information.

5.2.2 Three-point amplitudes

Scattering processes involving three massless on-shell states have no non-trivial kinematical invariants. At higher-points, complicated functions of kinematical invariants exist that allow rich perturbative structure at loop level. These invariants are absent at three-points. Poincare invariance, up to coupling constants, thus uniquely and totally fixes the kinematical structure of all three-point amplitudes for on-shell massless states.

The standard approach to solving for the three-point amplitudes (see for example Ref. [78]) involves first writing a general amplitude as:

$$A_3 = A_3^{(\lambda)}(\langle 12\rangle, \langle 23\rangle, \langle 31\rangle) + A_3^{(\tilde{\lambda})}([12], [23], [31]) \tag{5.2}$$

where (λ) denotes exclusive dependence on holomorphic spinors, and $(\tilde{\lambda})$ denotes the same for anti-holomorphic spinors. Imposing momentum conservation forces $[12] = [23] = [31] = 0$, and/or $\langle 12\rangle = \langle 23\rangle = \langle 31\rangle = 0$. Typically, only one of the two functions in Eq. (5.2) is smooth in this limit, and is thus selected as the physical one, while the other is discarded.

Explicitly, in these cases, the amplitudes become:

$$
\begin{aligned}
A_3(1_a^{h_1}, 2_b^{h_2}, 3_c^{h_3}) &= g_{abc}^{-} \langle 12 \rangle^{h_3-h_1-h_2} \langle 23 \rangle^{h_1-h_2-h_3} \langle 31 \rangle^{h_2-h_3-h_1} \text{ , for } \sum_{i=1}^{3} h_i < 0 \, , \\
A_3(1_a^{h_1}, 2_b^{h_2}, 3_c^{h_3}) &= g_{abc}^{+} [12]^{h_1+h_2-h_3} [23]^{h_2+h_3-h_1} [31]^{h_3+h_1-h_2} \quad \text{ , for } \sum_{i=1}^{3} h_i > 0 \, ,
\end{aligned} \tag{5.3}
$$

where $g_{abc}^{\pm}$ is the species dependent coupling constant.

However, this approach leads to ambiguities in the $\sum_{i=1}^{3} h_i = 0$ case. Consider for example a three-point interaction between two opposite-helicity fermions and a scalar. Equation (5.2) reads in this case:

$$
A_3\left(1^0, 2^{-\frac{1}{2}}, 3^{\frac{1}{2}}\right) = g^- \frac{\langle 12 \rangle}{\langle 13 \rangle} + g^+ \frac{[13]}{[12]} \tag{5.4}
$$

Imposing momentum conservation, for example by setting $\langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0$, is clearly ill-defined[2]. Because of this ambiguity, $\sum_{i=1}^{3} h_i = 0$ amplitudes have generally been ignored in most of the on-shell literature. However, the ambiguity is only superficial.

The inconsistencies arise because we first find the most general eigenfunction of the helicity operator, ie. Eq. (5.2), and only after that do we impose momentum conservation. However, this order of operations is arbitrary. Since we always only deal with on-shell amplitudes, we can simply first fix for example $\langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0$, and then look for solutions which are functions only of $\tilde{\lambda}$s. In this case, the amplitudes are perfectly well defined as:

$$
A_3 \quad = g_{abc}^{-} f^{-}(\lambda_i), \quad \text{when } [12] = [23] = [31] = 0 \tag{5.5}
$$

and

$$
A_3 \quad = g_{abc}^{+} f^{+}(\tilde{\lambda}_i), \quad \text{when } \langle 12 \rangle = \langle 23 \rangle = \langle 31 \rangle = 0 \tag{5.6}
$$

Ultimately, it will in fact turn out that none of these amplitudes are consistent with locality and unitarity, but this approach clears any ambiguities related to $\sum_{i=1}^{3} h_i = 0$ amplitudes.

Before moving on, we pause to consider the role of parity in the on-shell formalism. Parity conjugation swaps the left-handed and right-handed $SU(2)$s that define the (double-cover) of the four-dimensional Lorentz-group. As such, parity swaps the left-handed Weyl-spinors with the right-handed Weyl-spinors, $(1/2, 0) \leftrightarrow (0, 1/2)$. Therefore, within the spinor-helicity formalism, in the

[2]Attempting to impose momentum conservation by a well-defined limit leads to other inconsistencies as well, for example with the helicity operator.

context of Eq. (5.3),

$$\begin{aligned} g^-_{abc} = g^+_{abc} &\Longleftrightarrow \text{Parity}-\text{conserving interactions}\ ,\ \text{and} \\ g^-_{abc} = -g^+_{abc} &\Longleftrightarrow \text{Parity}-\text{violating interactions}\,. \end{aligned} \tag{5.7}$$

Further, as we associate the right-handed Weyl-spinors, i.e. the λs, with holomorphic degrees of freedom and the left-handed Weyl-spinors, i.e. the $\tilde{\lambda}$s, with anti-holomorphic degrees of freedom, we see that parity-conjugation swaps the holomorphic and anti-holomorphic variables. In other words, parity- and complex- conjugation are one-and-the-same. The conjugate of a given three-point amplitude is the same amplitude with all helicities flipped: the "conjugate" of $A_3(1^{+h_1}, 2^{+h_2}, 3^{+h_3})$ is $A_3(1^{-h_1}, 2^{-h_2}, 3^{-h_3})$.

We will find it useful to classify all such three-particle amplitudes by two numbers:

$$A = \left|\sum_{i=1}^{3} h_i\right|,\ H = \max\left\{|h_1|, |h_2|, |h_3|\right\}. \tag{5.8}$$

Comparing the relevant operator in ϕ^3-theory to its corresponding primitive three-point amplitude, we infer that three-point amplitudes with $A = 0$ correspond to *relevant* operators. Similarly, QCD's $A = 1$ three-point amplitude corresponds to *marginal* operators; GR has $A = 2$, and interacts via irrelevant, $1/M_{pl}$ suppressed, operators.

5.2.3 Four points and higher: Unitarity, Locality, and Constructibility

There are several, complimentary, ways to build up the full S-matrix of a theory, given its fundamental interactions. Conventionally, this is through Feynman diagrams, the work-horse of any perturbative analysis of a given field theory. However, this description of massless vector- (and higher-spin-) scattering via local interaction Lagrangians necessarily introduces unphysical, longitudinal, modes into intermediate expressions [16][86]. To project out these unphysical degrees of freedom, one must impose the gauge conditions.

On the other hand, recent developments have elucidated methods to obtain the full S-matrix, while keeping *all* states involved on-shell (and physical) throughout the calculation [78][79][54][20][41][89]. We refer to these methods, loosely speaking, as "constructive". Crucially, because all states are on-shell, all degrees of freedom are manifest, thus: *amplitudes that are directly constructed through on-shell methods are automatically gauge-invariant.* This simple fact dramatically increases both (a) the computational simplicity of calculations of scattering amplitudes, and (b) the physical transparency of the final results.

The cost is that amplitudes sewn together from on-shell, delocalized, asymptotic states do not appear to be manifestly local. Specifically, at the level of the amplitude, locality is reflected in the pole-structure of the amplitude. Scattering amplitudes in local theories have exclusively propagator-like, $\sim 1/K^2$, poles ($K = \sum_i p_i$ is a sum of external null momenta). Non-local poles correspond to higher-order poles, i.e. $1/(K^2)^4$, and/or poles of the form, $1/\langle i|K|j]$, where K is a sum of external momenta.[3] An on-shell S-matrix is local if its only kinematical poles are of the form $1/(\sum_i p_i)^2$.

Unitarity, as well, has a slightly different incarnation in the on-shell S-matrix. In its simplest guise, unitarity is simply the dual requirement that (a) the residue on each and every pole in an amplitude *must* have an interpretation as a physical factorization channel,

$$A^{(n)} \to \frac{1}{K^2} A_L^{(n-m+1)} \times A_R^{(m+1)} \,, \tag{5.9}$$

and (b) that any individual factorization channel, if it is a legitimate bridge between known lower-point amplitudes in the theory, must be a residue of a fully legitimate amplitude with the same external states in the theory. For example, given a factorization channel of the form

$$A_3(1^{-2}, 2^{-2}, P_{12}^{+2}) \frac{1}{s_{12}} A_3(P_{12}^{-2}, 3^{+2}, 4^{+2}) \,, \tag{5.10}$$

within a theory *constructed* from the three-point amplitude $A_3(+2, -2, +2)$ and its parity-conjugate, *then* this must be a factorization channel of the four-point amplitude $A_4(1^{-2}, 2^{-2}, 3^{+2}, 4^{+2})$.

Poincare invariance uniquely fixes the three-particle S-matrix in a theory, up to coupling constants. Constructive methods, such as the BCFW recursion relations, use these fixed forms for the three-point amplitudes as input to build up the entire S-matrix, without making reference to Feynman diagrams [20][41]. Basic symmetry considerations, residue theorems, and judicious application of tree-level/single-particle unitarity, fix the entire S-matrix![4]

Before closing, we motivate the most famous on-shell construction of massless scattering amplitudes: BCFW-recursion. In it, two null external momenta, p_1^μ and p_2^μ, are deformed by a complex null-momentum, $z \times q^\mu$. The shift is such that (a) the shifted momenta $p_1(z) = p_1 + qz$ and $p_2(z) = p_2 - qz$ remain on-shell (possible, as momentum q^μ is complex), and (b) the total sum of external momenta remains zero.

[3] Indeed, individual terms within gluon amplitudes generated by, for instance, BCFW-recursion[20] contain "non-local" poles, specifically of this second type, $\sim 1/\langle i|K|j]$. These non-local poles, however, always cancel in the total sum, and the final expression is manifestly local [16][41].

[4] Invocations of "unitarity" in this chapter do not refer to the standard two-particle unitarity-cuts.

As tree amplitudes are rational functions of their external kinematical invariants with, at most, simple poles, this deformation allows one to probe the analytic pole structure of the deformed amplitude, $A^{\text{tree}}(z)$:

$$A^{\text{tree}}(z) \equiv A^{\text{tree}}(p_1^{h_1}(z), p_2^{h_2}(z), ...p_n^{h_n})\,. \tag{5.11}$$

Kinematical poles in $A^{\text{tree}}(z)$ are either un-shifted, or scale as $1/K^2 \rightarrow 1/(2z(q\cdot K)+K^2)$, if K includes only one of $\hat{p_1}$ or $\hat{p_2}$. Cauchy's theorem then gives a simple expression for the physical amplitude, $A^{\text{tree}}(z=0)$,

$$A^{\text{tree}}(z=0) = \sum_{z_P} \text{Res}\left\{\frac{A_4(z)}{z}\right\}\Bigg|_{z_P=-\frac{K^2}{2q\cdot K}} + \big(\text{Pole at } z\rightarrow\infty\big)\,. \tag{5.12}$$

Existence of such a BCFW-shift, in both Yang-Mills/QCD and in General Relativity, that dies off at least as quickly as $1/z$ for large-z can be elegantly shown through imposing complex factorization[79], and allows the entire on-shell S-matrix to be built up from three-point amplitudes. Existence of valid BCFW-shifts were originally shown within local formulations of field theory [20][90][25]. In section 5.5 we develop a shift at four-points which is guaranteed to die off for large-z by simple dimensional analysis.

5.3 Ruling out constructible theories by pole-counting

Pedestrian counting of poles, mandated by constructibility in four-point amplitudes, strongly constrains on-shell theories. The number of poles in an amplitude *must* be less than or equal to the number of accessible, physical, factorization channels at four points. Tension arises, because the requisite number of poles in a four-point amplitude *increases* with the highest-spin particle in the theory, while the number of possible factorization channels is *bounded* from above by three, the number of Mandelstam variables.

This tension explicitly rules out the following theories as inconsistent with constructibility, locality, and unitarity: (1) all relevant interactions (A = 0), save ϕ^3 and an "exotic" Yukawa-like interaction, (2) all marginal interactions (A = 1) save those in YM, QCD, Yukawa theory, and scalar QED, and another "exotic" interaction between spin-3/2 particles and gluons, and (3) all first-order irrelevant interactions (A = 2) save those in GR. Further consistency conditions later rule out those two unknown, pathological, relevant ($A=0$) and marginal ($A=1$) interactions.

Further, incrementally more sophisticated pole-counting sharply constrains highly irrelevant ($A > 2$) higher-spin amplitudes. Specifically, save for two notable examples, they cannot consistently couple either to gravitational interactions or to more conventional Yang-Mills theories or "gauge"-theories. This is the subject of section 5.4. It is somewhat striking that simply counting poles in this way so powerfully constrains the palate of three-point amplitudes which may construct local and unitary S-matrices. The results of this pole-counting exercise are succinctly summarized in Fig. 5.1.

5.3.1 The basic consistency condition

Explicitly we find that four-particle S-matrices constructed from primitive three-particle amplitudes are inconsistent with locality and unitarity if there are more than three poles in any given term in an amplitude. More specifically, the number of poles in the simplest amplitudes has to be *at least* $N_p = 2H + 1 - A$. Thus, a theory is necessarily inconsistent if

$$2H + 1 - A = N_p > 3 \iff \text{Number of poles} > \text{cardinality of } \{s, t, u\} \,. \tag{5.13}$$

Recall that, in accordance with Eq. (5.8), $A = |h_1 + h_2 + h_3|$ and $H = \max\{|h_1|, |h_2|, |h_3|\}$.

We prove constraint (5.13 below.[5] First, we note there are A total spinor-brackets in three-point amplitudes of the type in Eq. (5.3):

$$A_3(1^{h_1}, 2^{h_2}, 3^{h_3}) = \kappa_A [12]^c [13]^b [23]^a \Rightarrow a + b + c = \sum_{i=1}^{3} h_i = A > 0 \,. \tag{5.14}$$

Thus, on a factorization channel of a four-point amplitude, A_4, *constructed* from a given three-point amplitude multiplied by its parity conjugate amplitude, $A_3 \times \bar{A}_3$, there will be A net holomorphic spinor-brackets and A net anti-holomorphic spinor-brackets: A $\langle\,\rangle$s and A $[\,]$s. Therefore, generically on such a factorization channel, the mass-squared dimension of the amplitude is:

$$A_4 \to \frac{\kappa_A^2}{s_{\alpha\beta}} A_3 \times \bar{A}_3 \Rightarrow \left[\frac{A_4}{\kappa_A^2}\right] = (K^2)^{A-1} \,. \tag{5.15}$$

By locality, an amplitude may only have $1/K^2$-type poles. Therefore the helicity information, captured by the non-zero little-group weight of the spinor-products, can only be present in an overall *numerator* factor multiplying the amplitude. Four-point amplitudes thus naturally split into three parts: a numerator, N, which encodes helicities of the states, a denominator, $F(s, t, u)$, which

[5] For expediency, we defer discussion of one technical point, proof of Eq. (5.17), to subsection 5.3.4.

encodes the pole-structure, and the coupling constants, κ_A^2, which encode the species-dependent characters of the interactions (discussed in section 5.5):

$$A_4 = \kappa_A^2 \frac{N}{F(s,t,u)} \Rightarrow \left[\frac{N}{F(s,t,u)}\right] = (K^2)^{A-1} \ , \tag{5.16}$$

where the last equality is inferred from Eq. (5.15). We prove in subsection 5.3.4, that *minimal* numerators N which accomplish this goal are comprised of exactly $2H$ holomorphic and $2H$ anti-holomorphic spinor-brackets, *none* of which can cancel against any pole in $F(s,t,u)$:

$$N \sim \langle\,\rangle_{(1)}...\langle\,\rangle_{(2H)} \ \ [\,]_{(1)}...[\,]_{(2H)} \Rightarrow [N] = (K^2)^{2H} \ . \tag{5.17}$$

Thus, by (5.15), (5.16), and (5.17), we see

$$\begin{aligned}\left[\frac{A_4}{\kappa_A^2}\right] = \left[\frac{N}{F(s,t,u)}\right] &= (K^2)^{A-1}\, , \text{ and } [N] = (K^2)^{2H} \\ &\Rightarrow [F(s,t,u)] = (K^2)^{2H+1-A} \\ &\Rightarrow N_p = 2H + 1 - A\,.\end{aligned} \tag{5.18}$$

Constraint (5.13) naturally falls out from Eq. (5.18), after observing that there can be at most three legitimate, distinct, factorization channels in any four-point tree amplitude. This specific constraint, and others arising from pole-counting from minimal numerators, is extremely powerful. The catalogue of theories they together rule out are succinctly listed in Fig. 5.1. We explore the consequences of this constraint below.

5.3.2 Relevant, marginal, and (first-order) irrelevant theories ($A \leq 2$): constraints

To begin with, note that constraint (5.13 immediately rules out all theories with $N_p > 3$. Beginning with relevant, $A = 0$, interactions, we see that $N_p = 2H + 1 \leq 3 \Rightarrow H \leq 1$. Already this rules out relevant interactions between massless spin-3/2 and spin-2 states.

Next, argument by contradiction rules out relevant amplitudes involving massless vectors, e.g. the $(H, A) = (1, 0)$-theory. Consider such a relevant amplitude, for example $A_3(+1, -1/2, -1/2)$. It constructs a putative four-point amplitude with external vectors,

$$A_4\left(-1, -\frac{1}{2}, \frac{1}{2}, 1\right) \ . \tag{5.19}$$

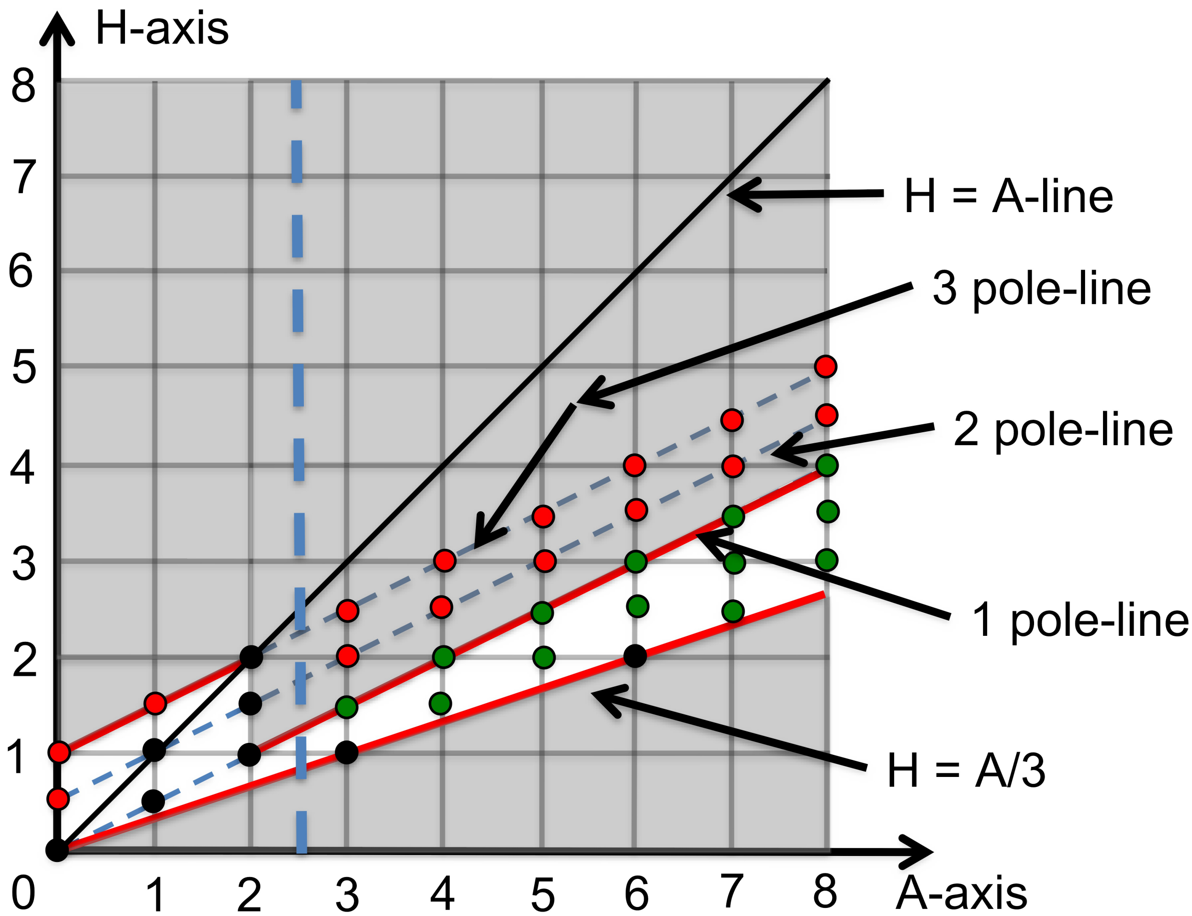


Figure 5.1: Summary of pole-counting results: with few notable exceptions (below) only three-point amplitudes on, or below, the $H = A$ line, for $A \leq 2$ can form tree-amplitudes that are consistent with locality and unitarity. Recall $N_p = 2H + 1 - A$, where $A = |\sum_{i=1}^{3} h_i|$ and $H = \max\{|h_i|\}$. (Color online.) Black-dots represent sets of three-point amplitudes that define self-consistent S-matrices that can couple to gravity; green-dots represent sets of three-point amplitudes which—save for two exceptions explicitly delineated in Eq. (5.45)—define S-matrices that *cannot* couple (in the sense defined in section 5.4) to any S-matrix defined by the black-dots; red-dots represent sets of three-point amplitudes that cannot ever form consistent S-matrices. Straightforward application of constraint (5.13, in subsection 5.3.2, rules out all A_3s with (H, A)-*above* the $N_p = 3$-line. More careful pole-counting, in subsection 5.3.3 and appendix 5.3.5, rules out all interactions above the $N_p = 1$ line, save for those with $(H, A) = (1/2, 0)$, $(1, 1)$, $(3/2, 2)$, and $(2, 2)$. Further, in section 5.4, a modified pole-counting rules out interaction between the $(H, A) = (2, 2)$ gravity theory and any other theory with a spin-2 particle, save the unique $(H, A) = (2, 6)$-theory. Similar results hold for gluon self-interactions: vectors present in any higher-spin amplitude with $A > 3$, save the unique $(H, A) = (1, 3)$-theory, cannot couple to the vectors interacting via leading $(H, A) = (1, 1)$ interactions. Section 5.5.4 rules out the $(H, A) = (1/2, 0)$-interaction. Amplitudes in the grey-shaded regions can never be consistent with locality and unitarity. Higher-spin, $A > 3$, amplitudes between the $H = A/2$ and $A = A/3$ lines may be consistent. However, they *cannot* be coupled either to GR or YM, save for $(H, A) = (1, 3)$ or $(2, 6)$. In section 5.6, we show inclusion of leading-order interactions between massless spin-3/2 states, at $A = 2$, promotes gravity to supergravity. Supergravity cannot couple to even these two $A > 2$ interactions, as seen in appendix 5.6.5.

This amplitude must have $2+1-0=3$ poles, each of which must have an interpretation as a *valid* factorization channel of the amplitude. So it must have valid s-, t-, and u-factorization channels, with *relevant* ($A=0$) three-point amplitudes on either side. However, on the $s\to 0$ pole, A_4 factorizes as,

$$A_4\left(-1,-\frac{1}{2},\frac{1}{2},1\right)\bigg|_{s\to 0}=\frac{1}{s}\bar{A}_3\left(-1,-\frac{1}{2},h\right)A_3\left(1,\frac{1}{2},-h\right)\,. \tag{5.20}$$

where h must be $3/2$ to make the interaction relevant. Thus, to be consistent with locality and unitarity, relevant vector couplings must also include spin-3/2 particles. But, as mentioned above, including these particles in the spectrum, and then taking them as external state invariably leads to too many poles. An identical argument shows that the remaining relevant vertex $A_3(+1,-1,0)$ requires spin 2 particles, again leading to an inconsistency. Thus all $(H,A)=(1,0)$ interactions are also ruled out. Thus the only admissible relevant three-point amplitudes are

$$A_3(0,0,0)\ ,\ \text{and}\ A_3\left(0,\frac{1}{2},-\frac{1}{2}\right)\,. \tag{5.21}$$

The first amplitude is the familiar one from ϕ^3-theory. We rule out the second amplitude in section 5.5.4.

Further, we see that marginal interactions cannot contain particles with helicities larger than 3/2. Directly, requiring $2H+1-A\le 3$ for $A=1$ forces $H\le 3/2$. $(H,1)$-type three-point amplitudes cannot build S-matrices consistent with locality and unitarity for $H>3/2$.

We rule out *marginal* $(H,A)=(3/2,1)$ amplitudes, i.e. marginal coupling to massless spin-3/2 states, using the same logic as above. This time, marginal amplitudes with external 3/2 particles require *all* three poles. Two factorization channels have consistent interpretations within the theory; however, the "third" channel does not. It necessitates exchange of a spin-2 state between the three-point amplitudes. But this violates constraint (5.13: marginal amplitudes with spin-2 states lead to amplitudes with four kinematic poles. Thus, the only admissible marginal three-point amplitudes are,

$$A_3\left(1,1,-1\right)\ ,\ A_3\left(1,\frac{1}{2},-\frac{1}{2}\right)\ ,\ A_3\left(1,0,0\right)\ ,\ \text{and}\ A_3\left(0,\frac{1}{2},\frac{1}{2}\right)\ , \tag{5.22}$$

and their conjugate three-point amplitudes. We refer to this set of three-point amplitudes, loosely, as "the $\mathcal{N}=4$ super Yang-Mills (SYM) interactions."

Finally, constraint (5.13 rules out leading-order gravitational coupling to particles of spin-$H>2$. Such three-point amplitudes, of the form $A_3(H,-H,\pm 2)$, have $A=2$ and $H>2$, and yield four-point amplitudes with $2H-1>3$ poles; this cannot be both unitary and local for $H>2$. Admissible $A=2$ amplitudes are restricted to:

$$A_3\left(2,2,-2\right)\,, \qquad A_3\left(2,\frac{3}{2},-\frac{3}{2}\right)\,, A_3\left(2,1,-1\right)\,, \; A_3\left(2,\frac{1}{2},-\frac{1}{2}\right)\,, \text{ and } A_3\left(2,\frac{1}{2},-\frac{1}{2}\right)\,, \tag{5.23}$$

$$A_3\left(\frac{3}{2},\frac{3}{2},-1\right)\,, \qquad A_3\left(\frac{3}{2},1,-\frac{1}{2}\right)\,, A_3\left(\frac{3}{2},\frac{1}{2},0\right)\,, \; A_3\left(1,\frac{1}{2},\frac{1}{2}\right)\,, \text{ and } A_3\left(1,1,0\right)\,, \tag{5.24}$$

and their conjugate three-point amplitudes. We refer to the amplitudes in (5.23) as "gravitational interactions." More generally, we refer to this full set of three-point amplitudes, loosely, as "the $\mathcal{N}=8$ SUGRA interactions."

5.3.3 Killing $N_p=3$ and $N_p=2$ theories for $A\geq 3$

It is relatively simple to show that any theory constructed from A_3s with $N_p=3$ poles, beyond $A=2$, cannot be consistent with unitarity and locality. To begin, we note that

$$\left\{N_p=3 \Longleftrightarrow 2H+1-A=3\right\} \Rightarrow H=A/2+1\,. \tag{5.25}$$

We label the helicities in the three-point amplitudes with $N_p=3$, as $A_3(H,g,f)$. Without loss of generality, we order them as $f\leq g\leq H=A/2+1$. As $A>2$, then $g+f$ must be positive: at a minimum $g>0$.

Now construct the four-point amplitude $A_4(H,-H,f,-f)$ from this three-point amplitude and its parity-conjugate. By assumption, this amplitude must have three poles, each of which must have an interpretation as a legitimate factorization channel within the theory constructed from $A_3(A/2+1,g,f)$ (or some mild extension of the theory/spectrum).

However, in order for the t-channel pole in the amplitude $A_4(H,-H,g,-g)$ to have a viable interpretation as a factorization channel, it requires a state with spin *greater* than $A/2+1=H$. Specifically, on this t-pole,

$$A_4(H,-H,g,-g)\Big|_{t\to 0}=\frac{1}{K_{14}^2}A_3\left(\frac{A+2}{2},-g,\frac{A+2}{2}+g\right)\bar{A}_3\left(-\frac{A+2}{2},g,-\frac{A+2}{2}-g\right) \tag{5.26}$$

By assumption, $g>0$: the intermediate state must have helicity $\tilde{H}=A/2+1+g$. Clearly this new state has helicity larger than $H=A/2+1$. A priori, there is no problem: new particles mandated by consistency conditions may be included into the spectrum of a theory without necessarily introducing inconsistencies. However, if *these* particles of spin $\tilde{H}>H=A/2+1$ are put as external states of the new three-point amplitudes in the modified theory, then these new four-point amplitudes will necessarily have $3 + 2g > 3$ poles, in violation of constraint (5.13).

Hence all theories constructed from A_3s with $N_p \geq 3$ and $A > 2$ are inconsistent. Similar arguments show that theories with $N_p = 2$ cannot be consistent for $A > 2$; they are however slightly more detailed, and involve several specific cases at low-A values. Proof of this extended claim is relegated to subsection 5.3.5.

5.3.4 Constructing minimal numerators

Here, we prove that the minimal numerator for a four-point amplitude of massless particles satisfies Eq. (5.17) for the special case where the sum of all four helicities vanishes.[6] Given $A_4(1^{h_1}, 2^{h_2}, 3^{h_3}, 4^{h_4})$, we re-label the external states by increasing helicity:

$$H_1 \geq H_2 \geq H_3 \geq H_4 \tag{5.27}$$

The total helicity vanishes, and thus $H_1 \geq 0$ and $H_4 \leq 0$. Now, define $H_1^+ = |H_1|$ and $H_4^- = |H_4|$: the numerator has at least $2H_1^+$ $\tilde{\lambda}$s, and at least $2H_4^-$ λs.

Now there must exist

$$N_{\tilde{\lambda}}^{\text{ext.}} = 2H_1^+ + N_{\tilde{\lambda}}^{\text{rest}} \geq 2H_1^+ \tag{5.28}$$

total external $\tilde{\lambda}$s in the numerator. Similarly, the numerator must contain a total of

$$N_{\lambda}^{\text{ext.}} = 2H_4^- + N_{\lambda}^{\text{rest}} \geq 2H_4^- \tag{5.29}$$

external λs.

By definition, the numerator is both (a) Lorentz-invariant, and (b) little-group covariant, and (c) encodes all of the helicity information of the asymptotic scattering states. Therefore, it must be of the form

$$\text{Numerator} \sim \langle\,\rangle_{(1)} ... \langle\,\rangle_{(n)} \;\; [\,]_{(1)} ... [\,]_{(m)} \;. \tag{5.30}$$

Notably, requiring $\sum_{a=1}^{4} H_a = 0$ directly implies that the numerator contains equal number of

[6] All four-point tree-amplitudes constructed from a set of three-point amplitudes and their conjugate amplitudes, A_3 & $\bar{A}_3$, have this property.

holomorphic and anti-holomorphic spinor-helicity variables,

$$\sum_{a=1}^{4} H_a = 0 \Leftrightarrow \left\{N_{\lambda}^{\text{ext.}} = N_{\tilde{\lambda}}^{\text{ext.}}\ ,\ \text{and}\ N_{\lambda}^{\text{total}} = N_{\tilde{\lambda}}^{\text{total}}\right\} \tag{5.31}$$

We note that because the numerator contains the same number of λs as $\tilde{\lambda}$s, and must be a product of spinor-brackets, then it must have the same number of each type of spinor-product:

$$\mathcal{N}_b = N_{\langle\ \rangle} = N_{[\ ]}\ . \tag{5.32}$$

The reason is as follows. First, note that only inner-products of spinor-helicity variables are both (a) Lorentz-invariant and (b) little-group covariant. Because the numerator has both of these properties, all of the λs and $\tilde{\lambda}$s which encapsulate the helicity information of the asymptotic scattering states must be placed within spinor-brackets. If we take $N_{\langle\ \rangle} \neq N_{[\ ]}$, then there would be a mis-match between the number of λs and $\tilde{\lambda}$s in the numerators. This contradicts the statement that the numerator must contain the same number of positive- and negative- chirality spinor-helicity variables. This proves Eq. (5.32).

Now, the *minimal* number of spinor-brackets $\mathcal{N}$ is simply given by

$$\mathcal{N}_b = 2 \times \max\{H_1^+, H_4^-\} \tag{5.33}$$

This can be seen in the following way. At the minimum, there must be $2H_1^+$ $[\ ]$s and $2H_4^-$ $\langle\ \rangle$s within the numerator. Otherwise, at least two of the $2H_1^+$ copies of $\tilde{\lambda}_1$ within the numerator would have to within the same spinor bracket, $[\tilde{\lambda}_1, \tilde{\lambda}_1]$. But this would force the numerator to vanish. As we are only concerned with non-trivial amplitudes, we thus require $N_{[\ ]} \geq 2H_1^+$. The same logic requires $N_{\langle\ \rangle} \geq 2H_4^-$.

But, by Eq. (5.32), we must have $N_{[\ ]} = N_{\langle\ \rangle}$. So we must have $\mathcal{N}_b \geq 2 \times \max\{H_1^+, H_4^-\}$. The minimal numerator saturates this inequality. This proves Eq. (5.33).

It is important to note that this minimal number of spinor-brackets of each type, $\mathcal{N}_b = 2\max\{H_1^+, H_4^-\}$, mandated by Eq. (5.33) to be present within the numerator is "large" enough to encode the helicity information of all of the external scattering states—not just the helicity information of $1^{+H_1^+}$ and $4^{-H_4^-}$.

In other words, there are "enough" $\langle\ \rangle$s and $[\ \ ]$s already present in the numerator to fit in the remaining λs and $\tilde{\lambda}$s required to encode the helicity information of the other two particles. I.e.,

$$N_{[\ ]} \geq N_{\tilde{\lambda}}^{\text{rest}}\ ,\ \text{and}\ N_{\langle\ \rangle} \geq N_{\lambda}^{\text{rest}} \tag{5.34}$$

Before proving this, first recall Eqs. (5.28), (5.29), and (5.33). By (5.33), $N_{[\ ]} = N_{\langle\ \rangle} = 2\text{max}\{H_1^+, H_4^-\}$. Now, how many λs and $\tilde{\lambda}$s must be present in the numerator to ensure all external helicity data is properly entered into the numerator? There are only three cases to consider. In all cases, Eq. (5.34) holds:

1. Particles 1 and 2 have positive helicity, while particles 3 and 4 have negative helicity. Now, by definition, we would like to show that of the $2\text{max}\{H_1^+, H_4^-\}$ [ ]s required by Eq. (5.33) are sufficiently numerous to allow inclusion of $2|H_2|$ more $\tilde{\lambda}_2$s. This is guaranteed by the orderings: $H_1 \geq H_2$. So there are enough empty slots in the anti-holomorphic spinor-brackets to encode the helicity of all positive-helicity particles. The same holds for H_3. For this case, Eq. (5.34) holds.

2. Only particle 4 has negative helicity. All others have positive helicity. Because the amplitudes under consideration have total helicity zero, we know that the sum of helicities of the particles with positive helicity must equal H_4^-. Hence there must be $2H_4^-\lambda_4$s and $2H_4^-$ physical $\tilde{\lambda}$s in the numerator. Further, as only particle 4 has negative helicity, it follows that $\mathcal{N}_b = 2\text{max}\{H_1^+, H_4^-\} = 2H_4^-$. And so there are $2H_4^-$ spinor-brackets of each kind. For this case Eq. (5.34) holds.

3. Only particle 1 has positive helicity. This case is logically equivalent to the above.

This proves Eq. (5.34), and therefore proves Eq. (5.17):

$$N \sim \langle\,\rangle_{(1)} ... \langle\,\rangle_{(2H)}\ \ [\,]_{(1)} ... [\,]_{(2H)} \Rightarrow [N] = (K^2)^{2H}\ , \tag{5.35}$$

where $H = \text{max}\{|h_1|, ..., |h_4|\}$. Establishing this result concludes the proof.

5.3.5 Ruling out theories with $N_p = 2$, for $A \geq 3$

In this section, we rule out self-interacting theories constructed from three-point amplitudes which necessitate two poles in the four-point amplitudes, for $A > 2$. This is simple pole-counting, augmented by constraint (5.13 and the results of subsection 5.3.3 for $N_p = 3$. Recall, for amplitudes within a self-interacting sector of a theory, we have $\sum_{i=1}^{4} h_i^{\text{ext}} = 0$, and $N_p = 2H + 1 - A$. For $N_p = 2$, we must have $H = \max\{|f|, |H|, |g|\} = (A+1)/2$.

Within this sector we may construct $A_4(1^{+H}, 2^{-H}, 3^{+f}, 4^{-f})$ from $A_3(H, g, f)$. By assumption, it has two factorization channels, specifically the t- and u-channels.[7] Without loss of generality, we take $f > 0$. The intermediary in the u-channel pole has spin $g = A - (H + f) = (A-1)/2 - f < H = (A+1)/2$, and poses no barrier to a unitary & local interpretation of the amplitude/theory.

However, for the t-channel's intermediary must have helicity $\tilde{H} = A + f - H = (A-1)/2 + f$. And so, for $f > 1$, we must include a new state with larger helicity $\tilde{H} = H + (f - 1) > H$. As discussed in subsection 5.3.3, this does not a priori spell doom for the theory. However, in this case it does: inclusion of this new, larger helicity, state within the theory forces inclusion of four-point amplitudes with these states on external lines. These *new* amplitudes have a larger number of poles: $2\tilde{h} + 1 - A = (2H + 1 - A) + 2f = 2 + 2f \geq 4 > 3$, for $f > 1$.

Theories with $H = (A+1)/2$, and $f > 1$ (or $g > 1$) cannot be consistent with unitarity and locality. Inspection reveals that all theories with $N_p = 2$ and $A > 2$ are of this type, save for three special examples. Explicitly, for $A = 3$, 4, 5, and $A = 6$, the $N_p = 2$ theories are defined by (H, g, f)s of the following types,

$$A = 3: \quad (2, 2, -1)\ ,\ \left(2, \frac{3}{2}, -\frac{1}{2}\right)\ ,\ (2, 1, 0)\ ,\ \left(2, \frac{1}{2}, \frac{1}{2}\right)\ , \tag{5.36}$$

$$A = 4: \quad \left(\frac{5}{2}, \frac{5}{2}, -1\right)\ ,\ \left(\frac{5}{2}, 2, -\frac{1}{2}\right)\ ,\ \left(\frac{5}{2}, \frac{3}{2}, 0\right)\ ,\ \left(\frac{5}{2}, 1, \frac{1}{2}\right)\ , \tag{5.37}$$

$$A = 5: \quad (3, 3, -1)\ ,\ \left(3, \frac{5}{2}, -\frac{1}{2}\right)\ ,\ (3, 2, 0)\ ,\ \left(3, \frac{3}{2}, \frac{1}{2}\right)\ ,\ (3, 1, 1)\ , \tag{5.38}$$

$$A = 6: \quad \left(\frac{7}{2}, \frac{7}{2}, -1\right)\ ,\ \left(\frac{7}{2}, 3, -\frac{1}{2}\right)\ ,\ \left(\frac{7}{2}, \frac{5}{2}, 0\right)\ ,\ \left(\frac{7}{2}, 2, \frac{1}{2}\right)\ ,\ \left(\frac{7}{2}, \frac{3}{2}, 1\right)\ , \tag{5.39}$$

and their conjugate amplitudes. Clearly, all but the last two entries on line (5.36) and the last entry on line (5.37) have $f > 1$ (thus $\tilde{H} > (A+1)/2$), and are inconsistent. Further, it is clear that all higher-A $N_p = 2$ theories may only have pathological three-point amplitudes, which indirectly lead to this same tension with locality and unitarity: except for those three special cases, all $N_p = 2$ theories must have fs that are larger than unity.

It is a simple exercise to show that these three pathological examples are inconsistent: playing around with the factorization channels of $A_4(H, -H, f, -f)$s reveals that again the t-channel is the problem. The t-channel requires the three-point amplitudes on lines (5.36) or (5.37) which are

[7] The s-channel in this amplitude is disallowed, as it would require a new particle with helicity $\tilde{H} = \pm A$, which would lead to amplitudes with $N_p = A + 1$.

directly ruled out, as they have $H = (A+1)/2$ and $f > 1$.

No three-point amplitude with $N_p = 2H + 1 - A = 2$ can lead to a constructible S-matrix consistent with locality and unitarity, for any A larger than two.

5.4 There is no GR (YM) but the true GR (YM)

In this section, we investigate further constraints imposed by coupling $A \geq 3$ theories to GR (YM) interactions. This is done by considering four-point amplitudes which factorize as $A_4 \to A_{GR} \times A_3$ and $A_4 \to A_{YM} \times A_3$, where A_3 is the vertex of some other theory. Note however that the arguments in this section apply only to three-point interactions which contain either a spin-2 or a spin-1 state. Other higher spin theories are not constrained in any way by this reasoning.

First, we find that all higher-spin theories are inconsistent if coupled to gravity. This is in addition to the previous section, where spin $s > 2$ theories with $A > 2$ were allowed if $N_p \leq 1$. Further, we show that massless spin-2 states participating in $A > 2$ three-point amplitudes must be identified with *the* graviton which appears in the usual $A = 2$ $A_3(+2, -2, \pm 2)$ three-point amplitudes defining the S-matrix of General Relativity. Pure pole-counting shows that *no* massless spin-2 state in *any* three-point amplitude with $A > 2$ can couple to GR, unless they are within the unique $(H, A) = (2, 6)$ three-point amplitudes, $A_3(2, 2, 2)$ and its complex conjugate. Similar results hold for gluons.[8]

To rule out higher-spin theories interacting with gravity, we show that amplitudes with factorization channels of the type,

$$A_4(1^{+2}, 2^{-2}, 3^{-H}, 4^{-h}) \to \frac{1}{K^2} A_3(2, -2, +2) \times A_3(-2, -H, -h) \tag{5.40}$$

cannot be consistent with unitarity and locality, unless $|H| \leq 2$ and $|h| \leq 2$.

It is relatively easy to see this, especially in light of the constraints from sections 5.3.2 and 5.3.3, which fix $H \leq A/2$ for $A \geq 3$. Note that, in order to even couple to GR's defining three-graviton amplitude, the three-point amplitude in question must have a spin-2 state. These two conditions

[8] We further show, in appendix 5.6.4 that theories with spin-3/2 states are also unique in a similar manner.

admit only three possible three-point amplitudes, for a given A:

$$\begin{aligned} A_3(A/2-1, A/2-1, 2) &\Rightarrow A_4(1^{+2}, 2^{-2}, 3^{-(A/2-1)}, 4^{-(A/2-1)}) \,, \\ A_3(A/2-1/2, A/2-3/2, 2) &\Rightarrow A_4(1^{+2}, 2^{-2}, 3^{-(A/2-1/2)}, 4^{-(A/2-3/2)}) \,, \text{ and} \\ A_3(A/2, A/2-2, 2) &\Rightarrow A_4(1^{+2}, 2^{-2}, 3^{-(A/2)}, 4^{-(A/2-2)}) \,. \end{aligned} \tag{5.41}$$

The minimal numerator which encodes the spins of the extenal states in, for instance, the first amplitude, must be,

$$N \sim [1|P|2\rangle^4 \left(\langle 34\rangle^2\right)^{(A/2-1)} \Rightarrow [N] = (K^2)^{3+A/2} \,. \tag{5.42}$$

However, by power-counting, the kinematic-dependent part of the amplitude must have mass-dimension,

$$\left[\frac{N}{f(s,t,u)}\right] = \left[\frac{1}{K^2} A^{(GR)}_{\text{Left}} A^{(A)}_{\text{Right}}\right] = \frac{(K^2)^{2/2}\,(K^2)^{A/2}}{(K^2)} = (K^2)^{A/2} \,, \tag{5.43}$$

and thus the denominator, $f(s,t,u)$, must have mass-dimension,

$$[f(s,t,u)] = (K^2)^3 \Rightarrow f(s,t,u) = s\,t\,u \,. \tag{5.44}$$

Casual inspection shows us that the "third" factorization channel, to be sensible, requires an intermediary with spin $A/2-1$ to couple directly via the leading $A=2$ gravitational interactions. This, and similar analysis for the other two classes of three-point amplitudes in Eq. (5.41), proves that the spin-2 particle associated with the graviton in the leading-order, $(H,A)=(2,2)$, gravitational interactions can only participate in three higher-derivative three-point amplitudes, namely,

$$A_3\,(+2,+1,+1) \ , \ A_3\left(+2,+\frac{3}{2},+\frac{3}{2}\right) \ , \ A_3\,(+2,+2,+2) \tag{5.45}$$

In the special case of the three-point amplitude $A_3(+2,+2,+2)$, the third channel simply necessitates an intermediate spin-2 state, the "graviton." Thus GR can couple to itself, or amplitudes derived from $R^a{}_b R^b{}_c R^c{}_a$, its closely related higher-derivative cousin [70][91].[9]

Second, we turn to gluons. Specifically, we show that gluons, i.e. the massless spin-1 particles

[9] The minimal numerators for the other candidate amplitudes in this theory, Eq. (5.41), have the same number of spinor-brackets in their numerator as that in Eq. (5.42); thus have the same mass-dimensions. Therefore all amplitudes must identical number of poles, and as in Eq. (5.44), they have $f(s,t,u) = stu$.

which couple to each-other at *leading* order via the $H = A = 1$ three-point amplitudes, can *not* consistently couple to any spin $s > 1$ within $A \geq 3$ amplitudes. This means that any constructible amplitude with factorization channels of the type,

$$A_4(1^{+1}, 2^{-1}, 3^{-H}, 4^{-h}) \to \frac{1}{K^2} A_3(1, -1, +1) \times A_3(-1, -H, -h) \tag{5.46}$$

cannot be consistent with unitarity and locality, unless $|H| \leq 1$ and $|h| \leq 1$.

Again, in light of the constraints from sections 5.3.2 and 5.3.3, which fix $H \leq A/2$ for $A \geq 3$, it is relatively easy to see this. To even possibly couple to this three-gluon amplitude, the three-point amplitude in question must have a spin-1 state. These two conditions allow only two possible three-point amplitudes, for a given A:

$$\begin{aligned} A_3(A/2 - 1/2, A/2 - 1/2, 1) &\Rightarrow A_4(1^{+1}, 2^{-1}, 3^{-(A/2-1/2)}, 4^{-(A/2-1/2)}), \text{ and} \\ A_3(A/2, A/2 - 1, 1) &\Rightarrow A_4(1^{+1}, 2^{-1}, 3^{-(A/2)}, 4^{-(A/2-1)}). \end{aligned} \tag{5.47}$$

The minimal numerator which encodes the spins of the extenal states in, for instance, the first amplitude, must be,

$$N \sim [1|P|2\rangle^2 (\langle 34\rangle^2)^{(A/2-1/2)} \Rightarrow [N] = (K^2)^{A/2+3/2}. \tag{5.48}$$

However, by power-counting, the kinematic-dependent part of the amplitude must have mass-dimension,

$$\left[\frac{N}{f(s,t,u)}\right] = \left[\frac{1}{K^2} A^{(YM)}_{\text{Left}} A^{(A)}_{\text{Right}}\right] = \frac{(K^2)^{1/2}\,(K^2)^{A/2}}{(K^2)} = (K^2)^{A/2-1/2}, \tag{5.49}$$

and thus the denominator, $f(s,t,u)$ must have mass-dimension two:

$$[f(s,t,u)] = (K^2)^2 \Rightarrow \frac{1}{f(s,t,u)} \text{ must have at least two poles.} \tag{5.50}$$

Again, casual inspection shows that, while the one pole—that in Eq. (5.46)—indeed has a legitimate interpretation as a factorization channel within this theory, the "second" channel generically does not: it requires the gluon to *marginally* couple to spin $A/2 - 1/2 \geq 1$ states. As seen in section 5.3.2, this cannot happen—unless $A/2 - 1/2 = 1 \Leftrightarrow A = 3$.

For the second amplitude in Eq. (5.47) the argument is a bit more subtle when $A = 3$. In this case, the u-channel is prohibited, but the t-channel is valid:

$$A_4(1^{+1}, 2^{-1}, 3^{-3/2}, 4^{-1/2}) \to \frac{1}{K^2} A_3(1, -1/2, 1/2) \times A_3(-1/2, -3/2, -1) \tag{5.51}$$

This interaction is ruled out through slightly more detailed arguments, involving the structure of the vector self-coupling constant in $A_3(1, -1, \pm 1)$—discussed in section 5.5. We pause to briefly describe how this is done, but will not revisit this particular, $A_3(1, -1/2, 1/2)$ interaction further (it is just a simple vector-fermion QED or QCD interaction). Simply, we note that $A_3(1, -1, \pm 1) \propto f_{abc}$, the structure-constant for a simple and compact Lie-Algebra; see Eq. (5.61). From here, it suffices to note that either by choosing the external vectors to be photons, or gluons of the same color, this amplitude vanishes, and then so does the original s-channel. Nothing is affected in Eq. (5.51) and so Eq. (5.50) cannot be fulfilled, implying that the coupling constant of $A_3(1, 1/2, 3/2)$ must vanish.

Thus at four-points YM can only couple to itself, gravity via the $A_3(\pm 2, 1, -1)$ three-point amplitude, or amplitudes derived from $F^a{}_b F^b{}_c F^c{}_a$, its closely related higher-derivative cousin.

5.5 Behavior near poles, and a possible shift

In this section we explain a new shift which is guaranteed to vanish at infinity. Using this shift, we re-derive classic results, such as (a) decoupling of multiple species of massless spin-2 particles [75], (b) spin-2 particles coupling to all particles (with $|H| \leq 2$, of course!) with identical strength, $\kappa = 1/M_{pl}$, (c) Lie Algebraic structure-constants for massless spin-1 self-interactions, and (d) arbitrary representations of Lie Algebra for interactions between massless vectors and massless particles of helicity $|H| \leq 1/2$.

Note that in section 5.3, we proved that a four point-amplitude, constructed from a given three-point amplitude and its parity conjugate, $A_3^{(H,A)}$ and $\bar{A}_3^{(H,A)}$, takes the generic form,

$$A_4 \sim \frac{\big(\langle\ \rangle[\ ]\big)^{2H}}{(K^2)^{2H-A+1}} \,. \tag{5.52}$$

Consequently in the vicinity of, say, the s-pole, the four-point amplitude behaves as,

$$\lim_{s\to 0} A_4 = \frac{1}{s}\frac{N}{t^{2H-A}} \,, \text{ where } N \sim \big(\langle\ \rangle[\ ]\big)^{2H} \,. \tag{5.53}$$

We exploit this scaling to identify a useful shift that allows us to analyze constraints on the coupling-constants, the "g_{abc}"-factor in three-point amplitudes [see Eq. (5.3)]. Complex deformation of the

Mandelstam invariants, which we justify in appendix 5.5.1, for arbitrary $\tilde{s}$ and $\tilde{t}$,

$$(s, t, u) \to (s + z\tilde{s}, t + z\tilde{t}, u + z\tilde{u}) \ , \tag{5.54}$$

grants access to the poles of $A_4(s, t, u)$ without deforming the numerator. Partitioning,

$$A_4(z = 0) \sim \kappa_A^2 \frac{N}{f(s, t, u)} \to A_4(z) \sim \kappa_A^2 \frac{N}{f(s(z), t(z), u(z))} \ , \tag{5.55}$$

accesses the poles in each term, while leaving the helicity-dependent numerator *un-shifted.* Basic power-counting implies that four-point amplitudes, constructed from three-point amplitudes of the type $A_3^{(H,A)} \times \bar{A}_3^{(H,A)}$, die off as $z \to \infty$ for $2H - A = 1, 2$ under this shift. Thus, four-point amplitudes are uniquely fixed by their finite-z residues under this deformation:

$$A_4(z = 0) = \sum_{z_P} \text{Res}\left(\frac{A_4(z)}{z}\right) \ . \tag{5.56}$$

Straightforward calculation of the residues on the s-, t-, and u-poles yields,

$$A(1_a, 2_b, 3_c, 4_d) = \left\{ \frac{(\tilde{s})^{2H-A}}{s} g^{abi} g^{icd} + \frac{(\tilde{t})^{2H-A}}{t} g^{adi} g^{ibc} + \frac{(\tilde{u})^{2H-A}}{u} g^{aci} g^{ibd} \right\} \frac{\text{Num}}{(\tilde{s}t - \tilde{t}s)^{2H-A}} \ . \tag{5.57}$$

Notably, this closed-form expression for the amplitude contains a non-local, spurious, pole which depends explicitly on the shift parameters, $\tilde{s}$ and $\tilde{t}$ (note: $\tilde{u} = -\tilde{s} - \tilde{t}$). Requiring these spurious parameters to cancel out of the final expression in theories, of self-interacting spin-1 particles, forces the Lie Algebraic structure of Yang-Mills [78][79]. Similarly, for theories of interacting spin-2 particles, we recover the decoupling of multiple species of massless spin-2 particles[75], and the equal coupling of all spin $|H| < 2$ particles to a spin-2 state [70][78][79].

5.5.1 Justifying the complex deformation.

One might worry about the validity of such a shift, and how it could be realized in practice. In other words, one could wonder whether-or-not shifting the Mandlsetam invariants, $(s, t, u) \to (s + z\tilde{s}, t + z\tilde{t}, u + z\tilde{u})$ would not also shift the numerator of the amplitude. Here, we prove that such a shift must always exist.

First, a concrete example. Suppose one desired to study the constraints on the f_{abc}s characterizing $A_3(1_a^{+1}, 2_b^{-1}, 3_c^{-1}) = f_{abc}\langle 23\rangle^3/\{\langle 31\rangle\langle 12\rangle\}$, through looking at the four-particle amplitude $A_4(1^{-1}, 2^{-1}, 3^{+1}, 4^{+1})$. The numerator must be $\langle 12\rangle^2[34]^2$. So, recognizing that $u = -s - t$ and

$\tilde{u} = -\tilde{s} - \tilde{t}$, we see if we shift

$$s = \langle 21\rangle[12] \to \langle 21\rangle([12] + z\,\tilde{s}/\langle 21\rangle) = s + z\tilde{s} \tag{5.58}$$

$$t = \langle 41\rangle[14] \to (\langle 41\rangle + z\,\tilde{t}/[14])[14] = t + z\tilde{t} \tag{5.59}$$

$$u = -s - t \to u + z\,\tilde{u} = -(s + t) - z(\tilde{s} + \tilde{t})\,. \tag{5.60}$$

Deforming the anti-holomorphic part of s and the holomorphic part of t allows the z-shift to probe the s-, t-, and u-poles of the amplitude while leaving the numerator $\langle 12\rangle^2[34]^2$ unshifted.

This is the general case for amplitudes with higher-spin poles, i.e. for amplitudes with $3 \geq 2H + 1 - A \geq 2$ (the only cases amenable to this general analysis); we prove this by contradiction. By virtue of having two or three poles in each term, we are guaranteed that the numerator does not have any complete factors of s, t, and/or u: if it did, then this would knock out one of the poles in a term, in violation of the assumption that $N_p = 2$ or 3.

5.5.2 Constraints on vector coupling $(A = 1)$

Here we derive consistency conditions on Eq. (5.57) for scattering amplitudes with external vectors, interacting with matter via leading-order, $A = 1$, couplings; $2H - A = 1$. Now, if the amplitude is invariant under changes of $\tilde{s} \to \tilde{S}$, then it necessarily follows that the same holds for re-definitions $\tilde{t} \to \tilde{T}$, and thus that unphysical pole cancels out of the amplitude.

Therefore, if $\frac{\partial A}{\partial \tilde{s}} = 0$, then it indeed follows that the amplitude is invariant under redefinitions of the shift parameter, $\tilde{s}$, and the unphysical pole has trivial residue. Beginning with the all-gluon amplitude, where the three-point amplitudes are $A_3(1_a^{+1}, 2_b^{-1}, 3_c^{\pm 1}) \propto f_{abc}$, we see that the derivative is,

$$\left.\frac{\partial A_4}{\partial \tilde{s}}\right|_{(H,A)=(1,1)} \propto f^{abi} f^{icd} + f^{aci} f^{ibd} + f^{adi} f^{ibc}\,. \tag{5.61}$$

Requiring this to vanish is equivalent to imposing the Jacobi identity on these f_{abc}s. Thus, requiring the amplitude to be physical forces the gluon self-interaction to be given by the adjoint representation of a Lie group [78][79][86].

Next, considering four-point amplitudes with two external gluons and two external fermions or scalars, we are forced to introduce a new type of coupling: $A_3(1_a^{\pm 1}, 2_b^{+h}, 3_c^{-h}) \propto (T_a)_{bc}$. Concretely, we wish to understand the invariance of $A_4(1_a^{+1}, 2_b^{-1}, 3_c^{-h}, 4_d^{+h})$, constructed from the shift (5.54), under redefinitions $\tilde{s} \to \tilde{S}$.

Factorization channels on the t- and u-poles are given by the products of two A_3s with one gluon and two spin-h particles, and thus are proportional to $(T_a)_{ci}(T_b)_{di}$ and $(T_a)_{di}(T_b)_{ci}$, respectively–while the s-channel is proportional to $f_{abi}(T_i)_{cd}$. So, $\frac{\partial A}{\partial \tilde{s}}$ is proportional to,

$$(T_a)_{ci}(T_b)_{id} - (T_a)_{di}(T_b)_{ic} + f_{abi}(T_i)_{cd}\,. \tag{5.62}$$

This is nothing other than the definition of the commutator of two matrices, T_a and T_b in an arbitrary representation of the Lie group "defined" by the gluons in Eq. (5.61) [78][79][86].

5.5.3 Graviton coupling

Four-point amplitudes with two external gravitons have $2H - A = 2$, and so Mandelstam deformation (5.54) yields,

$$A(1_a^{-2}, 2_b^{-h}, 3_c^{+2}, 4_d^{+h}) = \left\{\frac{\tilde{s}^2}{s}\kappa_h^{abi}\kappa_h^{icd} + \frac{\tilde{t}^2}{t}\kappa_h^{adi}\kappa_h^{ibc} + \frac{\tilde{u}^2}{u}\kappa_{h=2}^{aci}\kappa_h^{ibd}\right\}\frac{(\langle 12\rangle[34])^{4-2h}\langle 1|2-3|4]^{2h}}{(\tilde{s}t - \tilde{t}s)^2}\,, \tag{5.63}$$

where κ_h^{abc} is the coupling constant in $A_3(1_a^{\pm 2}, 2_b^{-h}, 3_c^{+h})$. Demanding this amplitude be independent of redefinitions of $\tilde{s} \to \tilde{S}$, again reduces to the constraint that the partial derivative of Eq. (5.63) must vanish. Evaluating the derivative, we see,

$$\left.\frac{\partial A_4}{\partial \tilde{s}}\right|_{(H,A)=(2,2)} \propto \tilde{s}\left(\kappa_h^{abi}\kappa_h^{icd} - \kappa_{h=2}^{adi}\kappa_h^{ibc}\right) + \tilde{t}\left(\kappa_h^{aci}\kappa_h^{ibd} - \kappa_{h=2}^{adi}\kappa_h^{ibc}\right)\,, \tag{5.64}$$

which vanishes only if,

$$\kappa_h^{abi}\kappa_h^{icd} = \kappa_{h=2}^{adi}\kappa_h^{ibc}\,,\ \text{and}\ \kappa_h^{aci}\kappa_h^{ibd} = \kappa_{h=2}^{adi}\kappa_h^{ibc}\,. \tag{5.65}$$

As noted in [78], for $h = 2$, this implies that the κ_h^{abc}s are a representation of a commutative, associative algebra. Such algebras can be reduced to self-interacting theories which decouple from each other. In other words, multiple gravitons, i.e. species of massless spin-2 particles interacting via the leading-order $(H, A) = (2, 2)$ three-point amplitudes, necessarily decouple from each-other. This is the perturbative casting of the Weinberg-Witten theorem [75]. As the multiple graviton species decouple, we refer to the *diagonal* graviton self-interaction coupling as, simply, κ.

Diagonal gravitational self-coupling powerfully restricts the class of solutions to Eq. (5.65) for $h < 2$. Directly, it implies that any individual graviton can only couple to a particle-antiparticle pair. In other words, $\kappa_h^{gab} = 0$, for different particle flavors a and b on the spin-$\pm h$ lines. Similar to the purely gravitational case, we write simply $\kappa_h^{gaa} = \kappa_h$. Furthermore, to solve Eq. (5.65) for $h \neq 2$ then it also must hold that $\kappa_h = \kappa_{h=2} = \kappa$. In other words, the graviton self-coupling constant κ is a simple constant; all particles which interact with a given unique graviton do so diagonally and with identical strengths. Thus multiple graviton species decouple into disparate sectors, and, within a given sector, gravitons couple to all massless states with identical strength, κ—the perturbative version of the equivalence principle [70][78][79].

5.5.4 Killing the relevant $A_3\big(0, \frac{1}{2}, -\frac{1}{2}\big)$-theory

This shift neatly kills the S-matrix constructed from the three-point amplitudes $A_3(\frac{1}{2}, -\frac{1}{2}, 0)$. Just as before, we will see that in order for $A_4(0, 0, \frac{1}{2}, -\frac{1}{2})$ to be constructible (via complex Mandelstam - deformations) and consistent, the coupling constant in the theory must vanish.

As in YM/QCD, in this theory $2H - A = 1$. Invariance of $A_4(0, 0, \frac{1}{2}, -\frac{1}{2})$ under deformation redefinitions $\tilde{s} \to \tilde{S}$ again boils down to a constraint akin to Eq. (5.61)—with one exception. Namely, there are only two possible factorization channels in this theory and not three: any putative s-channel pole would require a ϕ^3 interaction, not present in this minimal theory. And so invariance under redefinitions $\tilde{s} \to \tilde{S}$ reduces to,

$$\left.\frac{\partial A_4}{\partial \tilde{s}}\right|_{(H,A)=(\frac{1}{2},0)} = f^{acp} f^{bdp} + f^{adp} f^{bcp} \,. \tag{5.66}$$

The only solution to this constraint is for $f^{acp} f^{bdp} = 0 = f^{adp} f^{bcp}$, i.e. for the coupling constant to be trivially zero.[10]

5.6 Interacting spin-$\frac{3}{2}$ states, GR, and supersymmetry

Supersymmetry automatically arises as a consistency condition on four-point amplitudes built from leading-order three-point amplitudes involving spin-3/2 states. In a sense, this should be more-or-less obvious from inspection of the leading-order spin-3/2 amplitudes in Eqs. (5.23), and (5.24). For convenience, they are,

$$A_3\left(\frac{3}{2}, \frac{1}{2}, 0\right) \,,\; A_3\left(\frac{3}{2}, 1, -\frac{1}{2}\right) \,,\; A_3\left(\frac{3}{2}, \frac{3}{2}, -1\right) \,,\; \text{and } A_3\left(\frac{3}{2}, 2, -\frac{3}{2}\right) \,. \tag{5.67}$$

[10] One may wonder why such an argument does not also rule out conventional well-known theories, such as spinor-QED or GR coupled to spin-1/2 fermions, as inconsistent. The resolution to this question is subtle, but boils down to the fact that amplitudes involving fermions in these $A > 0$ theories have extra, *antisymmetric* spinor-brackets in their numerators. These extra spinor-brackets introduce a relative-sign between the two terms, and in effect modify the condition (5.66) from $\{ff + ff = 0 \Rightarrow f = 0\}$ to $ff - ff = 0$, which is trivially satisfied.

Clearly, every non-gravitational $A = 2$ amplitude with a spin-3/2 state involves one boson and one fermion, with helicity (magnitudes) that differ by exactly a half-unit. This should be unsurprising, as $A - 3/2 = 1/2$. Nonetheless, we should expect supersymmetry to be an emergent phenomena: throughout the previous examples, mandating a unitary interpretation of a factorization channel within novel four-point amplitudes in a theory forced introduction of new states with new helicities into the spectrum/theory. In a sense, the novelty of $A = 2$ amplitudes with external spin-3/2 states is that these new helicities do *not* lead to violations of locality and unitarity.

In amplitudes with external spin-3/2 states (and no external gravitons), each term in the amplitude must have $2H + 1 - A \to 3 + 1 - 2 = 2$ poles. Generically, one of these two poles will mandate inclusion of states with *new* helicities into the spectrum of the theory. Fundamentally, we see that the *minimal* $A = 2$ theory with a single species of spin-3/2 state is given by the two three-point amplitudes (and and their parity-conjugates):

$$A_3\left(\frac{3}{2}, 2, -\frac{3}{2}\right), \text{ and } A_3(2, 2, -2)\,. \tag{5.68}$$

These interactions *define* pure $\mathcal{N} = 1$ SUGRA, and are indicative of all other theories which contain massless spin-3/2 states (at leading order). All non-minimal extensions of any theory containing spin-3/2 states necessarily contain the graviton. As we will make precise below, supersymmetry necessitates gravitational interactions—supersymmetry requires the graviton.

Minimally, consider a four-particle amplitude which ties together four spin-3/2 states, two with helicity $h = +3/2$, and two with helicity $h = -3/2$, via leading-oder $A = 2$ interactions: $A_4^{(A=2)}(1^{+\frac{3}{2}}, 2^{+\frac{3}{2}}, 3^{-\frac{3}{2}}, 4^{-\frac{3}{2}})$. As this is a minimal amplitude, we consider the case where the like-helicity spin-3/2 states are identical: there is only one flavor/species of a spin-3/2 state. How many poles would such an amplitude have? By Eq. (5.13), there must be

$$2H + 1 - A = N_p \longrightarrow N_p = 2 \tag{5.69}$$

poles in any four-point amplitude constructed from $A = 2$ three-point amplitudes which has spin-3/2 states as its highest-spin external state. The key point here is really only that $N_p > 0$: the amplitude must have a factorization channel. Because it has two poles, at least one of them must be mediated by graviton exchange. In this minimal theory, as (a) gravitons can only be produced through particle-antiparticle annihilation channels and (b) the like-helicity spin-3/2 states are identical, both channels occur via graviton exchange. See Fig. 5.2(a) for specifics.

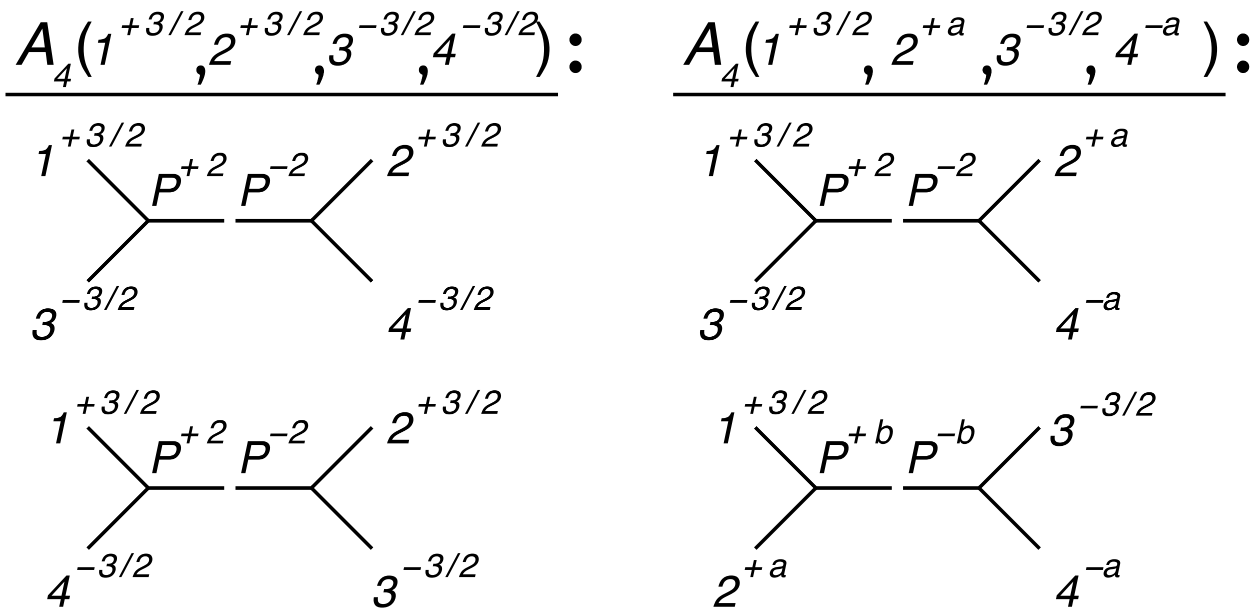


Figure 5.2: Factorization necessitates gravitation in theories with massless spin-3/2 states. Specifically, figure (a) represents the two factorization channels in the minimal four-point amplitude, $A_4(1^{+\frac{3}{2}}, 2^{+\frac{3}{2}}, 3^{-\frac{3}{2}}, 4^{-\frac{3}{2}})$ in an S-matrix involving massless spin-3/2 states. Further, figure (b) shows the two factorization channels present in the amplitude $A_4(3/2, -3/2, +a, -a)$.

Because this set of external states should always be present in *any* theory with leading-order interactions between any number of spin-3/2 states, S-matrices of these theories must *always* include the graviton.

This can be made even more explicit. Consider an S-matrix constructed, at least in part, from a three-point amplitude, $A_3(3/2, a, b)$, and its conjugate, $A_3(-3/2, -a, -b)$, where $H = 3/2$ and $A = 2$. These three-point amplitudes tie-together a spin-3/2 state with two other states which, collectively, have helicity-magnitudes $|H| \leq 3/2$. This theory necessarily contains the four-point amplitude,

$$A_4(1^{+\frac{3}{2}}, 2^{+a}, 3^{-\frac{3}{2}}, 4^{-a})\,. \tag{5.70}$$

As noted in Eq. (5.69), the denominator within this amplitude has two kinematic poles. Clearly the s-channel is has the spin-b state for an intermediary. However, as (a) the three-point amplitudes in the theory all have $A = 2$, and (b) the opposite-helicity spin-3/2 states (equivalently, the spin-a states) are antiparticles, the u-channel factorization channel must be mediated by a massless spin-2 state: the graviton. This is depicted in Fig. 5.2.

Note that the t-channel is also possible, mediated by a helicity $a + 1/2$ particle. However, repeating the above reasoning for the new $A_3(3/2, a+1/2, -a)$ amplitude will eventually lead to the necessity of introducing a graviton. This is because in each step the helicity of a is increased by $1/2$, and this process stops once a reaches $3/2$, when both the t and u-channels can only be mediated by a graviton. This pattern of adding particles with incrementally different spin will be investigated further in the following sections.

Before delving into details of the spectra in theories with multiple species of spin-3/2 states, we note one final feature of these theories. Analysis of their four-particle amplitudes, e.g. the amplitude in Eq. (5.70), via on-shell methods such as the Mandelstam deformation introduced in the previous section, straightforwardly shows that the coupling constants in this theory [the $g^{\pm}_{abc}s$ in the language of Eq. (5.3)] are *equal* to $\kappa = 1/M_{pl}$, the graviton self-coupling constant. More generally, in any $A = 2$ theory with spin-3/2 states, *each and every* defining three-point amplitude, $A_3(1_a^{h_a}, 2_b^{h_b}, 3_c^{h_c}) = \kappa_{abc} M_{abc}(\langle\, , \rangle)$ has an identical coupling constant, $\kappa_{abc} = \kappa = 1/M_{pl}$, up to (supersymmetry (SUSY) preserving Kronecker) delta-functions in flavor-space.

It is important to emphasize here that, as the spin-3/2 gravitinos only interact via $A = 2$ three-point amplitudes, they cannot change the $A < 2$ properties of any state within the same amplitude. Concretely, a bosonic (fermonic) state which transforms under a given specific representation of a compact Lie Algebra, i.e. a particle which interacts with massless vectors (gluons) via leading order ($A = 1$) interactions, can only interact with a fermonic (bosonic) state which transforms under the *same* representation of the Lie Algebra when coupled to spin-3/2 states within $A = 2$ three-point amplitudes. From the point of view of the marginal interactions, only the spin of the states which interact with massless spin-3/2 "gravitino(s)" may change. This is the on-shell version of the statement that all states within a given supermultiplet have the same quantum-numbers, but different spins.

We now consider the detailed structure of interactions between states of various different helicities which participate in S-matrices that couple to massless spin-3/2 states. Minimally, such theories include a single graviton and a single spin-3/2 state (and its antiparticle). Equipped with this, we can ask what the next-to-minimal theory might be. There are two ways one may enlarge the theory: (1) introducing a state with a *new* spin into the spectrum of the theory, or (2) introducing another species of massless spin-3/2 state. We pursue each in turn.

5.6.1 Minimal extensions of the $\mathcal{N} = 1$ supergravity theory

First, we ask what the minimal enlargement of the $\mathcal{N} = 1$ SUGRA theory is, if we require inclusion of a single spin-1 vector into the spectrum. In other words, what three-particle amplitudes must be added to,

$$\mathcal{N} = 1\,\mathrm{SUGRA} \Longleftrightarrow \left\{A_3(+2,\pm 2,-2)\,,\; A_3(+3/2,\pm 2,-3/2)\right\}, \tag{5.71}$$

in order for all four-particle amplitudes to factorize properly on all possible poles, once vectors are introduced into the spectrum. Clearly, inclusion of a vector requires inclusion of,

$$A_3(+1,\pm 2,-1)\,, \tag{5.72}$$

into the theory. It is useful to consider the four-particle amplitude, $A_4(-3/2,+3/2,+1,-1)$; its external states are only those known from the minimal theory and this extension, i.e. a particle-antiparticle pair of the original spin-3/2 "gravitino" and a particle-antiparticle pair of the new massless spin-1 vector.

By Eq. (5.69), this amplitude must have two poles. The s-channel pole is clearly mediated by graviton-exchange, as the amplitude's external states are composed of two distinct pairs of antiparticles. The *second* pole brings about new states. There are two options for which channel the second pole is associated with: the u-channel pole, or the t-channel pole. On the t-channel, the amplitude must factorize as,

$$A_4(-3/2,+3/2,+1,-1)\Big|_{t\to 0} \to A_3(1^{-\frac{3}{2}},4^{-1},P_{14}^{+\frac{1}{2}})\frac{1}{(p_1+p_4)^2}A_3(2^{+\frac{3}{2}},3^{+1},-P_{14}^{-\frac{1}{2}})\,, \tag{5.73}$$

and we see that, by virtue of the fact that the three-point amplitudes must have $A=2$, the new particle introduced into the spectrum is a spin-1/2 fermion. This option corresponds to the spectrum for $\mathcal{N}=1$ SUGRA that is invariant under charge conjugations, parity flips, and time-reversals (CPT complete), combined with the CPT complete spectrum for $\mathcal{N}=1$ SYM. In this case, the full $A=2$ sector of the theory would be,

$$\left\{A_3(2,2,-2)\,,\; A_3\left(2,\frac{3}{2},-\frac{3}{2}\right)\,,\; A_3\left(2,1,-1\right)\,,\; A_3\left(2,\frac{1}{2},-\frac{1}{2}\right)\,,\; A_3\left(\frac{3}{2},1,-\frac{1}{2}\right)\right\}\,, \tag{5.74}$$

Note that, as the spin-2 and spin-3/2 states interact gravitationally, one can add extra, leading-order $A=1$ ("gauge") interactions between the spin-$H\leq 1$—but it is not necessary.

If the second pole is in the u-channel, then the amplitude must factorize as,

$$A_4(-3/2,+3/2,+1,-1)\Big|_{u\to 0} \to A_3(1^{-\frac{3}{2}},3^{+1},P_{13}^{-\frac{3}{2}})\frac{1}{(p_1+p_3)^2}A_3(2^{+\frac{3}{2}},4^{-1},-P_{13}^{+\frac{3}{2}})\,, \tag{5.75}$$

and we see that, by virtue of the fact that the three-point amplitudes must have $A = 2$, the new particle introduced into the spectrum must be *another* spin-3/2 gravitino. This option corresponds to the CPT-complete spectrum for $\mathcal{N} = 2$ SUGRA.

It is not immediately obvious that this internal spin-3/2 state must be distinguishable from the original spin-3/2 state. Distinguishability comes from the fact that the factorization channel $A_3(1_a^{3/2}, 2_b^{3/2}, P^{-1})A_3(3_{\bar{a}}^{-3/2}, 4_{\bar{b}}^{-3/2}, -P^{+1})/K_{12}^2$ must be part of a four-particle amplitude with both the "new" and the "old" spin-3/2 species as external states. This amplitude also has only two poles. As one of them is mediated by vector exchange, we see that there is only *one* graviton-exchange channel. Therefore the new and old spin-3/2 states cannot be identical.

Constructing theories in this way is instructive. As a consequence of requiring a unitary interpretation of all factorization channels in non-minimal S-matrices involving massless spin-3/2 states, we are forced to introduce a new fermion for every new boson and vice-versa. Further, we see that through allowing minimal extensions to this theory, we can either have extended supergravity theories, i.e. $\mathcal{N} = 2$ SUGRA theories, truncations of the full $\mathcal{N} = 8$ SUGRA multiplet, or supergravity theories and supersymmetric Yang-Mills theories in conjunction, i.e. $\mathcal{N} = 1$ SUGRA $\times\, \mathcal{N} = 1$ SYM theories. The same lessons apply for more extended particle content. However, it is difficult to make such constructions systematic. Below, we discuss the second, more systematic, procedure which hinges upon the existence of $\mathcal{N}$ distinguishable species of spin-3/2 fermions.

5.6.2 Multiple spin-$\frac{3}{2}$ states and (super)multiplets

Another way to understand these constructions is as follows: specify the number $\mathcal{N}$ of distinguishable species of spin-3/2 states, and then specify what else (besides the graviton) must be included into the theory. This amounts to specifying the number of supersymmetries and the number and type of representations of the supersymmetry algebra. In the above discussion, the two minimal extensions to the $\mathcal{N} = 1$ SUGRA theory were: (a) $\mathcal{N} = 1$ SUGRA $\times\, \mathcal{N} = 1$ SYM, and (b) $\mathcal{N} = 2$ SUGRA, with *two* gravitinos and one vector.

To render this construction plan unique, we require that all spins added to the theory besides the graviton and the $\mathcal{N}$ gravitinos, i.e. all *extra* supermultiplets included in the theory, be the "top" helicity component of whatever comes later. So, again, the discussion in subsection 5.6.1 would cleanly fall into two pieces: (A) a *single* gravitino (in the graviton supermultiplet) together with a gluon and its descendants, and (B) *two* distinct gravitinos (in the graviton supermultiplet) and their descendants. Clearly this procedure can be easily extended (see subsection 5.6.3).

The general strategy is to look at amplitudes which tie together gravitinos and lower-spin descendants (ascendants) of the "top" ("bottom") helicities in the theory, of the type

$$A(1_x^{+3/2}, 2_y^{-3/2}, 3_u^{-s}, 4_v^{+s})\,. \tag{5.76}$$

Here, the x and y labels describe the species information of the two gravitinos, and u and v describe the species information of the lower-spin particles in the amplitude. Note graviton-exchange can only happen in the s-channel. Unitary interpretation of one of the other channels generically forces the existence of a new spin-$s-\frac{1}{2}$ state into the spectrum. For pure SUGRA (i.e. no spin-1, 1/2, or 0 "matter" supermultipets), this works as follows,

1. One gravitino ($\{a\}$). The unique amplitude to consider, after the archetype in Eq. (5.76), is $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_a^{3/2}, 4_a^{-3/2}\big)$. Both s- and t-channels occur via graviton-exchange. The theory is self-complete: the other amplitude does not require any new state.

2. Two gravitinos ($\{a, b\}$). The unique amplitude to consider is $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_b^{3/2}, 4_b^{-3/2}\big)$. Here, the t-channel is disallowed; the u-channel needs a vector with gravitino-label $\{ab\}$. Inclusion of this state completes the theory.

3. Three gravitinos ($\{a, b, c\}$). Two classes amplitudes of the type in Eq. (5.76) to consider. First, $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_c^{3/2}, 4_c^{-3/2}\big)$ requires a vector with gravitino-label ac in its u-channel; as there are three amplitudes of this type, there are three distinguishable vectors: $\{ab, ac, bc\}$. Second, $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_{bc}^{+1}, 4_{bc}^{-1}\big)$ needs a fermion with gravitino-label $\{abc\}$ in the u-channel. No other amplitudes require any new states.

4. Four gravitinos ($\{a, b, c, d\}$). Here, the structure is slightly more complicated, but similarly hierarchical. Three classes of amplitudes, each following from its predecessor. First, there are $\binom{4}{2}$ distinct $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_c^{3/2}, 4_c^{-3/2}\big)$s. They require vectors with gravitino labels $\{ab, ac, ad, bc, bd, cd\}$. Second, there are $\binom{4}{3}$ distinct $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_{bc}^{+1}, 4_{bc}^{-1}\big)$s, which require spin-1/2 fermions with labels $\{abc, abd, acd, bcd\}$. Third and finally, we consider $A_4\big(1_a^{3/2}, 2_a^{-3/2}, 3_{bcd}^{1/2}, 4_{bcd}^{-1/2}\big)$. On its u-channel, it requires a spin-0 state with gravitino-label $\{abcd\}$.

Crucially, we observe that all spins present in the graviton supermultiplet (the graviton and all of its descendants) with $\mathcal{N}$ gravitinos are still present in the graviton supermultiplet with $\mathcal{N}+1$ gravitinos—but with higher multiplicities. These descendant states are explicitly labeled by the gravitino species from whence they came. Spin-s states in the graviton multiplet have $\binom{\mathcal{N}}{s}$ distinct gravitino/SUSY labels.

Importantly, if h is the unique lowest helicity descendant of the graviton with $\mathcal{N}$ gravitinos, then inclusion of an extra gravitino allows for $\mathcal{N}$ *new* helicity-h descendants of the graviton. Now, studying $A_4\big(1^{3/2}_{\mathcal{N}+1}, 2^{-3/2}_{\mathcal{N}+1}, 3^{+h}_{ab...\mathcal{N}}, 4^{-h}_{ab...\mathcal{N}}\big)$, we see that again the u-channel requires a single new descendant with helicity $h-1/2$ and SUSY-label $\{ab...\mathcal{N}, \mathcal{N}+1\}$.

This logic holds for the descendants of all "top" helicity states: isomorphic tests and constructions, for example, allow one to construct and count the descendants from the gluons of SYM theories. We see below that, by obeying the consistency conditions derived from pole-counting and summarized in Fig. 5.1, this places strong constraints on the number of distinct gravitinos in gravitational and mixed gravitational and $A<2$-theories.

5.6.3 Supersymmetry, locality, and unitarity: tension and constraints

As we have seen, inclusion of $\mathcal{N}$ distinguishable species of massless spin-3/2 states into the spectrum of constructible theories forces particle helicities $\{H, H-1/2, ..., H-\mathcal{N}/2\}$ into the spectrum. But, as we have seen in sections 5.3 and 5.4, the $A=2$ gravitational interactions cannot consistently couple to helicities $|h|>2$. And so, within the supersymmetric gravitational sector, we must have (a) $H=2$, and (b) $H-\mathcal{N}/2 \geq -2$. Otherwise, we must couple a spin-$5/2>2$ to gravity–which is impossible. Locality and unitarity constrains $\mathcal{N} \leq 8$.

So there is tension between locality, unitarity, and supersymmetry. We now ask about the spectrum of next-to-minimal theories coupled to spin-3/2 states. There are two options for such next-to-minimal theories: either (relevant) self-interacting scalars or (marginal) self-interacting vectors coupled to $\mathcal{N}$ flavors of spin-3/2 particles. Immediately, we see that ϕ^3 cannot be consistently coupled to spin-3/2 states. Coupling the spin-0 lines in ϕ^3 to even one spin-3/2 state would force the existence of non-zero $A_3(1/2, -1/2, 0)$-type interactions. But these interactions, as discussed in section 5.5.4, are not consistent with unitarity and locality. So *relevant* interactions cannot be supersymmetrized in flat, four-dimensional, Minkowski space.

However, for ($A=1$) self-interacting gluons, the story is different. By the arguments above, unitarity and locality dictate that if $\mathcal{N}$ spin-3/2 particles are coupled to gluons, then gluons must couple via marginal interactions, to spin-$\pm|1-\mathcal{N}/2|$ states. Again basic pole-counting in section 5.3.2, $A=1$ interactions are only valid for $|h| \leq 1$. And so, we must have $\mathcal{N} \leq 4$, if we would like to couple interacting vectors to multiple distinct spin-3/2 particles while also respecting locality and unitarity of the S-matrix.

5.6.4 Uniqueness of spin-3/2 states

In this appendix, we will use similar arguments to those in section 5.4 to show that massless spin-3/2 states can only couple consistently to massless particles with helicities $|H| \leq 2$. Recall that for $A > 2$, no constructible theory can be consistent with unitarity and locality unless $A/3 \leq H \leq A/2$. To see whether-or-not the gravitino discussed in section 5.6 can couple to any higher-A amplitude, we simply study four-point amplitudes with factorization channels of the type,

$$A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-c}, 4^{-d}) \to A_3^{(GR)}(1^{+2}, 2^{-\frac{3}{2}}, P^{+\frac{3}{2}}) \frac{1}{s} A_3^{(A)}(P^{-\frac{3}{2}}, 3^{-c}, 4^{-d})\,. \tag{5.77}$$

Note that only two three-point amplitudes are consistent with the dual requirements (a) $A/3 \leq H \leq A/2$ and (b) $3/2 \in \{h_1, h_2, h_3\}$ for $A > 2$. These theories, and their corresponding four-particle amplitudes, are:

$$\begin{aligned} A_3(A/2-1/2, 3/2, A/2-1) &\Rightarrow A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-(A/2-1/2)}, 4^{-(A/2-1)})\,, \text{ and} \\ A_3(A/2, 3/2, A/2-3/2) &\Rightarrow A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-A/2}, 4^{-(A/2-3/2)})\,. \end{aligned} \tag{5.78}$$

Now, the minimal numerator which encodes the helicity information of, for instance, the first amplitude is,

$$N \sim [1|P|2\rangle^3 [1|Q|3\rangle \left(\langle 3,4\rangle^2\right)^{A/2-1} \Rightarrow [N] = \left(K^2\right)^{(A/2)+3}\,. \tag{5.79}$$

However, by power-counting, the kinematic part of the amplitude must have mass-dimension,

$$\left[\frac{N}{f(s,t,u)}\right] = \left[\frac{1}{K^2} A_{\text{Left}}^{(GR)} A_{\text{Right}}^{(A)}\right] = \frac{(K^2)^{1/2}(K^2)^{A/2}}{(K^2)} = (K^2)^{A/2}\,, \tag{5.80}$$

and thus the denominator, $f(s,t,u)$ must have mass-dimension three:

$$[f(s,t,u)] = (K^2)^3 \Rightarrow f(s,t,u) = s\,t\,u\,! \tag{5.81}$$

However, as is obvious from inspection of any amplitude for $A \geq 4$, two of these factorization channels require inclusion of states with helicities which violate the most basic constraint (5.13). Thus, no spin-3/2 state in any three-point amplitude with $A \geq 4$ can be identified with the gravitino of section 5.6. The sole exception to this is the three-point amplitude $(H, A) = (3/2, 3)$:

$$A_3(1^0, 2^{+\frac{3}{2}}, 3^{+\frac{3}{2}}) \Rightarrow A_4(1^{+2}, 2^{-\frac{3}{2}}, 3^{-\frac{3}{2}}, 4^0)\,. \tag{5.82}$$

Factorization channels in this putative amplitude necessitate only either scalar or gravitino exchange, and are thus not in obvious violation of the consistency condition (5.13).

5.6.5 F^3- and R^3-theories and SUSY

Basic counting arguments show us that the F^3- and R^3- theories, i.e. the S-matrices constructed from $A_3(1,1,1)$ and $A_3(2,2,2)$ and their conjugates, are not compatible with leading-order (SUGRA) interactions with spin-3/2 states. The argument is simple.

First, we show that F^3-theories are not supersymmetrizable. Begin by including the minimal $\mathcal{N}=1$ SUGRA states, together with the three-particle amplitudes which couple gluons to the single species of spin-3/2 (gravitino) state that construct the $\mathcal{N}=1$ SYM multiplet. Additionally, allow the F^3-three-point amplitude as a building-block of the S-matrix. In other words, begin consider the four-particle S-matrix constructed from,

$$A_3\left(2,\frac{3}{2},-\frac{3}{2}\right)\ ,\ A_3\left(\frac{3}{2},1,-\frac{1}{2}\right)\ ,\ A_3\left(1,\frac{1}{2},-\frac{3}{2}\right)\ ,\ \text{and}\ A_3\left(1,1,1\right)\ , \tag{5.83}$$

where all spin-1 states are gluons, and all spin-1/2 states are gluinos. Now, consider the four-particle amplitude, $A_4(+\frac{3}{2},-\frac{1}{2},-1,-1)$. On the s-channel, it factorizes nicely:

$$A_4\left(1^{+\frac{3}{2}},2^{-\frac{1}{2}},3^{-1},4^{-1}\right)\Big|_{s\to 0}\to\frac{1}{s}A_3\left(1^{\frac{3}{2}},2^{-\frac{1}{2}},P^{+1}\right)A_3\left(P^{-1},3^{-1},4^{-1}\right)\ . \tag{5.84}$$

Clearly, it fits into the theory defined in Eq. (5.83). To proceed further, we note that its minimal numerator must have the form,

$$N\sim[1|P|2\rangle[1|Q|3\rangle[1|K|4\rangle\langle 34\rangle,\ . \tag{5.85}$$

Now, this amplitude must have kinematic mass-dimension,

$$\left[\frac{A_3^{\text{SUGRA}}A_3^{F^3}}{K^2}\right]=\left[\frac{[\ ]^2\langle\ \rangle^3}{\langle\rangle[\ ]}\right]=\left(K^2\right)^{1+\frac{1}{2}}\ . \tag{5.86}$$

Combining Eq. (5.85) and Eq. (5.86), we see that $1/f(s,t,u)$ must have two poles. On the *other* pole, say on the $t\to 0$ pole, it takes the form

$$A_4\left(1^{+\frac{3}{2}},2^{-\frac{1}{2}},3^{-1},4^{-1}\right)\Big|_{t\to 0}\to\frac{1}{t}A_3\left(1^{\frac{3}{2}},4^{-1},P^{+\Delta}\right)A_3\left(P^{-\Delta},2^{-\frac{1}{2}},3^{-1}\right)\ . \tag{5.87}$$

Now, one of these two sub-amplitudes must have $A = 3$. However, recall that in section 5.4 we showed that the only three-point amplitude which may have spin-1 states identified with the gluons is the $A_3(1,1,1)$ amplitude. Observe that, regardless of which amplitude has $A = 3$, *both* amplitudes contain one spin-1 state and another state with spin-$s \neq 1$. Therefore *neither* amplitude can consistently couple to the $A = 1$ gluons. Therefore we conclude that $\mathcal{N} = 1$ supersymmetry is incompatible with F^3-type interactions amongst gluons.

Similar arguments show that the three-point amplitudes arising from R^3-type interactions cannot lead to consistent S-matrices, once spin-3/2 gravitinos are included in the spectrum. Again, we first specify the four-particle S-matrix as constructed from the following primitive three-particle amplitudes:

$$A_3\left(2, \frac{3}{2}, -\frac{3}{2}\right)\ , \text{ and } A_3\left(2,2,2\right)\ , \tag{5.88}$$

Now, consider the four-particle amplitude, $A_4(+\frac{3}{2}, -\frac{3}{2}, -2, -2)$, an analog to that considered in the F^3-discussion. On the s-channel, it factorizes nicely:

$$A_4\left(1^{+\frac{3}{2}}, 2^{-\frac{3}{2}}, 3^{-2}, 4^{-2}\right)\Big|_{s\to 0} \to \frac{1}{s} A_3\left(1^{\frac{3}{2}}, 2^{-\frac{3}{2}}, P^{+2}\right) A_3\left(P^{-2}, 3^{-2}, 4^{-2}\right)\ . \tag{5.89}$$

Clearly, it fits into the theory defined in Eq. (5.88). To proceed further, we note that its minimal numerator must have the form,

$$N \sim [1|P|2\rangle^3 \left(\langle 34\rangle^2\right)^2\ . \tag{5.90}$$

Now, this new amplitude must have kinematic mass-dimension,

$$\left[\frac{A_3^{\text{SUGRA}} A_3^{R^3}}{K^2}\right] = \left[\frac{[\ ]^2\langle\ \rangle^6}{\langle\rangle[\ ]}\right] = \left(K^2\right)^3\ . \tag{5.91}$$

Combining Eq. (5.90) and Eq. (5.91), we see that $1/f(s,t,u)$ must have two poles, again, as in the F^3-discussion above. On the *other* pole, say on the $t \to 0$ pole, it takes the form

$$A_4\left(1^{+\frac{3}{2}}, 2^{-\frac{3}{2}}, 3^{-2}, 4^{-2}\right)\Big|_{t\to 0} \to \frac{1}{t} A_3\left(1^{\frac{3}{2}}, 4^{-2}, P^{+\Delta}\right) A_3\left(P^{-\Delta}, 2^{-\frac{3}{2}}, 3^{-2}\right)\ . \tag{5.92}$$

Reasoning isomorphic to that which disallowed $\mathcal{N} = 1$ SUSY and F^3-gluonic interactions rules out the compatibility of this given factorization channel with $\mathcal{N} = 1$ SUSY and R^3-effective gravitational interactions. Namely, it must be that one of these two sub-amplitudes must have $A = 6$. However, recall that in section 5.4 we showed that the only three-point amplitude which may have spin-1 states identified with the gluons is the $A_3(1,1,1)$ amplitude. Observe that, regardless of which amplitude has $A = 6$, *both* amplitudes contain one spin-1 state and another state with spin-$s \neq 2$. Therefore *neither* amplitude can consistently couple to the $A = 2$ gravitons. Therefore we conclude that $\mathcal{N} = 1$ supersymmetry is *also* incompatible with R^3-type interactions amongst gravitons.

5.7 Future directions and concluding remarks

Our results can be roughly separated into two categories. First, we classify and systematically analyze all possible three-point massless S-matrix elements in four-dimensions, via basic pole-counting. The results of this analysis are succinctly presented in Fig. 5.1. Second, we study the couplings and spectra of the few, special, self-interactions allowed by this first, broader-brush, analysis. In this portion of the chapter, we reproduce standard results on the structures of massless S-matrices involving higher-spin particles, ranging from the classic Weinberg-Witten theorem and the Equivalence Principle to the existence of supersymmetry, as consequences of consistency conditions on various S-matrix elements. We recap the main results below.

Locality and constructibility fix the generic pole-structure of four-point tree-amplitudes constructed from fundamental higher-spin three-point massless amplitudes. Tension between the number of poles mandated by these two principles, and unitarity, which bounds the number of poles in an amplitude from above ($N_p \leq 3$), eliminates all but a small (yet infinite) sub-class of three-point amplitudes as leading to four-particle tree-level S-matrices that are inconsistent with locality and unitarity.

Already from this point of view we see that, for low $A = |\sum_{i=1}^{3} h_i|$, (Super-)Gravity, (Super-)Yang-Mills, and ϕ^3-theory are the unique, leading, interactions between particles of spin-$|h| \leq 2$. Further, we see that gravitational interactions cannot *directly* couple to particles of spin-$|H| > 2$. Similarly, massless vectors interacting at leading-order ($A = 1$) cannot consistently couple to massless states with helicity-$|H| > 1$.

In light of these constraints, we study higher-A theories. The upper-bound on the number of poles in four-particle amplitudes, imposed by unitarity and locality, is even stronger for higher-spin *self-interactions* ($N_p \leq 1$ for $A > 2$). The set of consistent three-point amplitudes with $A > 2$ is further reduced to lie between the lines $H = A/2$ and $H = A/3$.

Exploiting this, we re-examine whether-or-not the primitive amplitudes which define the S-matrices of General Relativity and Yang-Mills can *indirectly* couple to higher-spin states in a consistent manner. As they cannot directly couple to higher-spin states, this coupling can only happen within non-primitive four-point (and higher) amplitudes, which factorize into GR/YM self-interaction amplitudes (with $H = A$), multiplied by an $A > 2$ three-point amplitude with an external tensor or vector. Again, simple pole-counting shows that amplitudes which couple the $A \leq 2$ to the $A > 2$ theories generically have poles whose unitary interpretation mandates existence of a particle with spin $\tilde{H} > A/2$.

Having such a high-spin particle contradicts the most basic constraint (5.13), and thus invalidates the interactions—save for two special examples. These examples are simply the higher-derivative amplitudes which also couple three like-helicity gravitons, $A_3(2,2,2)$, and/or like-helicity gluons, $A_3(1,1,1)$. There is a qualitative difference between massless spin-2(1) particles participating in lower-spin ($A \leq 2$) amplitudes, and massless spin-2(1) particles participating in higher-spin ($A > 3$) amplitudes. The graviton is unique. Gluons are also unique. They cannot be coupled to particles of spin-$|h| > 2$!

Equipped with the (now) *finite* list of leading interactions between spin-1, spin-2, and lower-spin states, we then analyze the structure of their interactions—i.e. their coupling constants. To do this, we set up, and show the validity of, the Mandelstam shift (5.54). Assuming parity-invariance, and thus $g^{+}_{ijk} = +g^{-}_{ijk}$ [in the notation of Eq. (5.3)], we perform the Mandelstam shift on four-point amplitudes in these theories. Invariance with respect to redefinitions of the unphysical shift-parameter directly implies the Lie Algebraic structure of the marginal ($A = 1$) coupling to massless vectors; similarly massless tensors must couple (a) diagonally (in flavor space), and (b) with equal strength to all states.

Finally, we analyze consistency conditions on four-point amplitudes which couple to massless spin-3/2 states. The minimal theory/set of interacting states, at leading order, which include a single spin-3/2 state is the theory with a single graviton and a single spin-3/2 state, at $A = 2$. From this observation, we identify the spin-3/2 state with the gravitino. The gravitino also couples to matter with strength $\kappa = 1/M_{pl}$, but as it is not a boson, it does not couple "diagonally": coupling to non-graviton states within the leading-order $A = 2$ interactions automatically necessitates introduction of a fermion for every boson already present in the theory, and vice-versa. We recover the usual supersymmetry constraints, such as fermion-boson level matching, and the maximal amount of distinguishable gravitinos which may couple to gluons and/or gravitons; above these bounds, the theories becomes inconsistent with locality and unitarity.

We close with future directions. Clearly, it would be interesting to discuss on-shell consistency conditions for theories which have primitive amplitudes which begin at *four* points, rather than at three-points. Certain higher-derivative theories, such as the nonlinear sigma model[92], are examples of this type of theory: in the on-shell language, derivative interactions between scalars can only act to give factors of non-trivial kinematical invariants within the numerator of a given amplitude. All kinematical invariants are identically zero at three points. So the first non-zero S-matrix elements in derivatively-coupled scalar theories must be at four-points. Supported by the existence of semi- on-shell recursions in these theories[93], it is conceivable that these theories are themselves constructible. Straightforwardly, this leads to the on-shell conclusion that all S-matrix elements in these theories have an even number of external legs. Much more could be said, and is left to future work.

Besides theories with derivative interactions, there is also a large of class of higher-spin theories *not* constrained by any of the arguments presented in this chapter. These are $A \geq 3$, $N_p \leq 1$ theories which do not contain any spin-1 or spin-2 states, for example $A_3(3/2, 3/2, 0)$. It is not clear from this on-shell perspective whether such theories are completely compatible with locality and unitarity, or more sophisticated tests can still rule them out.

Indeed, an exhaustive proof of the spin-statistics theorem has yet to be produced through exclusively on-shell methods. Proof of this theorem usually occurs, within local formulations of field theory, through requiring no information propagation outside of the light-cone. In the manifestly on-shell formalism, all lines are on their respective light-cones; superluminal propagation, and (micro-)causality violations are naively inaccessible. Ideally some clever residue theorem, such as that in [52], should prove the spin-statistics theorem in one fell swoop. Further, it would be interesting to prove that parity-violation, with $g^{+}_{abc} = -g^{-}_{abc}$, within three-point amplitudes only leads to consistent four-particle amplitudes for parity-violating gluon/photon-fermion amplitudes ($A = 1$).

Further, one may reasonably ask what the corresponding analysis would yield for *massive* states in constructible theories. As is well known, massive vectors must be coupled to spinless bosons (such as the Higgs), to retain unitarity at $E_{\mathrm{CM}} \sim s \geq m_V$ [86]. It would be extremely interesting to see this consistency condition, and analogous consistency conditions for higher-spin massive particles, naturally fall out from manifestly on-shell analyses.

Finally, one may wonder whether-or-not similar analysis to that presented in this chapter could apply to loop-level amplitudes in massless theories. In the on-shell language, loops and trees have varying degrees of transcendental dependence on kinematical invariants. Concretely, tree amplitudes have at most simple poles. However, loop amplitudes have both polylogarithms which have branch-cuts and are functions of ratios of kinematic invariants, and rational terms with higher-order poles in the kinematical invariants, the existence of which is (almost) solely to cancel the higher-order poles in these same kinematic invariants arising from these polylogarithms. See, for example, the discussion in chapters 3 and 4. These branch-cuts and higher-order poles would dramatically complicate any attempt to use the reasoning championed in this chapter, at loop-level. Nonetheless, there is a very well-known consistency condition which arises between the interference of tree- and loop-amplitudes: the classic Green-Schwarz anomaly cancellation mechanism! It would be very interesting to pursue, exhaustively, the extent to which similarly inconsistent-seeming tree-amplitudes could be made consistent upon including loops. But this is left for future work.

In conclusion, these results confirm the Coleman-Mandula and the Haag-Lopuszanski-Sohnius theorems for exclusively massless states in four-dimensions [94][95]. Through assuming a constructible, non-trivial, S-matrix that is compatible with locality and unitarity, we see that the maximal structure of non-gravitational interactions between low-spin particles is that of compact Lie groups. Only through coupling to gravitons and gravitinos can additional structure be given to the massless tree-level S-matrix (at four-points). This additional structure is simply supersymmetry; it relates scattering amplitudes with asymptotic states of different spin, within the same theory. Further, no gravitational, marginal, or relevant interaction may consistently couple to massless asymptotic states with spin greater than two.

5.8 Acknowledgements

The authors thank Nima Arkani-Hamed for suggesting this topic, and for his continual encouragement and guidance throughout the process. We thank Clifford Cheung, Yu-tin Huang, David Skinner, and Jaroslav Trnka, and Alexander Zhiboedov for helpful and encouraging discussions on various topics. Finally, we thank Stanley Deser, Sylvester James Gates, Massimo Porrati, Augusto Sagnotti, Mirian Tsulaia and collaborators, for helpful discussions of previous, related, work.

Chapter 6

Gravitons, Permutation Invariance, and Bonus-Scaling

Mysteries abound at the interface between General Relativity and Quantum Field Theory. Particularly, graviton scattering amplitudes in maximally supersymmetric $\mathcal{N}=8$ Supergravity have surprisingly soft behavior in the deep ultraviolet (UV). To four loops, it has been shown that the critical dimension of supergravity is the same as $\mathcal{N}=4$ Super Yang-Mills, a conformally invariant theory free of UV divergences [39]. This result was obtained through the peculiar BCJ duality between color and kinematics, which relates graviton amplitudes to the squares of gluon amplitudes [96][34]. Other arguments, based the non-linearly realized $E_{7(7)}$ symmetry of $\mathcal{N}=8$ supergravity, predict UV finiteness to six-loops [97]. Yet others hint at a full finiteness (see e.g. [98]).

Standard perturbative techniques, i.e. Feynman diagrams, lead to incredibly complicated expressions, and obfuscate general features of the theory. Reframing the discussion in terms of the modern analytic S-matrix has so far proven incredibly useful for discussing Yang-Mills theory (for example, in Ref. [99]), and may provide crucial insights into quantum gravity as well. The on-shell program offers a different perspective on the principles of locality and unitarity, and their powerful consequences [79][78, 80]. It also provides a computational powerhouse, the BCFW on-shell recursion relation [20].

Briefly, if two external momenta in the amplitude A_n are subjected to the on-shell BCFW shift:

$$p_1^\mu \to p_1^\mu + zq^\mu \qquad p_2^\mu \to p_2^\mu - zq^\mu \tag{6.1}$$

and $A_n(z) \to 0$ for large z, then $A_n(z=0)$ can be recursively constructed from lower-point on-shell amplitudes:

$$A_n = \oint \frac{dz}{z} A_n(z) = \sum_{\{L\}} \frac{A_L(\hat{1}, \{L\}, \hat{P}) A_R(\hat{P}, \{R\}, \hat{n})}{P^2}. \tag{6.2}$$

Initial proofs required sophisticated Feynman diagram analyses, and found that gluon amplitudes have the minimum scaling of z^{-1}, but that graviton amplitudes have a "bonus", seemingly unnecessary, scaling of z^{-2} [20][100][101][90][25][102][16]. Surprisingly, Ref. [79] found that a fully on-shell proof of BCFW constructability actually requires this improved scaling for gravitons, in order for Eq. (6.2) to satisfy unitarity. The bonus scaling is not just a "bonus", but a critical property of General Relativity. This z^{-2} scaling, also present in the case of non-adjacent gluon shifts [103], implies new residue theorems:

$$0 = \oint A_n(z) dz = \sum_{\{L\}} z_P \frac{A_L(\hat{1}, \{L\}, \hat{P}) A_R(\hat{P}, \{R\}, \hat{n})}{P^2}, \tag{6.3}$$

i.e., new relations between terms in Eq. (6.2): the bonus relations. The bonus scaling and the bonus relations have a number of important implications. In [104], it was shown that BCJ relations can be extracted from bonus relations. In the case of gravity, bonus relations have been used to simplify tree level calculations [105]. At loop level, the large z scaling of the BCFW shift corresponds to the high loop momenta limit; unsurprisingly improved scaling implies improved UV behavior [16][35] (see also chapter 2).

In this chapter, we prove that the inherent Bose-symmetry between gravitons directly implies this improved bonus scaling, completing the arguments of Ref. [79]. Bose-symmetry in General Relativity endows it with a purely on-shell description and constrains its UV divergences[1]. We further apply the same argument to gauge theories and gravity in various dimensions.

6.1 Completing on-shell constructability.

Reference [79] first assumes n-point and lower amplitudes scale as z^{-1}—thereby ensuring Eq. (6.2) holds—and then checks if the BCFW expansion of the $(n+1)$-point amplitude factorizes correctly on all channels. Factorization on all channels is taken to define the amplitude. Correct factorization in most channels requires z^{-1} scaling of lower point amplitudes. However, some channels do not

[1]The better than expected UV behavior was also at least partially understood from the "no-triangle" hypothesis of $\mathcal{N}=8$ supergravity, as a consequence of crossing symmetry and the colorless nature of gravitons in Ref. [101]

factor correctly without improved z^{-2} scaling, as well as a z^6 scaling on the "bad" shifts. In the following, we present a proof for both of these scalings.

Essentially, the argument rests on a very simple observation: any symmetric function $f(i,j)$, under deformations $i \to i + zk$, $j \to j - zk$, must scale as an even power of z. In particular, any function with a strictly better than $\mathcal{O}(1)$ large z behavior (no poles at infinity), is automatically guaranteed to decay at least as z^{-2}.

Although straightforward, this is not manifest when constructing the amplitude. BCFW terms typically scale as z^{-1}, but only specific pairs have canceling leading z^{-1} pieces. Similarly, the bad BCFW shift behavior of z^6 is only obtained when the leading z^7 pieces cancel in pairs.

Consider the five point amplitude in $\mathcal{N} = 8$ SUGRA, exposed by the $[1,5\rangle$ BCFW shift, where $|1] \to |1] - z|5]$ and $|n\rangle \to |n\rangle + z|1\rangle$,

$$\begin{aligned} M_5 =& \frac{M(123P)\times M(P45)}{P_{123}^2} + (4 \leftrightarrow 3) + (4 \leftrightarrow 2) \\ =& \frac{[23][45]}{\langle 12\rangle\langle 13\rangle\langle 23\rangle\langle 24\rangle\langle 34\rangle\langle 45\rangle\langle 15\rangle^2} + \frac{[24][35]}{\langle 12\rangle\langle 14\rangle\langle 24\rangle\langle 23\rangle\langle 43\rangle\langle 35\rangle\langle 15\rangle^2} + \frac{[43][25]}{\langle 14\rangle\langle 13\rangle\langle 43\rangle\langle 42\rangle\langle 32\rangle\langle 25\rangle\langle 15\rangle^2}, \end{aligned} \tag{6.4}$$

with the SUSY-conserving delta-function stripped out.

Under a $[2,3\rangle$ shift, the first term scales as z^{-2}, while the other two scale as z^{-1}. However, their sum (now symmetric in 2 and 3) scales as z^{-2}: the whole amplitude has the correct scaling. This pattern holds true in general. Where present, z^{-1}'s cancel between pairs of BCFW terms, $M_L(K_L, i, P) \times M_R(-P, j, K_R)$ and $M_L(K_L, j, P) \times M_R(-P, i, K_R)$. Further, terms without pairs over saturate the bonus scaling.

One such example is $M(1^-2^-3^+P^+)M(P^-4^-5^+6^+)$, appearing in M_6^{NMHV}. Under a $[4,3\rangle$ shift, it has no corresponding pair: $M(1^-2^-4^-P)M(P\,3^+5^+6^+)$ vanishes for all helicities h_P. Luckily, it turns out these types of terms have a surprisingly improved scaling of z^{-9}. Hence, they never spoil the scaling of the full amplitude.

In the next section we classify and prove the scalings of all possible BCFW terms. Following this, we demonstrate how leading z pieces cancel between BCFW terms.

6.1.1 BCFW terms under secondary z-shifts.

Consider the $[1, n\rangle$ BCFW expansion of a n-point GR tree amplitude M_n (where $\tilde{\lambda}_1 \to \tilde{\lambda}_1 - w\tilde{\lambda}_n$, $\lambda_n \to \lambda_n + w\lambda_1$):

$$M_n = \sum_{L,R} \frac{M_L(\hat{1}, \{L\}, \hat{P}) M_R(-\hat{P}, \{R\}, \hat{n})}{P^2} \tag{6.5}$$

We would like to understand how BCFW terms in $\mathcal{M}_n$ scale under secondary $[i, j\rangle$ z-shifts

$$\tilde{\lambda}_i \to \tilde{\lambda}_i - z\tilde{\lambda}_j \qquad\qquad \lambda_j \to \lambda_j + z\lambda_i. \tag{6.6}$$

We recall two features of these terms as they appear in Eq. (6.5). First, the value of the primary deformation parameter $w = w_P$, which accesses a given term, is

$$w_P = \frac{P^2}{\langle 1|P|n]}, \tag{6.7}$$

and, on this pole, the intermediate propagator factorizes:

$$\hat{P}^{\alpha\dot{\alpha}} = \frac{\{[\tilde{\lambda}_n|P\}^{\alpha} \ \{\langle\lambda_1|P\}^{\dot{\alpha}}}{\langle\lambda_1|P|\tilde{\lambda}_n]} = \frac{\big|\lambda_P\big\rangle\big[\tilde{\lambda}_P\big|}{\langle 1|P|n]} \equiv |\hat{P}\rangle[\hat{P}|. \tag{6.8}$$

The little-group ambiguity amounts to associating the denominator with either λ_P, $\tilde{\lambda}_P$, or some combination of them. In what follows, we find it easiest to associate it entirely with the anti-holomorphic spinor, $|\hat{P}] = |\tilde{\lambda}_P]/\langle 1|P|n]$—see Eq. (6.9), below.

With this in hand, we now turn to the large z scalings of the various BCFW terms, subjected to the secondary z-shifts in Eq. (6.6). There will be two different types of BCFW terms: those with both i and j within the same subamplitude, and those with i and j separated by the propagator. The former inherit all z dependence from the lower point amplitudes in the theory, since the secondary shift acts like a usual BCFW shift on the subamplitude. The latter are more complicated, since the z shift affects the subamplitudes in several ways besides the simple shifts on i and j.

Specifically, both w_P and the factorized form of the internal propagator acquire z dependence:

$$\begin{aligned}
w_P = \frac{P^2}{\langle 1|P|n]} &\quad\longrightarrow\quad \frac{P^2 + z\langle i|P|j]}{\langle 1|P|n] + z\langle 1i\rangle[jn]}, \\
\big|\widehat{P}\big\rangle^{\alpha} \equiv \frac{\big|\big([n|P)\big\rangle^{\alpha}}{[jn]} &\quad\longrightarrow\quad \big|\widehat{P}\big\rangle^{\alpha} + z\big|i\big\rangle^{\alpha}, \\
\big|\widehat{P}\big]^{\dot\alpha} \equiv \frac{\big|\big(\langle 1|P)\big]^{\dot\alpha}}{\langle 1|P|n]/[jn]} &\quad\longrightarrow\quad \frac{\big|\tilde\lambda_P\big]^{\dot\alpha} - z\langle 1i\rangle|j]^{\dot\alpha}}{\langle 1|P|n]/[jn] - z\langle 1i\rangle}.
\end{aligned} \tag{6.9}$$

With this factorized form of the propagator, it turns out that the left- and right-hand subamplitudes have well defined individual z scalings, which depend only on the helicity choices for i^{h_i}, j^{h_j} and P^h:

$$\begin{aligned}
M_L(i^- P^-) &\sim z^{-2} & M_R(j^- P^-) &\sim z^{+2} \\
M_L(i^- P^+) &\sim z^{-2} & M_R(j^- P^+) &\sim z^{+2} \\
M_L(i^+ P^-) &\sim z^{+6} & M_R(j^+ P^-) &\sim z^{+2} \\
M_L(i^+ P^+) &\sim z^{-2} & M_R(j^+ P^+) &\sim z^{-6}.
\end{aligned} \tag{6.10}$$

The scaling of a full BCFW term $M_L M_R/P^2$ can then be easily determined from these values, which we prove in two steps.

First, note that the large z scalings on the left of Eq. (6.10) match the familiar BCFW scalings of full amplitudes. We prove this by showing that the large z behavior of the left-hand subamplitude maps isomorphically onto a BCFW shift of M_L. Looking at Eq. (6.9), we see that, in the large z limit, the spinors of i and P become

$$\begin{aligned}
\lambda_i &\longrightarrow \lambda_i \lambda_P \longrightarrow z\lambda_i \\
\tilde\lambda_i &\longrightarrow -z\tilde\lambda_j\tilde\lambda_P \longrightarrow \tilde\lambda_j,
\end{aligned} \tag{6.11}$$

which is just a regular BCFW $[i, P\rangle$ shift within the left-hand subamplitude.

Now we turn to the slightly unusual scalings on the right-hand side of Eq. (6.10). With the little-group choice in Eq. (6.9), the left-hand subamplitude has exactly the correct spinor variables to map onto the usual BCFW shift. Now observe that, starting with the other little-group choice for the spinors on the z shifted internal propagator, we obtain the usual BCFW scalings on this side:

$$\begin{aligned}
M_R(j^-P^-) &\sim z^{-2} \\
M_R(j^-P^+) &\sim z^{+6} \\
M_R(j^+P^-) &\sim z^{-2} \\
M_R(j^+P^+) &\sim z^{-2}.
\end{aligned} \tag{6.12}$$

Proving these results is identical to the previous reasoning for the left-hand subamplitude.

It becomes clear now that to get the other half of the scalings, we need only account for the change in z scaling when switching the $1/\langle 1|P(z)|n]$ factor between λ_P and $\tilde{\lambda}_P$. Assume that the spinors of the propagator appear with weights[2]:

$$M_R \propto \big(|P\rangle\big)^a \big(|P]\big)^b, \tag{6.13}$$

where $-a+b=2h_P$, and h_P is the helicity of the internal propagator as it enters the right-hand subamplitude. Now, in the limiting cases where $1/\langle 1|P(z)|n]$ is entirely associated with $|\lambda_P\rangle$ or $|\tilde{\lambda}_P]$ the amplitude scales as:

$$M_R \propto \left(\frac{|\lambda_P\rangle}{\langle 1|P|n]}\right)^a \big(|\tilde{\lambda}_P]\big)^b \to z^s, \quad \text{or} \tag{6.14}$$

$$M_R \propto \big(|\lambda_P\rangle\big)^a \left(\frac{|\tilde{\lambda}_P]}{\langle 1|P|n]}\right)^b \to z^t, \tag{6.15}$$

where s is the BCFW large z scaling exponent, obtained in Eq. (6.12), and t is the related scaling, for the other internal little-group choice. It follows that $s-t=b-a=2h_P$, and so the t scalings can be easily derived as $t=s\pm 4$, depending on the helicity of the propagator.

Having proven all eight scaling relations in Eq. (6.10), we can classify the scaling behavior of all possible types of BCF terms with i and j in different subamplitudes. For these terms the propagator contributes a z^{-1} to each term, and so from Eq. (6.10) we obtain eight possible types of terms:

- $M_L(i^+P^-)M_R(j^-P^+)/P^2$ scales as z^{+7}, (6.16)
- $M_L(i^-P^-)M_R(j^+P^+)/P^2$ scales as z^{-9}, (6.17)
- The other six BCFW terms scale as z^{-1} . (6.18)

[2]In general, the spinors need not appear with uniform homogeneity. The analysis below still holds, but must be applied term by term. The same caveat applies to Eqs. (6.24) and (6.26).

In the next section we will see how pairing terms improves these scalings by one power of z, such that we recover the required z^{-2} and z^6 scalings.

Finally, while the individual scalings in Eq. (6.10) are not invariant under z dependent little-group rescalings on the internal line $\widehat{P}(z)$, the above results for full BCFW terms are invariant under these rescalings.

6.1.2 Improved behavior from symmetric sums.

We first study $[+,+\rangle$ and $[-,-\rangle$ shifts, with scalings in Eq. (6.18). Define $M_L(K_L,i,P)\times M_R(-P,j,K_R)/P^2 \equiv M(i|j)$, where K_L is the momenta from the other external states on the left-hand subamplitude. We wish to show that in the large z limit

$$M(i|j) = -M(j|i). \tag{6.19}$$

so the leading z^{-1} pieces cancel in the symmetric sum of BCFW terms, $M(i|j)+M(j|i)$.

Because i and j have the same helicity, $M(j|i)$ is obtained directly from $M(i|j)$ by simply swapping labels:

$$M(i|j) = M(\lambda_i,\tilde\lambda_i,\lambda_j,\tilde\lambda_j) \tag{6.20}$$

$$M(j|i) = M(\lambda_j,\tilde\lambda_j,\lambda_i,\tilde\lambda_i) \tag{6.21}$$

In the large z limit, these become

$$M(i|j) = M(\lambda_i,-z\tilde\lambda_j,z\lambda_i,\tilde\lambda_j) \tag{6.22}$$

$$M(i|j) = M(z\lambda_i,\tilde\lambda_j,\lambda_i,-z\tilde\lambda_j) \tag{6.23}$$

The two have equal z scaling, and so can only differ by a relative sign. The spinors appear with weights

$$\begin{aligned} M(i|j) &\propto \langle ij\rangle^F[ij]^G(\lambda_i)^a(\tilde\lambda_i)^b(\lambda_j)^c(\tilde\lambda_j)^d \\ M(j|i) &\propto \langle ji\rangle^F[ji]^G(\lambda_j)^a(\tilde\lambda_j)^b(\lambda_i)^c(\tilde\lambda_i)^d, \end{aligned} \tag{6.24}$$

while in the large z limit, the leading terms are

$$\begin{aligned} M(i|j) &\propto z^{b+c}\left(\langle ij\rangle^F[ij]^G(\lambda_i)^a(-\tilde{\lambda}_j)^b(\lambda_i)^c(\tilde{\lambda}_j)^d\right) \\ M(j|i) &\propto z^{a+d}\left(\langle ji\rangle^F[ji]^G(\lambda_i)^a(\tilde{\lambda}_j)^b(\lambda_i)^c(-\tilde{\lambda}_j)^d\right). \end{aligned} \tag{6.25}$$

These cancel if and only if $F+G+b+d=\text{odd}$. First, from Eq. (6.18), $M(a|b)$'s scale as z^{odd}. So $b+c=a+d=\text{odd}$. Second, by helicity counting in Eq. (6.24), we know $-F+G-c+d=2h_j=\text{even}$. Therefore, we obtain the required result, and the leading z^{-1} pieces cancel.

For the $[-,+\rangle$ and $[+,-\rangle$ shifts a simple modification of the above argument is required. This is because we now expect the cancellation to occur between the pair terms $M_L(K_L,i^-,P^+)\times M_R(-P^-,j^+,K_R)/P^2$ and $M_L(K_L,j^+,P^-)\times M_R(-P^+,i^-,K_R)/P^2$. Switching different helicity particles requires us to flip the propagator's helicity as well. It can be shown that, in the large z-limit, $M_L(K_L,i^-,P^+)=M_L(K_L,j^+,P^-)$; likewise for the right-hand subamplitude. Note that switching i^- and j^+ requires more care now: functionally, the correct label swaps for M_L are $i\to P$, $P\to j$ while for M_R $j\to P$ and $P\to i$. Therefore we can write, as above,

$$\begin{aligned} M_L(i^-,P^+) &\propto \langle iP\rangle^F[iP]^G(\lambda_i)^a(\tilde{\lambda}_i)^b(\lambda_P)^k(\tilde{\lambda}_P)^l \\ M_L(j^+,P^-) &\propto \langle Pj\rangle^F[Pj]^G(\lambda_P)^a(\tilde{\lambda}_P)^b(\lambda_j)^k(\tilde{\lambda}_j)^l. \end{aligned} \tag{6.26}$$

Crucially, the large z limit is also different for the two subamplitudes, since the limits (6.11) were obtained with $i\in P$. The second subamplitude instead has $j\in P$, and in this case the limits are $\lambda_P\to -z\lambda_i$ and $\tilde{\lambda}_P\to\tilde{\lambda}_j$. In the large z limit then identical counting as above shows that $a+b=\text{even}$, and the same will hold for M_R. The propagator is antisymmetric in the large z limit under swapping i and j, and therefore the leading z pieces cancel as expected. This cancellation reduces the leading z^{-1} and z^{+7} scalings for the opposite helicity shifted BCF terms in the previous section, down to the well known z^{-2} and z^{+6} BCFW scalings for GR. This completes the proof of the bonus scaling for GR, and closes the final gap in the on-shell proof of BCFW in GR Ref. [79].

6.1.3 Analysis of the full amplitude.

The simple argument we used above can be applied directly to the whole amplitude, if we restrict to like-helicity shifts. Consider

$$A_n(i,j)\propto\langle ij\rangle^F[ij]^G(\lambda_i)^a(\tilde{\lambda}_i)^b(\lambda_j)^c(\tilde{\lambda}_j)^d. \tag{6.27}$$

If this amplitude is manifestly symmetric under exchange of two (bosonic) particle labels, then $A_n(i,j) = A_n(j,i)$, which fixes $a = c$, $b = d$, and $F+G = \text{even}$. By helicity counting, $-F+G-a+b = 2h_i = \text{even}$, and then $a+b = \text{even}$. So, under a $[i,j\rangle$ shift,

$$A_n(i(z), j(z)) \sim z^{b+c} = z^{a+b} = z^{\text{even}}. \tag{6.28}$$

This same logic holds in Eq. (6.27), even if the shifted lines are identical fermions. Permuting labels i and j again forces $a = c$, and $b = d$, and $F + G = \text{odd}$. But so must $2h_i = -F + G - a + b$. Hence $a + b$ remains even. BCFW shifts of identical particles, bosons or fermions, fix z^{even} scaling at large z.

To understand the opposite-helicity shifts, we are led to consider pure GR as embedded within maximal $\mathcal{N} = 8$ SUGRA. Amplitudes in maximal supergravity do not distinguish between positive and negative helicity graviton states. Using the methods of [17] to truncate to pure GR, we recover the usual BCFW scalings.

As an interesting corollary of our four-dimensional analysis, the large z scaling of gravity amplitudes in three dimensions is drastically improved to z^{-4}. Due to the fact that the little group in three dimensions is a discrete group, the BCFW deformation is non-linear. In particular the three dimensional spinors shift as [106]:

$$\lambda_i(z){=}\mathrm{ch}(z)\lambda_i{+}\mathrm{sh}(z)\lambda_j\;,\;\lambda_j(z){=}\mathrm{sh}(z)\lambda_i{+}\mathrm{ch}(z)\lambda_j \tag{6.29}$$

where $\mathrm{ch}(z) = (z + z^{-1})/2$ and $\mathrm{sh}(z) = (z - z^{-1})/2i$. Thus, momenta shift as

$$p_i(z) = \overline{P_{ij}} + yq + \frac{1}{y}\tilde{q}\;\;,\;\; p_j(z) = \overline{P_{ij}} - yq - \frac{1}{y}\tilde{q} \tag{6.30}$$

where $\overline{P_{ij}} = \frac{p_i+p_j}{2}$, $y = z^2$, and $q, \tilde{q}$ can be read off from Eq. (6.29). Now let's consider three-dimensional gravity amplitudes that arise from the dimension reduction of four-dimensional gravity theory. The degrees of freedom are given by a dilaton and a scalar. Since both are bosons, little group dictates that one must have even power of λ_i. Thus the large z behavior of gravity amplitudes is completely dictated by Eq. (6.30). Permutation invariance then requires the function to be symmetric under $y \leftrightarrow -y$, and so must be an even power of y. Thus if gravity amplitudes can be constructed via BCFW shift, the large z asymptotic behavior must be at most $y^{-2} = z^{-4}$. Indeed it is straightforward to check that the four-point $\mathcal{N} = 16$ supergravity amplitude behaves as z^{-4} under a super-BCFW shift. This is to be compared with the z^{-1} scaling of superconformal

Chern-Simons theory [106].

More generally, BCFW shifts in $d \geq 4$ take the form,

$$p_i^\mu(z) = p_i^\mu + z\, q^\mu \qquad p_j^\mu(z) = p_j^\mu - z\, q^\mu, \tag{6.31}$$

where q is null and orthogonal to p_i and to p_j. External wave-functions of shifted boson lines also shift [25]. For identical bosons, Bose-symmetry disallows z^{odd} scaling, as it would introduce a sign change under label swaps. Identical fermions shift similarly; here the antisymmetric contraction of the identical spinor wave-functions absorbs their exchange-sign. BCFW shifts of identical particles must scale as z^{even} for large z in dimensions $d \geq 4$.

Symmetry between identical particles is crucial for these cancellations to occur. Gluon partial amplitudes are not permutation invariant: distinct gluons generally have different colors. This spoils the permutation invariance—as is clear from z^{-1} drop-off of adjacent shifts of a color-ordered tree amplitude in Yang-Mills. Gravitons, however, are unique: they cannot have different "colors" [75]. Thus graviton amplitudes are invariant under permutations from the outset: the discrete symmetry group of graviton amplitudes is larger than for gluon amplitudes. Consequently, gravity amplitudes are softer in the deep-UV than Yang-Mills amplitudes.

6.2 Bose-symmetry and color in Yang-Mills.

Finally, we explore the interplay between color and the large z structure of Yang-Mills amplitudes. For ease, we focus on $A_4^{\text{tree}}(1^-, 2^-, 3^+, 4^+)$. It can be written in terms of color-ordered partial amplitudes as

$$\frac{A_4(1^-2^-3^+4^+)}{\langle 12\rangle^2[34]^2} = \frac{Tr(1234)}{st} + \frac{Tr(1243)}{su} + \frac{Tr(1324)}{tu}. \tag{6.32}$$

Under a $[1,2\rangle$ shift, only t and u shift, and in opposite directions: $\widehat{t}(z) = t + z\langle 1|4|2]$, and $\widehat{u}(z) = u - z\langle 1|4|2]$. The term proportional to $Tr(1324)$ scales as z^{-2}, while the other two scale as z^{-1}. The leading z terms,

$$\frac{A_4(\widehat{1}^-, \widehat{2}^-, 3^+, 4^+)}{\langle 12\rangle^2[34]^2} \sim \frac{Tr(1234) - Tr(1243)}{z\langle 1|4|2]\, s} + \cdots\,, \tag{6.33}$$

cancel when gluons 1 and 2 are identical, and $T_1 = T_2$.

Cancellation of z^{-1} terms must hold for general tree amplitudes when the gluons have the same color labels. However, only BCFW shifts of lines that are *adjacent* in color-ordering cancel pairwise as in Eq. (6.33). For color-orderings where this shift is non-adjacent, there are no pairs of BCF terms with canceling z^{-1}-terms. This implies that good non-adjacent BCFW shifts in gluon partial amplitudes must scale as z^{-2}.

6.3 Future directions and concluding remarks.

We have shown the z^{-2} bonus scalings/relations, crucial for consistent on-shell contraction of Gravitational S-matrices, follow from Bose-symmetry. Similar z^{-1} cancellations occur in QED and GR [107]. Further, Bose-symmetry alone implies z^{-2} drop-off of non-adjacent BCFW shifts in Yang-Mills. More broadly, BCFW shifts of identical particles—bosons and fermions—must scale as z^{even} in general settings, beyond $d = 4$.

Graviton amplitudes in Refs. [108][109][110], which manifest permutation symmetry, also manifest z^{-2} drop-off. This is not a coincidence: permutation symmetry automatically implies bonus behavior. A better understanding of gravity should be tied to more natural manifestations of permutation invariance. However, not all improved scalings obviously come from permutation invariance. Notably, Hodges' observation that BCFW-terms, built from "bad" "opposite helicity" z^{-1} $\mathcal{N} = 7$ SUGRA shifts, term-by-term scale as z^{-2} [111]. As the legs are not identical, permutation invariance is not prominent in the proof [112].

Permutation invariance has unrecognized and powerful consequences even at tree level. Do new constraints appear when accounting for it in other shifts? Does it have non-trivial consequences at high-loop orders in $\mathcal{N} = 8$ SUGRA, or $\mathcal{N} = 4$ SYM? Would mandating it expose new facets of the "Amplituhedron" of Ref. [99]?

6.4 Acknowledgements:

We thank Aleksey Cherman, Jaroslav Trnka and Song He for careful readings and comments. We specially thank Yu-tin Huang for thoughtful conversations, especially regarding color in QCD, the discussion on z^{-4} scaling in 3d gravity and generalizing Eq. (6.28) to arbitrary dimension, and Nima Arkani-Hamed for discussions and encouraging us to clarify the scalings of individual BCFW terms.

Chapter 7

T-reflections

7.1 Introduction

In this chapter we observe that the partition functions $Z(\beta = 1/T)$ of a number of physical systems have a curious symmetry under a 'reflection' of the temperature, $T \to -T$:

$$Z(+\beta) = e^{i\gamma} Z(-\beta)\,, \tag{7.1}$$

where $e^{i\gamma}$ is a theory-dependent but temperature-independent complex number of modulus one. Note that asking about the behavior of $Z(\beta)$ under T-reflection requires information about all of the energy levels, E_n, and degeneracies, d_n, in

$$Z(\beta) = \sum_{n=1}^{\infty} d_n e^{-\beta E_n}\,. \tag{7.2}$$

Naively, the sum representation of $Z(\beta)$ in Eq. (7.2) converges only for $\beta > 0$, and directly testing Eq. (7.1) by sending $\beta \to -\beta$ in $Z(\beta)$ hinges on being able to resum the series into some sort of closed form. Systems where such resummations are known are rare.

In this chapter we consider a variety of quantum systems where exact expressions for $Z(\beta)$ are available and show that they have the T-reflection symmetry summarized in Eq. (7.1). Many, but not all, of our example systems are conformal field theories (CFTs). The presence of T-reflection symmetry requires a unique choice of ground-state energy; shifting it by $-\Delta$ multiplies $Z(\beta)$ by a non-invariant factor of $e^{\beta\Delta}$. As T-reflection fixes the ground-state energy, it resembles a space-time symmetry. However, it seems to persist in interacting theories without conformal- or supersymmetry, and may be of help in resolving the cosmological constant problem [1].

Our most most intricate and surprising examples are two-dimensional CFT minimal models. Here, the role of the temperature is played by the modular parameter τ ($\mathrm{Im}(\tau) > 0$) of the torus, and T-reflection then acts on τ as [2]

$$R : \tau \to -\tau \,. \tag{7.3}$$

We show that all characters of the Virasoro algebra which appear in any given minimal model are invariant under R-transformations/T-reflections. These characters, denoted $\chi^{p,p'}_{r,s}(q)$ and $\bar{\chi}^{p,p'}_{m,n}(\bar{q})$, are the building blocks of two-dimensional CFT minimal model torus partition functions, and hence all minimal models are invariant under T-reflections. This R-transformation should be contrasted with the familiar modular transformation [116] of the torus. The modular group $PSL(2, Z)$ is generated by two generators S and T, which act on τ as

$$S : \tau \to -\frac{1}{\tau} \,, \quad T : \tau \to \tau + 1 \,. \tag{7.4}$$

The transformation R does not commute with the generators of $PSL(2, Z)$. In fact R, S, T generate the extended modular group $PGL(2, Z)$.

7.2 Oscillators in Quantum Mechanics

Let us begin with the simplest example, a bosonic simple harmonic oscillator. If we measure energy in units of the oscillator frequency, and denote the ground state energy by Δ, the partition function is given by

$$Z_{\mathrm{SHO}}(\beta) = \sum_{n=0}^{\infty} e^{-\beta(n+\Delta)} = \frac{e^{\beta(1/2-\Delta)}}{2\sinh(\beta/2)} \,. \tag{7.5}$$

This is T-reflection symmetric, with $\gamma = \pi$, only if the ground-state energy takes the naive value, $\Delta = +1/2$.

Similarly, the thermal partition function for a fermionic oscillator can be written as,

$$Z(\beta) = 2e^{\beta(1/2+\Delta)} \cosh(\beta/2) \,. \tag{7.6}$$

If $\Delta = -1/2$ this has a T-reflection symmetry ($\gamma = 0$). As emphasized above, shifting the ground-state energy breaks any symmetry under T-reflections. One can also check that the twisted partition

[1] Some potentially related ideas have appeared in e.g. [113],[114]

[2] The same transformation was considered in, e.g., Ref. [115], albeit in a different context.

function $\tilde{Z} = \mathrm{Tr}\,(-1)^{\mathcal{F}} e^{-\beta H}$ (where $\mathcal{F}$ is fermion number) for the fermionic oscillator has T-reflection symmetry if and only if $\Delta = -1/2$.

7.3 Examples in Field Theory

7.3.1 Free $d = 2$ CFTs

Next we consider some of the simplest QFT examples of systems with T-reflection symmetry, which are free $d = 2$ CFTs. The torus partition functions of $d = 2$ CFTs decompose into linear combinations of products of holomorphic and anti-holomorphic Virasoro characters, which can themselves be thought of as partition functions for left-moving and right-moving modes. Because $\tau = i\beta/L$ is the shape modulus of the torus — $T^2 = S^1_L \times S^1_\beta$ — and $q = e^{2\pi i \tau}$, the T-reflection properties of the partition functions are fixed by the transformation properties of the characters under $q \to q^{-1}$.

We now examine the q-inversion properties of the single holomorphic character

$$\chi_{\mathrm{s}}(q) = q^{-\frac{1}{24}} \prod_{n=1}^{\infty} \frac{1}{(1-q^n)} = \frac{1}{\eta(q)}\,. \tag{7.7}$$

contributing to the free scalar CFT (the analysis for anti-holomorphic sector is completely analogous). We find

$$\begin{aligned} \chi_{\mathrm{s}}\left(q^{-1}\right) &= q^{+\frac{1}{24}} \prod_{n=1}^{\infty} \frac{1}{1-q^{-n}} \\ &= \left\{ q^{\frac{1}{12}} \prod_{a=1}^{\infty} q^a \prod_{b=1}^{\infty} \frac{1}{(-1)} \right\} \chi_{\mathrm{s}}(q) \\ &= \frac{q^{\frac{1}{12}} q^{\zeta(-1)}}{(-1)^{\zeta(0)}} \chi_{\mathrm{s}}(q) = i \chi_{\mathrm{s}}(q)\,, \end{aligned} \tag{7.8}$$

where we used zeta-function regularization [$\zeta(s)$ is the Riemann zeta function]. So $\chi_{\mathrm{s}}(q)$ has T-reflection symmetry with $\gamma = \pi/2$.

For additional insight into this result — especially on the reason for the appearance of a regulator above — we can compute the same partition function directly from the Euclidean path integral for a scalar field. This gives

$$-\ln Z(\beta) = V_0 L \beta + \sum_{m=1}^{\infty} \left[\frac{\beta \omega_m}{2} + \log(1 - e^{-\beta \omega_m}) \right] \tag{7.9}$$

where $\omega_m = 2\pi m/L$ and V_0 is the bare vacuum energy from the Lagrangian. One can evaluate

(7.9) as written, which yields $Z(\beta) = \chi_s(\beta)$, or evaluate (7.9) *after* sending $\beta \to -\beta$, which yields $Z(-\beta) = \chi_s(-\beta)$. Note that this amounts to what we did in the SHO example in (7.5) for each of the infinite number of oscillators in the field theory. It is important to note that UV divergences appear in the calculations of both $Z(\beta)$ and $Z(-\beta)$ in QFTs, and hence the frequency sums must be regularized and renormalized, with the divergences absorbed in counterterms in *e.g.* V_0. Indeed, the second sum in (7.9) is precisely the same quantity as the infinite product in (7.7), while the UV divergences and renormalization issues which appear in the first sum in (7.9) are hidden in the definition of the Casimir energy $\Delta = -1/24$ in (7.7). One can then verify that $Z(\beta)$ has T-reflection symmetry so long as the renormalized vacuum energy is set to the Casimir energy, just as in (7.8), and also reproduce the $\gamma = \pi/2$ factor which we found using ζ-function regularization above. Similar remarks apply to all of our QFT examples.

We next check the free fermion CFT partition functions on T^2 (see *e.g.* [117]). We can take periodic (R) or anti-periodic (NS) boundary conditions for each of the two cycles of T^2, which yields three non-trivial distinct partition functions R-NS, NS-R, and NS-NS. (R-R fermions have a zero mode which nullifies their associated partition function.) The difference between R-NS and NS-R comes because we take the Hamiltonian to be associated with isometries along one of the S^1's in T^2.

For NS-NS fermions, the character takes the form

$$\chi_{\mathrm{NS-NS}}(q) = q^{-\frac{1}{48}} \prod_{n=0}^{\infty} (1 + q^{n+1/2})\,. \tag{7.10}$$

Under T-reflection, we have

$$\begin{aligned} \chi_{\mathrm{NS-NS}}(q^{-1}) &= q^{+\frac{1}{48}} \prod_{n=0}^{\infty} (1 + q^{-(n+1/2)}) \\ &= \left\{ q^{+\frac{1}{24}} \prod_{a=0}^{\infty} q^{-(n+1/2)} \right\} \chi_{\mathrm{NS-NS}}(q) \\ &= q^{\frac{1}{24} - \zeta(-1,\frac{1}{2})} \chi_{\mathrm{NS-NS}}(q) = \chi_{\mathrm{NS-NS}}(q)\,, \end{aligned} \tag{7.11}$$

where $\zeta(s,a)$ is the Hurwitz zeta function, and $\zeta(-1,x) = -\frac{x^2}{2} + \frac{x}{2} - \frac{1}{12}$.

Similar arguments allow us to verify that the R-NS and NS-R partition functions,

$$\chi_{\mathrm{R-NS}}(q) = q^{\frac{1}{24}} \textstyle\prod_{n=0}^{\infty} (1 + q^n)\,, \tag{7.12}$$

$$\chi_{\mathrm{NS-R}}(q) = q^{-\frac{1}{48}} \textstyle\prod_{n=1}^{\infty} (1 - q^{n-1/2})\,, \tag{7.13}$$

are each separately covariant under T-reflection.

7.3.2 Free gauge theories in $d = 4$-dimensions

We now consider gauge theories on $S^3_R \times S^1_\beta$ where R and β are the radii of the 3-sphere and the thermal circle respectively. Asymptotically-free gauge theories with gauge groups of rank N with a strong scale Λ become arbitrarily weakly coupled when $R\Lambda \to 0$. Indeed, in the $R\Lambda \to 0$ limit they behave as if they were free CFTs on $S^3_R \times S^1_\beta$ with a color-singlet constraint from the Gauss law. Refs. [118][119][120] showed how to compute their partition functions in this free CFT limit. The gauge theory partition functions can be written in terms of so-called 'single-particle' partition functions, which can be calculated using CFT techniques and are given by

$$z_S(x) = \frac{(x^{1/2} + x^{-1/2})}{(x^{-1/2} - x^{1/2})^3}\,, \tag{7.14}$$

$$z_F(x) = \frac{2^3}{(x^{-1/2} - x^{1/2})^3}\,, \tag{7.15}$$

$$z_V(x) = 1 + \frac{(x^2 - x^{-2}) - 4(x - x^{-1})}{(x^{-1/2} - x^{1/2})^4} \tag{7.16}$$

for real scalars, Majorana fermions, and vectors respectively and $x = e^{-\beta/R}$. We observe that z_S, z_F and $1 - z_V$ are T-reflection symmetric with $\gamma = \pi$.[3]

The single-trace and full multi-trace confining-phase partition functions of (nearly-free) large-N gauge theories with adjoint matter on $S^3_R \times S^1_\beta$ have T-reflection symmetry, as they depend only on the covariant functions $1 - z_V(x)$, $z_S(x)$, and $z_F(x)$.[118][119][120] For example, for a theory with n_f and n_s massless adjoint fermions and conformally-coupled scalars, the confining-phase thermal partition function of the gauge theory in the large-N limit can be written as

$$Z_G(\beta) = \prod_{n=1}^{\infty} \frac{1}{1 - z_V(x^n) - n_s z_S(x^n) + (-1)^n n_f z_F(x^n)}. \tag{7.17}$$

The values $n_s = 6$ and $n_F = 4$ correspond to $\mathcal{N} = 4$ Super Yang-Mills theory. T-reflection maps each single particle partition function, (7.14)-(7.16), and hence each factor in the product (7.17), into itself, up to a factor of (-1). To compute the value of γ for the transformation of Z_G we must deal with the formal expression

$$\prod_{n=1}^{\infty}(-1) = (-1)^{\sum_{n=1}^{\infty} 1} \equiv (-1)^{\zeta(0)} = e^{-i\pi/2} \tag{7.18}$$

[3]Note that while the free Maxwell theory is conformal only in $d = 4$ [121], but z_S and z_F can be computed in any $d > 2$ via the state-operator correspondence, and they transform with $\gamma = (d - 1)\pi$ under T-reflection.

where in the last step we used zeta function regularization to define the value of the infinite sum. Hence these gauge theories are T-reflection symmetric with $\gamma = \pi/2$.

We have checked that the free conformally-coupled massless scalar, massless fermion, and $U(1)$ gauge theories without matter on $S^3 \times S^1$ all have T-reflection symmetry, but only if the vacuum energy is set to the appropriate Casimir energy. Since Z_G describes a family of confining theories[120], one might have expected that the vacuum energy would be of order the confinement scale $1/R$, which in this case is also the order of a Casimir energy on S^3. However, from (7.17) one can see that Z_G has T-reflection symmetry only if its vacuum energy *vanishes*. We do not know the reason for this phenomenologically tantalizing result.

7.3.3 Superconformal Indices

Next we consider twisted partition functions on $\mathcal{M} \times S^1_\beta$ for supersymmetric field theories, where the compact manifold $\mathcal{M}$ and the boundary conditions on the circle S^1_β are chosen such that supersymmetry is preserved. This requires that fermions have periodic boundary conditions on S^1_β, and S^1_β can be interpreted as a spatial circle. Our discussion applies to $d = 4$, $\mathcal{N} = 1$ theories with arbitrary matter content (with the rank of the gauge group N kept finite), as long as the theory has a $U(1)$ R-symmetry.

The virtue of such generalized partition functions is that they are independent of the continuous coupling constants and can be evaluated exactly even for strongly interacting theories, by taking the free coupling limit. We again essentially have a set of decoupled oscillators. In the examples we have been able to check, these partition functions are also invariant under T-reflections.

For $\mathcal{M} = S^3$, we have the so-called superconformal index [122][123] defined for $d = 4$, $\mathcal{N} = 1$ gauge theories on $S^3 \times S^1$, which is defined to be

$$I = \mathrm{Tr}\left[(-1)^{\mathcal{F}} p^{\frac{E+j_2}{3}+j_1} q^{\frac{E+j_2}{3}-j_1} \prod_i u_i^{F_i}\right], \tag{7.19}$$

where E, j_1, j_2 are the Cartan generators of $U(1) \times SU(2) \times SU(2)$ isometry of $S^3 \times S^1$, F_i are the flavor Cartan generators of the theory, and p, q, u_i are the relevant fugacities. These fugacities depend exponentially on β, and T-reflection inverts them.

As tabulated in e.g. [122, 124], the single-particle indices for vector-multiplets z_V and chiral-multiplets z_S are

$$z_V(p,q) = 1 - \frac{(pq)^{1/2} + (pq)^{-1/2}}{(p^{1/2} - p^{-1/2})(q^{1/2} - q^{-1/2})}\,,$$
$$z_S(p,q,u) = \frac{(pq)^{1/2}u^{-1} + (pq)^{-1/2}u}{(p^{1/2} - p^{-1/2})(q^{1/2} - q^{-1/2})}\,. \tag{7.20}$$

Although T-reflections invert the fugacities, we see that the rational functions $(1 - z_V)$ and z_S are invariant.

As was the case for the free CFTs on $S^3_R \times S^1_\beta$ above, the full superconformal indices are functions of z_S and $1 - z_V$. From this one can show that the entire superconformal index, a quantity associated with *interacting* SUSY gauge theories, has a T-reflection symmetry with $\gamma = 0$.

We have also verified that a similar argument works for the more general indices on $S^1 \times S^3/Z_p$ from [125], which in turn implies a T-reflection symmetry for their dimensionally-reduced counterparts (e.g. three-dimensional index on $S^1 \times S^2$ [126]). For $S^1 \times S^3/Z_p$, T-reflection also flips the discrete holonomies along S^3/Z_p.

7.3.4 Minimal models

We next consider $d = 2$ minimal model CFTs, which give an infinite family of interacting but solvable field theories. These CFTs describe the critical behavior of a wide variety of models from statistical physics, many of which are experimentally realizable. A minimal model $\mathcal{M}(p,p')$ has a finite set of primary operators, characterized by two integers r and s, satisfying $1 \le r < p, 1 \le s < p'$. The characters of minimal models are given by [127][128]

$$\chi^{p,p'}_{(r,s)} = K^{(p,p')}_{r,s}(q) - K^{(p,p')}_{r,-s}(q)\,, \tag{7.21}$$
$$K^{(p,p')}_{r,\pm s}(q) = \frac{1}{\eta(q)} \sum_{n=-\infty}^{+\infty} q^{\frac{(2pp'\,n + (pr \mp p's))^2}{4pp'}}\,. \tag{7.22}$$

The sum in (7.22) takes the form of a theta function—$\vartheta_{00}(z,Q)$—which has an infinite product representation:

$$\eta(q) K^{(p,p')}_{r,\pm s}(q) = q^{\frac{(pr \mp p's))^2}{4pp'}} \sum_{n=-\infty}^{+\infty} (q^{2pp'})^{\frac{n^2}{2}} (q^{pr \mp p's})^n$$
$$= Q^{\frac{\alpha^2}{2}} \prod_{n=1}^{\infty} (1 - Q^n)(1 + Q^{n+\alpha-\frac{1}{2}})(1 + Q^{n-\alpha-\frac{1}{2}}) \tag{7.23}$$

where $Q = q^{2pp'}$, $\alpha = \frac{pr \mp p's}{2pp'}$, and we used Jacobi's triple-product identity for $\vartheta_{00}(Q^\alpha, Q)$ in the last line. $Q(q)$-inversion maps $\eta(q) K^{(p,p')}_{r,\pm s}(q)$, i.e. $Q^{\frac{\alpha^2}{2}} \vartheta_{00}(Q^\alpha, Q)$, into

$$
\begin{aligned}
Q^{-\frac{\alpha^2}{2}} \vartheta_{00}(Q^{-\alpha}, Q^{-1}) &= \frac{Q^{-\frac{\alpha^2}{2}} \vartheta_{00}(Q^{+\alpha}, Q^{+1})(-1)^{\zeta(0)}}{Q^{\zeta(-1)+\zeta(-1,\frac{1}{2}-\alpha)+\zeta(-1,\frac{1}{2}+\alpha)}} \\
&= -i\, Q^{+\frac{\alpha^2}{2}} \vartheta_{00}(Q^\alpha, Q)\,. \qquad (7.24)
\end{aligned}
$$

The T-reflection symmetry of the character $\chi^{p,p'}_{(r,s)}$ then follows immediately from Eqs. (7.7), (7.8), (7.23) and (7.24), and we find $\gamma = 0$. We have carried out similar checks for the characters of $\mathcal{N} = 1, 2$ super-minimal models. All these characters have T-reflection symmetry.

These examples can be run backwards to show that T-reflection invariance fixes the ground state energy to take the values which coincide with the values mandated by other physical principles, such as modular invariance.

7.3.5 Lattice Models

At least some exactly solved lattice models in two-dimensional statistical mechanics also have T-reflection symmetry. For instance, Onsager's exact solution [129] of the two-dimensional Ising model on the square lattice, with periodic boundary conditions and zero external field, is

$$
\begin{aligned}
\ln\left[Z(\beta)\right] &= N \ln(2\cosh 2\beta J) \\
&+ \frac{N}{\pi} \int_0^{\pi/2} dw \, \ln\left[\frac{1}{2}\left\{1 + (1 - K^2 \sin^2 w)^{1/2}\right\}\right], \\
K &= \frac{2\sinh(2\beta J)}{(\cosh(2\beta J))^2}, \qquad (7.25)
\end{aligned}
$$

where J is the nearest neighbor interaction, and N is the number of sites on the lattice ($N \gg 1$). This solution has the T-reflection symmetry for arbitrary T, even away from the critical point. (At the critical point the Ising model is described by the minimal model $\mathcal{M}(4, 3)$, for which we already verified T-reflection.) Note that this is an explicit example of an interacting non-supersymmetric and non-conformal many-body model with T-reflection symmetry. The symmetry would be broken if there is a shift of the ground state energy in (7.25).

As shown in Refs. [130][131][132], superconformal indices of $\mathcal{N} = 1$ quiver gauge theories can be identified with partition functions of two-dimensional exactly solvable statistical mechanics models. Exploiting the T-reflection symmetry of the $\mathcal{N} = 1$ four-dimensional superconformal indices explored above, one could show a wide variety of models in two-dimensional statistical mechanics also have

T-reflection symmetry.

7.4 Discussion

Several comments are now in order. It is important to keep in mind that T-reflection symmetry can only hold for a special class of theories; for arbitary E_n the partition function (7.2) cannot have any simple T-reflection transformation. For instance, a three level system with unevenly spaced energy levels will not be T-reflection symmetric. For interacting systems with an infinite number of energy levels it is not easy to check T-reflection, as we have already remarked in the introduction. One of the exceptions is the case where E_n is a quadratic polynomial in n, where we can we can relate the transformation of (7.2) to the inversion properties of theta functions. While many of our examples were free theories, we again emphasize that T-reflection symmetry is not necessarily spoiled by interactions, as is highlighted by the minimal model examples.

It is conceivable that in general T-reflection symmetry may require simultaneous transformations of other parameters of the system. As a simple example, consider adding an anharmonic interaction $\frac{\lambda}{4!}\phi^4$ to the harmonic oscillator. The first-order correction to the partition function is (e.g. Ref. [133])

$$Z(\beta,\lambda) = Z_{\mathrm{SHO}}(\beta)\left[1 - \frac{\beta\lambda(1+x)^2}{8(1-x)^2}\right] \,, \tag{7.26}$$

where $x = e^{-\beta}$ and Z_{SHO} was given earlier in (7.5) (we take $\Delta = 1/2$). One then finds that $Z \to -Z$ under $\beta \to -\beta, \lambda \to -\lambda$. Note that the expression above has the λ-corrected vacuum energy

$$\Delta(\lambda) = \frac{1}{2}\left(1 + \frac{\lambda}{4}\right). \tag{7.27}$$

Just as in the non-interacting theory, one is not allowed to shift the vacuum energy from this particular value without ruining the T-reflection 'symmetry'. Similar conclusions seem to hold in higher orders of perturbation theory. Nonetheless, it is uncertain whether conclusions based on finite-order perturbation theory carry any real information in this context: T-reflection covariance depends crucially on the structure of the entire spectrum. Higher-lying states in the spectrum depend more sensitively on the anharmonicities. So it is an open question whether perturbative studies of deformations of exactly-solvable systems will have definitive lessons for the survival of T-reflection symmetry in more general settings.

All of our examples and remarks raise several obvious interrelated questions: what is the origin of the T-reflection symmetry? How general is the symmetry? What are the broader implications of T-reflection symmetry? The answers to these questions are currently unknown. It is conceivable that T-reflection symmetry is a previously unappreciated consequence of known symmetries, such as e.g. time-reversal symmetry, or it might be something entirely new. All we can say for now is that T-reflection symmetry is not a consequence of conformal invariance, nor of supersymmetry, since we have given examples of T-reflection symmetric systems which are neither conformal nor supersymmetric.

In some examples the reflection symmetry has a simple group-theoretical explanation. As an example, consider the character of the spin j-representation of $SU(2)$:

$$\mathrm{Tr}\, q^{J_3} = \frac{q^{2j+1} - q^{-2j-1}}{q - q^{-1}}\,, \tag{7.28}$$

This is invariant under $q \to q^{-1}$, which amounts to an exchange of the highest-weight state with the lowest-weight state. So here T-reflection is an involution (inner automorphism) of the $SU(2)$ algebra which exchanges the raising and lowering operators $J_\pm := J_1 \pm iJ_2$ via $J_3 \to -J_3\,, \quad J_\pm \to J_\mp$, which is simply a π rotation along the axis 1. Hence, for $SU(2)$, T-reflection symmetry is implied directly by the the algebraic structure itself. Similar involutions (called Cartan-Chevalley involutions) exist for arbitrary semisimple Lie algebras. One might then naturally wonder whether T-reflection symmetry for CFTs might have something to do with the fact that the Virasoro algebra

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}\delta_{m+n,0}\,, \tag{7.29}$$

has an outer automorphism $L_n \to -L_{-n}$. Such group-theory-based arguments do not seem to have a chance to explain T-reflection in general, however. The involution exchanges the highest energy state with the lowest energy state, but many of our examples correspond to infinite-dimensional representations without highest energy states.

7.5 Acknowledgements

We would like to thank A. Altuğ, N. Arkani-Hamed, J. L. Evans, Y.-t. Huang, Z. Komargodski, J. Maldacena, Y. Oshima, A. Polyakov, H. Verlinde, A. Zhiboedov for discussions. This work is supported by the U.S. Department of Energy under the grants DE-FG-88ER40388 (G. B.), DE-SC0011842 (A. C.).

Chapter 8

Casimir energy of confining large N gauge theories

8.1 Introduction

In typical quantum field theories (QFTs) with a mass gap $M_0 > 0$, the mass M of the heaviest particle species sets the natural size of the vacuum energy $V \sim M$. The Standard Model (SM) contains a variety of gapped sectors, and the electron contribution to the vacuum energy density $\mathcal{O}(m_e^4) \sim 6 \times 10^{-2}\,\mathrm{MeV}^4$ is already much larger than the value $\sim 1 \times 10^{-36}\,\mathrm{MeV}^4$ inferred from the accelerating expansion of the universe[134]. The apparent need to fine-tune V against M is the cosmological constant problem.

In gapped QFTs the only known mechanism which naturally gives $V = 0$ is linearly realized supersymmetry (SUSY). But if the SM is the low energy limit of a SUSY QFT, SUSY must be broken at some scale $\mu_{\mathrm{SUSY}} \gg m_e$ (see e.g. [135][136]), and the cosmological constant problem remains severe. This strongly motivates a search for other mechanisms that would force V to vanish.

If a QFT has a finite number of particle species it seems difficult to escape the conclusion that $V \sim M$, but what sets the scale of V if there are an infinite number of species with increasing masses[1]? This is the situation in weakly-coupled string theories and in confining large N gauge theories, which are believed to have a dual string description[142]. In this chapter we compute the vacuum energy of a variety of *non-supersymmetric* $SU(N)$ gauge theories at $N = \infty$, including pure

[1]See [137][138][139][140][141] for earlier discussions of this type of question.

Yang-Mills theory. The calculations are done using a compactification of spacetime to $S^3_R \times S^1_\beta$, where these theories develop an analytically tractable confining regime[120] if the S^3 radius R is much smaller than the strong scale $1/\Lambda$, and if the temperature $T = 1/\beta$ is below a critical value. In this regime V is simply the Casimir energy C of the theory on $S^3 \times \mathbb{R}$. It was recently observed[6] that temperature-reflection (T-reflection) symmetry predicts that the vacuum energy associated with the $N = \infty$ spectrum of these confining theories should *vanish*.

Our calculations confirm this prediction. Since the result holds in a variety of large N gauge theories, it seems unlikely to be an accident. It is possible that confining gauge theories have emergent symmetries in the large N limit which force V to vanish.

8.2 T-reflection

For QFTs on $S^3_R \times S^1_\beta$ the spectrum of single-particle excitations is discrete, and in our cases of interest the partition function can be written as

$$\begin{aligned} -\log Z(\beta) = -V_0\,\beta\mathcal{V} &+ \sum_{\pm,n=1}^{\infty}\left[\pm\frac{\beta}{2} d_n^{\pm}\omega_n^{\pm}\right] \\ &+ \sum_{\pm,n=1}^{\infty}\left[\pm d_n^{\pm}\log\left(1 \mp e^{-\beta\omega_n^{\pm}}\right)\right] \end{aligned} \tag{8.1}$$

where V_0 is the bare vacuum energy, $\mathcal{V}$ is the spatial volume, and $\omega_n^{\pm}, d_n^{\pm}$ are the energies and degeneracies of bosonic (+) and fermionic (-) states. We study theories where $\omega_n^{\pm}$ only depends on the scale R. The sum in the upper line is UV divergent and must be regulated and renormalized to obtain a physical expression. The renormalized contribution explicitly depends on R and is the Casimir energy. In [6] we noted that one can also formally define the quantity $Z(-\beta)$ by sending $\beta \to -\beta$ in (8.1):

$$\begin{aligned} -\log Z(-\beta) = V_0\,\beta\mathcal{V} + \log(-1)\sum_{n=1}^{\infty} d_n^{+} \qquad\qquad (8.2)\\ + \sum_{\pm,n=1}^{\infty}\left[\pm\frac{\beta}{2} d_n^{\pm}\omega_n^{\pm}\right] + \sum_{\pm,n=1}^{\infty}\left[\pm d_n^{\pm}\log\left(1 \mp e^{-\beta\omega_n^{\pm}}\right)\right] \end{aligned}$$

Of course, $Z(-\beta)$ also has UV divergences, and requires the same type of regularization and renormalization as $Z(\beta)$. With renormalized expressions for both $Z(\beta)$ and $Z(-\beta)$ in hand it can be shown that there is a T-reflection symmetry[6]

$$Z(\beta) = e^{i\gamma} Z(-\beta) \tag{8.3}$$

where $\gamma = \pi \, \text{Finite}[\sum_{n=1} d_n^+]$ [2], provided that the R-independent part of the vacuum energy from V_0 is set to zero. Hence (8.3) holds only if the renormalized vacuum energy V coincides with the Casimir energy $C = 1/2 \sum_{\pm,n} d_n^\pm \omega_n^\pm$. For instance, on $S_R^3 \times S_\beta^1$, Eq. (8.3) holds for a real conformally-coupled scalar field when $V = 1/(240R)$ and $\gamma = 0$, while for an Abelian vector field T-reflection holds with $V = 11/(120R)$ and $\gamma = \pi$.

8.2.1 Non-abelian gauge theories on $S_R^3 \times S_\beta^1$

We analyze $SU(N)$ gauge theories with n_F adjoint Majorana fermions and n_S real adjoint scalars on $S_R^3 \times S_\beta^1$. For moderate n_F, n_S, these theories are asymptotically free with a strong scale Λ, and are weakly coupled if $R\Lambda \ll 1$. Indeed, in the $\Lambda R \to 0$ limit where we will work, the 't Hooft coupling λ goes to 0, and these theories develop a conformal symmetry at the microscopic level. However, no matter how small λ becomes, the Gauss law constraint on the compact manifold S^3 only allows color-singlet operators to be part of the space of finite-energy states, and these operators must include one or more color traces.

As explained in detail in [120] (see also [118, 119]) in the large N limit such theories have at least two distinct phases. In particular, there is a low temperature confining phase, dominated by the dynamics of an infinite number of stable single-trace hadronic states, and a mass gap of order $1/R$. The confined phase has a free energy scaling as N^0 and unbroken center symmetry.

In this chapter, we focus on the weakly-coupled large N confining phase, since we wish to compute the vacuum energy of the theory on $S^3 \times \mathbb{R}$. The Casimir energy is dictated by the energies and degeneracies of the states of the theory, which are in turn encoded within the thermodynamic partition function, $Z(\beta) = \text{Tr}\, e^{-\beta H}$. We shall use the spectrum of states in the $N = \infty$ limit to compute the Casimir energy. Before proceeding to the vacuum energy computation, we review and expand on the remarks in [6] concerning the T-reflection properties of $Z(\beta)$ in $N = \infty$ confining gauge theories on $S^3 \times S^1$.

In large N confining phases, the physical excitations are created by single-trace operators which generate the physical single-particle states. Hence the thermodynamic partition function associated to the spectrum of excitations on $S^3 \times \mathbb{R}$ is given by (8.1) with the spectral data $\omega_n^\pm, d_n^\pm$ taken from

[2] There are branch cuts ambiguities in the definition of γ, and one can choose it to live in $[0, \pi]$.

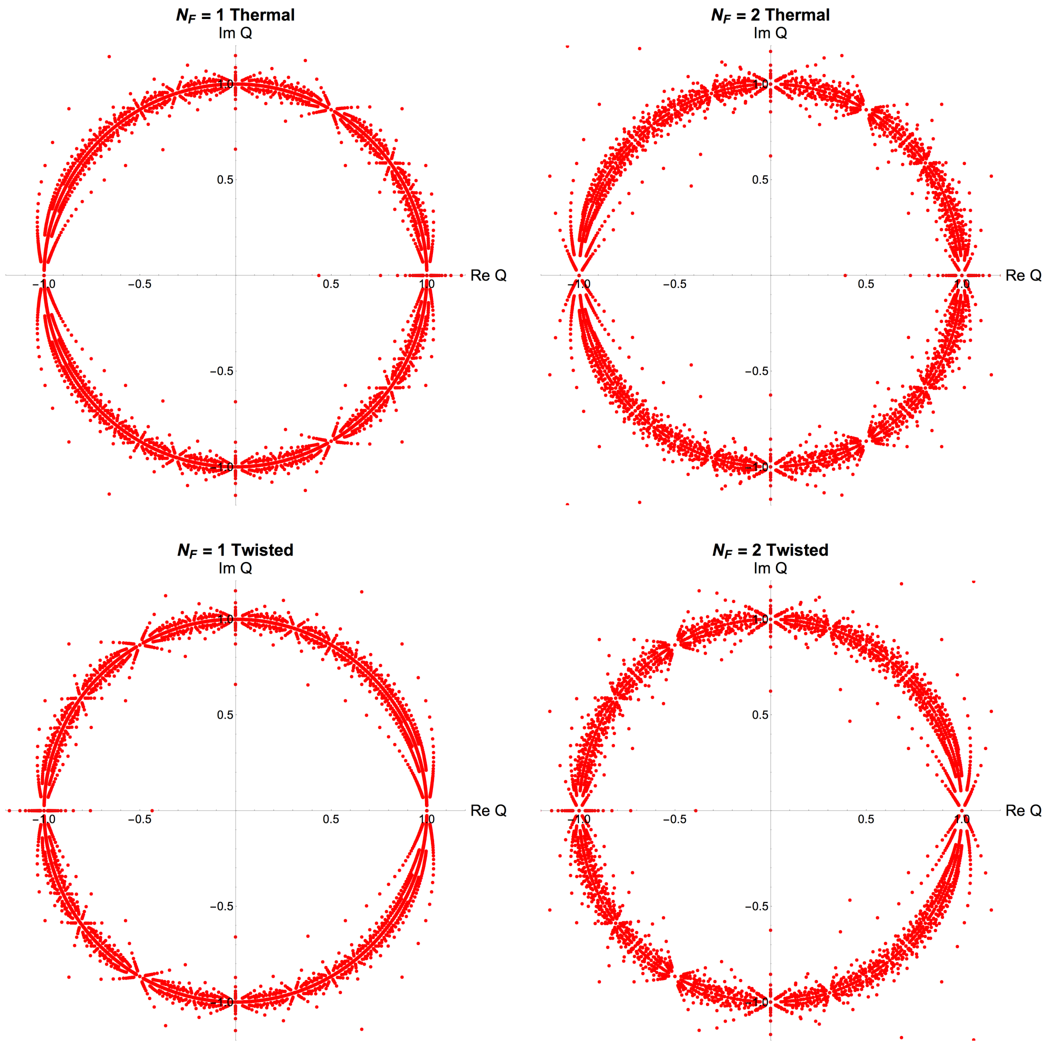


Figure 8.1: (Color Online.) Structure of singularities (red dots) coming from the first 45 terms in (9.26) in the large N confining-phase partition functions of gauge theories with adjoint matter on $S^3 \times S^1$, in the complex plane for $y = e^{-\beta/(2R)}$. The blue curve is an example of a path from $y = 0$ to $y = 1$ which does not pass through any singularities. **Left:** Yang Mills ($n_F = 0, n_S = 0$) theory. **Right:** Gauge theory with $n_F = 1, n_S = 2$.

the single-trace thermodynamic partition function[120]

$$-Z_{\mathrm{ST}}(\beta) = \sum_{k=1}^{\infty} \frac{\varphi(k)}{k} \log\left[1 - z_V(x^k) - n_S z_S(x^k) + (-1)^k n_F z_F(x^k)\right] =: \sum_{n=1}^{\infty} d_n y^n \tag{8.4}$$

where $\varphi(k)$ is the Euler totient function, $x = e^{-\beta/R}$, $y = x^{1/2}$, states with even/odd labels n are bosons/fermions, and

$$z_S(x) = \frac{x^2 + x}{(1-x)^3}, \; z_F(x) = \frac{4x^{3/2}}{(1-x)^3}, \; z_V(x) = \frac{6x^2 - 2x^3}{(1-x)^3}$$

are the so-called single-letter partition functions for respectively the conformally-coupled real scalar, Majorana fermion and Maxwell vector fields on S^3.

To relate this to (8.1), which includes contributions from multi-particle states, recall that for bosonic systems with integer-spaced levels we can write

$$\begin{aligned} -\log Z^{(0)}(\beta) &= \sum_{n=1}^{\infty} d_n \log(1-x^n) = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \frac{d_n}{k} x^{kn} \\ &= \sum_{k=1}^{\infty} \frac{Z_{\mathrm{SP}}(x^k)}{k} \end{aligned} \tag{8.5}$$

where $Z_{\mathrm{SP}}(\beta)$ is the single-particle partition function, with a similar final expression for a fermionic system. $Z^{(0)}(\beta)$ is only a part of the expression (8.1) for $Z(\beta)$, since it leaves out the Casimir vacuum energy. Hence unless the Casimir energy happens to be zero, $Z^{(0)}(\beta)$ will not enjoy T-reflection symmetry. Indeed, for most QFTs, $Z^{(0)}(\beta)$ is not T-reflection symmetric, and the Casimir energy must be included in $Z(\beta)$ to satisfy T-reflection, as can be checked for a free scalar field theory on $S^3_R \times S^1_\beta$.

Nevertheless, consider the $N=\infty$ confined-phase gauge theory partition function *without* the vacuum energy contribution[120]:

$$\begin{aligned} Z_G(\beta) &:= \exp\left[-\sum_{k=1}^{\infty} \frac{Z_{\mathrm{ST}}(x^k)}{k}\right] \\ &= \prod_{n=1}^{\infty} \frac{1}{1 - z_V(x^k) - n_S z_S(x^k) + (-1)^k n_F z_F(x^k)} \end{aligned} \tag{8.6}$$

Since $z_S(1/x) = -z_S(x)$, $z_F(1/x) = -z_F(x)$, and $1 - z_V(1/x) = -[1 - z_V(x)]$, we see that

$$Z_G(\beta) = e^{i\pi/2} Z_G(-\beta) \tag{8.7}$$

with the prefactor obtained from a zeta-function regularization of $(-1)^{\sum_{n=1}^{\infty} 1}$. So $Z_G(\beta)$ enjoys T-reflection symmetry. This is consistent with the general argument for T-reflection symmetry after (8.1) only if the renormalized Casimir vacuum energy of the $N=\infty$ theory *vanishes*.

8.2.2 Vacuum energy

To check the T-reflection prediction we calculate the Casimir vacuum energy C

$$C = \frac{1}{2}\sum_{n=1}^{\infty} d_n \omega_n \tag{8.8}$$

with $R\omega_n = n/2$ and d_n are drawn from (9.26). The sum is divergent, and must be regularized and renormalized to find the physical value of C. In many QFTs the simplest way to do this is to observe that C is encoded in the behavior of the physical single particle partition function[3], which for us is $Z_{\mathrm{ST}}(y)$, through

$$C[y] = \left[\frac{1}{4R} y \frac{d}{dy} Z_{\mathrm{ST}}(y^2)\right] = \frac{1}{2}\sum_{n=1}^{\infty} d_n \omega_n y^n. \tag{8.9}$$

Then C would normally be given by the finite part of $C[y \to 1]$ in the simple class of theories we work with, which have no microscopic mass terms. This amounts to defining C via an especially natural analytic continuation, and also resembles a heat-kernel regularization, since it effectively introduces the damping factor $e^{-\omega_n/\mu}$, with $\mu = 1/\beta$ playing the role of the UV cutoff. If we were dealing with a system where $d_n \to q\, n^p$ once $n \gg 1$ for some fixed $p, q \in \mathbb{R}^+$, then $C[y]$ would be well-defined for any $y \in [0, 1)$, and we would expect to find

$$C[y \to 1] = R^3\mu^4 + R\mu^2 + C + \mathcal{O}(\mu^1) \tag{8.10}$$

and the leading power of μ is tied to the spacetime dimension $d = 4$. Here, the thermodynamic degeneracy factors d_n from (9.26) are associated with confining large N gauge theories, and it is known that d_n grows *exponentially* with n, $d_n \sim p\, n^q\, h^n, n \gg 1$ with $p, q, h \in \mathbb{R}^+$ and $h > 1$. This is the famous Hagedorn scaling of the density of states. Consequently, if we keep $\mu \in \mathbb{R}^+$, $Z_{\mathrm{ST}}(\mu)$ is only well-defined for $\mu < \mu_H$. Physically, if the temperature is increased past T_H there is a Hagedorn instability, and a consequent phase transition to a deconfined phase. So at first glance it is not clear how to use (8.9) to compute C for confining large N theories.

To circumnavigate this roadblock, note that we do not have to take the $y \to 1$ limit of Z_{ST} along the real axis. We can approach $y = 1$ along any smooth path in the complex plane which does not

[3]A related but different prescription gives the a-anomaly coefficient[143][144].

go through any singularities. The singularities of $Z_{\rm ST}[y]$ are set by the roots of

$$p[y] = 1 - z_V(y^2) \pm n_F z_F(y^2) - n_S z_S(y^2) \tag{8.11}$$

If $p[y]$ has a root $y_H \in [0,1]$, then the logarithms in (9.26) (which depend on $p[y^k]$) become singular at $y = y_H, y_H^{1/2}, y_H^{1/3} \cdots$, and (9.26) ceases to be well-defined for $y \geq y_H$. Such roots are present for any integer $n_F, n_S \geq 0$, which is the origin of the Hagedorn instability. Figure 8.1 shows the location of the singularities of the Yang-Mills (left) and $N_f = 1, N_s = 2$ (right) single-trace partition functions as red dots, with the blue curve illustrating an example of one of the many approach trajectories to $y = 1$ along which there are no singularities.

Armed with this observation, we can evaluate C numerically or analytically. Let us begin with the analytical approach. The first step to isolate the part that diverges as $y \to 1$ from the rest in $y\, dZ_{ST}/dy$ in (8.9);

$$\begin{aligned} &y \frac{\partial}{\partial y} \log\left[1 - z_V(y^{2m}) + n_F(-1)^m z_F(y^{2m}) - n_S z_S(y^{2m}))\right] \\ &= \frac{2my^{2m}\left(3y^{4m} - 2(n_S+3)y^{2m} + 6n_F(-y)^m - n_S - 3\right)}{y^{6m} - (3+n_S)y^{4m} + 4n_F(-y)^{3m} - (3+n_S)y^{2m} + 1} \\ &\quad + \frac{6my^{2m}}{1-y^{2m}} = 3m + \frac{6my^{2m}}{1-y^{2m}}\,, \end{aligned} \tag{8.12}$$

where in the last step we substituted $y = 1$ in the finite term. This substitution should be understood as a limit in the complex plane that avoids any singularities along its path, as described above. By using Eqs. (9.26), (8.9), and (9.67) we obtain the formally divergent expression

$$C = -\frac{3}{4R}\left(\sum_{m=1}^{\infty} \varphi(m) + 2\lim_{\beta\to 0}\sum_{m=1}^{\infty} \frac{\varphi(m)\, y^{2m}}{1-y^{2m}}\right). \tag{8.13}$$

After regulating the first term using a spectral zeta function via the identity $\sum_{m=1}^{\infty}\varphi(m)m^{-s} = \zeta(s-1)/\zeta(s)$, and regulating the second using the Lambert series, $\sum_{m=1}^{\infty}\varphi(m)q^m/(1-q^m) = q/(1-q)^2$, we obtain

$$C = -\frac{1}{4R}\left(\frac{3\zeta(-1)}{\zeta(0)} + \frac{6R^2}{\beta^2} - \frac{1}{2}\right) = -\frac{3R}{2\beta^2} \tag{8.14}$$

up to $\mathcal{O}(\beta^2)$. The divergent contribution is cancelled by a $\int d^4x\sqrt{g}\,\mathcal{R}$ counter-term, and the lack of a finite term in (8.14) means that the renormalized C is zero. A similar calculation gives $\gamma = -3\pi/2$. At first glance, splitting terms in (8.9) and regularizing them individually might seem worrisome, but since we have used a spectral zeta function and the cutoff functions depend only on the spectrum throughout, these manipulations are justified.

To compute C numerically we examine the $\beta \to 0$ limit of $Z_{\mathrm{ST}}[e^{-\beta e^{-i\alpha}}]$ with a cutoff $k_{\max}$ on the k sum. Increasing $k_{\max}$ allows probing Z_{ST} at lower β. The leading divergence in Z_{ST} as $\beta \to 0$ turns out to scale as $1/\beta$, rather than $1/\beta^3$ as one might have expected from (8.10)[4]. The coefficient of β in a small-β expansion of Z_{ST} is C. Extracting this coefficient using least-squares fits in Yang-Mills theory gives e.g. $CR = (-0.95 + 2.22i) \times 10^{-4}$ with $k_{\max} = 2.5 \times 10^5, \alpha = \pi/6$ from sampling Z_{ST} in the range $\beta \in [5 \times 10^{-5}, 6.5 \times 10^{-2}]$. C decreases as $k_{\max}$ is increased, which is consistent with the analytic result $C = 0$. We have also checked that the analytic results for C and γ (up to branch choice ambiguities) are reproduced numerically for theories with $N_f = 0, N_s \geq 0$. We have not succeeded in getting stable numerical results for C once $N_f \geq 1$, so for this subclass of theories our conclusions rely on our two analytic arguments.

8.3 Conclusions

The confining-phase Casimir vacuum energy in non-supersymmetric large N gauge theories with adjoint matter turns out to be zero. This result cannot be attributed to cancellations between bosons and fermions, since it holds even in Yang-Mills theory, which has a purely bosonic spectrum. Since we find a zero vacuum energy in a variety of examples, it is unlikely to be an accident. It appears that there is a mechanism other than SUSY that can make vacuum energies vanish, at least in a class of $N = \infty$ gauge theories, and consequently also in their string duals.

Obviously the most pressing task suggested by our results is to understand them in terms of some symmetry principle. This may involve some novel emergent large N symmetry of confined phases of gauge theories, or some previously unrecognized $N = \infty$ consequence of an already known symmetry, such as center symmetry. It will be valuable to gather further clues by generalizing the analysis, and to explicitly compute $1/N$ corrections to the vacuum energy. Depending on how broadly the results generalize, it is possible that they may find phenomenological applications. It is important to see whether the vacuum energy continues to vanish if additional scales are introduced into the problem, for instance by working with a squashed S^3, and to understand the consequences of including contributions from other matter field representations. Finally, we note that there may be some relations between our results and the recent observation that the $S^3 \times S^1$ Casimir energy vanishes in non-interacting conformal higher-spin theories[145].

[4] A similar result was discussed in [146].

8.4 Acknowledgements:

We thank O. Aharony, J. L. Evans, C. P. Herzog, K. Jensen, I. Klebanov, M. Ünsal, and H. Verlinde for discussions. This work is supported in part by the U. S. Department of Energy under the grants DE-FG-88ER40388 (G. B.) and DE-SC0011842 (A. C.).

Chapter 9

Fermionic symmetries in QCD[Adj]

9.1 Introduction

In this chapter we explore adjoint QCD, an $SU(N)$ gauge theory with N_f flavors of massless Majorana quarks in the adjoint representation of $SU(N)$. Working in a weakly coupled and analytically tractable regime, we show that for any $N_f \geq 1$ there are large cancellations between bosonic and fermionic contributions to the $(-1)^F$-twisted partition function at large N. The cancellations are so strong that when large N adjoint QCD is compactified on a spatial circle of size L, with periodic boundary conditions for the fermions, it has no Hagedorn instabilities and stays in a confined phase for *any* $L \sim N^0$, and enjoys large N volume independence for any $L \sim N^0$.

The weakly coupled regime used in our calculations opens up when the theory is compactified on $S^3 \times S^1$ and the S^3 radius is made small[118, 119, 120]. When the S^1 is large, the large N theory can be shown to be in a confined phase, with the physical spectrum consisting of weakly coupled 'hadron' states created by single-trace operators and an order N^0 free energy. If the S^1 circle is spatial, with periodic boundary conditions for the fermions, the Euclidean path integral computes the twisted partition function[147]

$$\tilde{Z}(L) = \mathrm{Tr}(-1)^F e^{-LH} = \int dE \tag{9.1}$$

where $\rho_{B,F}$ are the bosonic and fermionic densities of states and L is the circumference of the S^1. We verify that as a consequence of the Hagedorn phenomenon, both ρ_B and ρ_F grow exponentially in E. In principle ρ_B and ρ_F might be expected to be quite different from each other. Remarkably, we find that ρ_B and ρ_F have the same asymptotic behavior, with all exponentially-growing parts coinciding exactly for any $N_f \geq 1$. Such a relation between the bosonic and fermionic densities of states leads to the dramatic consequence that adjoint QCD on $S^3 \times S^1$ *does not* have a Hagedorn instability, and the theory stays in the confined phase for *any* spatial circle size $L \sim N^0$ for any $N_f \geq 1$. This is due to the fact that (9.1) involves $\rho_B - \rho_F$, in contrast to the thermal partition function, which involves $\rho_B + \rho_F$. The boson-fermion degeneracies lead to strong cancellations in (9.1), and keep $\tilde{Z}(L)$ a smooth function of L for any $L \sim N^0$. Our results provide physical insight into the result of [147], which found that adjoint QCD on $S^3 \times S^1$ enjoys large N volume independence for any L.

The observation of degeneracies between bosonic and fermionic spectra normally suggests that the theory has a fermionic symmetry. But at any finite N, adjoint QCD on $S^3 \times S^1$ is not supersymmetric. The S^3 curvature breaks the flat-space $\mathcal{N} = 1$ supersymmetry of the $N_f = 1$ theory, while if $N_f > 1$ the theory has $2(N^2 - 1)$ bosonic and $2N_f(N^2 - 1)$ fermionic degrees of freedom at the microscopic level, and hence cannot be supersymmetric in any conventional sense even in flat space. Since the degeneracies we observe appear in the large N limit, our results are consistent with the conjecture posed in [148] that adjoint QCD should have an *emergent* fermionic symmetry in the large N limit even away from $N_f = 1$ if the theory enjoys volume independence. Emergent fermionic symmetries in the large N limit of otherwise non-supersymmetric theories do not contradict the Coleman-Mandula and Haag-Lopuszanski-Sohnius theorems, since the S-matrix elements of physical states vanish in the large N limit.

The chapter is organized as follows. In Section 9.2 we review some relevant properties of adjoint QCD, and summarize the arguments of [148] concerning Hagedorn instabilities and large N volume independence which motivated our search for spectral degeneracies in adjoint QCD. In Section 9.3 we describe the calculation of the twisted and thermal partition functions for adjoint QCD in the large N limit on $S^3 \times S^1$, using the technology of [118, 119, 120]. Section 9.4 is the key part of the chapter, and describes the behavior of the twisted and thermal densities of states which are relevant for spatial and thermal compactifications respectively. Figure 9.1 gives a visual summary of our story. Thermally-compactified adjoint QCD has Hagedorn instabilities, as shown in Section 9.4.1, but there are no Hagedorn instabilities for spatial compactification as shown in Section 9.4.2. We compute the twisted Casimir energy in adjoint QCD at large N and show that it vanishes in Section 9.4.3, while Section 9.4.5 comments on the connections between our results and misaligned supersymmetry. Finally, in Section 9.5, we make some remarks on the relation of our findings to the underlying symmetries of adjoint QCD, and conclude in Section 9.6.

9.2 Properties of large N adjoint QCD

In this section we briefly review two properties of large N gauge theories — and in particular of adjoint QCD — which play a key role in the rest of our analysis. These properties are the presence of Hagedorn instabilities in generic confining large N gauge theories, and the phenomenon of large N volume independence, which is special to adjoint QCD. The tension between Hagedorn instabilities and volume independence motivate our study of adjoint QCD on $S^3 \times S^1$.

9.2.1 Hagedorn instability

Large N gauge theories with a confinement scale Λ_c are believed to have a density of states $\rho(E)$ with a Hagedorn scaling [149]

$$\rho(E \gg \Lambda_c) \to e^{\beta_H E}, \qquad \beta_H \sim \Lambda_c^{-1} \tag{9.2}$$

A heuristic argument for this relation is that large N theories have an infinite number of stable hadronic states, and highly-excited states can be thought of as excitations of confining strings, see e.g. [150]. Relativistic string theories famously have Hagedorn densities of states, motivating (9.2). A more rigorous argument in favor of (9.2) based directly on the known properties of large N gauge theories was recently given in [151, 152].

If such a theory is compactified on $M \times S^1_\beta$, where S^1_β is a thermal circle, then the associated partition function can be written as

$$Z(\beta) = \mathrm{Tr}\, e^{-\beta H} = \int dE\, [\rho_B(E) + \rho_F(E)]\, e^{-\beta E} \tag{9.3}$$

with $\rho_{B,F}$ being the bosonic and fermionic densities of states respectively. If $\rho_B + \rho_F = \rho$ satisfies (9.2), then the sum over states in $Z(\beta)$ will diverge for $\beta \leq \beta_H$. This is known as a Hagedorn instability. Consequently, it is believed that all confining large N theories undergo a deconfinement phase transition at some inverse temperature $\beta_d \geq \beta_H$.

9.2.2 Large N volume independence

Consider a confining gauge theory with one or more directions compactified on a spatial torus T with periodic boundary conditions for fermions, and suppose the theory is in the confining phase. In general, connected correlation functions of single-trace color singlet operators will depend on the volume of T, with the dependence taking the form $e^{-L\Lambda}$ where Λ is the mass gap and $L \sim N^0$ is the scale of the volume[1]. Large N volume independence is the statement that in the 't Hooft large N limit, the connected correlation functions of topologically trivial single-trace operators do not depend on L, provided center symmetry and translation invariance are not broken [153, 154, 155, 156, 157, 158][2]. Volume independence implies that the connected parts of $n \geq 1$-point correlation functions of single-trace topologically-trivial operators are L-independent up to $1/N$ corrections. For zero point-functions such as $\log Z$ (the free energy), volume independence forces their $\mathcal{O}(N^2)$ parts to be volume independent. Of course, in the confining phase, where center symmetry is unbroken and volume independence is valid, $\log Z$ is $\mathcal{O}(N^0)$. Hence the validity of volume independence for $L \in [L_{\min}, \infty)$ implies that a theory must not have any Hagedorn instabilities for $L \in [L_{\min}, \infty)$, since these would drive the appearance of an $\mathcal{O}(N^2)$ volume-dependent part in $\log Z$.

Recently, convincing numerical and analytic evidence[159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174] has appeared that adjoint QCD with massless quarks is special in the sense that, when compactified on $M \times S^1_L$, it enjoys large N volume independence for *any* circle size $L \sim N^0$[157] so long as the circle is a *spatial* one, with periodic boundary conditions for fermions. That is, in adjoint QCD, large N volume independence is believed to hold for $L \in (0, \infty)$ for any $N_f \in [1, 5.5)$. [3]

9.2.3 The tension

Volume independence for any L implies the absence of phase transitions as a function of L. As a result, one might worry that large N volume independence for any L is not consistent with the well-established existence of Hagedorn instabilities at $L_H \sim \Lambda_c^{-1}$ in confining theories. Indeed, in many theories there truly is a clash between volume independence and the Hagedorn instability, which is resolved by the failure of volume independence at $\beta = \beta_d$[156, 175]. From a modern perspective, this gives a simple heuristic explanation for the failure of the original large N volume independence

[1]The restriction to $L \sim N^0$ is important, since in general volume *dependence* is expected to set in once $L \sim N^{-1}$, with e.g. possible chiral phase transitions at $L \sim 1/(N\Lambda)$ where Λ is the strong scale. The restriction to toroidal compactifications is also important, since on e.g. $S^3_R \times S^1_L$ the physics depends on R even at large N, in contrast to what sometimes happens to the dependence on L.

[2]There is a simple heuristic picture behind the phenomenon of large N volume independence. The way a given hadron knows that it is a periodic box is to interact with the 'image' hadrons introduced by the boundary conditions on the walls. If we take an 't Hooft large N limit, with $N \to \infty$ with all physical scales fixed, then the interactions between hadrons become $1/N$ suppressed, and the finite volume effects must disappear at leading order in the $1/N$ expansion. So as long as a large N theory is in its confining phase, it will enjoy volume independence for toroidal compactifications.

[3]When $N_f < 5.5$, adjoint QCD is asymptotically-free and has a strong scale Λ as determined from the IR Landau pole in the one-loop beta function. For $N_f < 4$ adjoint QCD on $\mathbb{R}^4$ is believed to develop a mass gap of order Λ. If $5.5 > N_f \gtrsim 4$, it is believed that adjoint QCD on $\mathbb{R}^4$ flows to a conformal fixed point in IR, and for $N_f = 5$ this fixed point can be seen in the two-loop beta function, and occurs at weak coupling.

proposal of Eguchi and Kawai in the context of pure Yang-Mills theory[153, 154]. However, adjoint QCD does not necessarily suffer from this issue[148]. To see this, recall that the modern formulation of large N volume independence is a statement about the sensitivity of observables to the size of spatial circles[157]. The Euclidean path integral for a theory compactified on a spatial circle computes the twisted partition function, $\tilde{Z}(L)$, defined in (9.1); it does not compute the thermal partition function $Z(\beta)$. The twisted and thermal partition functions are sharply different in theories with bosonic and fermionic states of similar energies. This is the case in $SU(N)$ adjoint QCD with massless fermions. In contrast, in QCD with N_f fundamental fermions, with even N there are no fermionic states at all, while for odd N the only fermionic states are baryons, which become parametrically heavy in the large N limit. The general statement is that the twisted and thermal partition functions are qualitatively similar for $\beta \sim L \sim N^0$ for large N gauge theories with complex-representation fermions, but they are very different in theories with light adjoint fermions.

The relevance of $\tilde{Z}(L)$ rather than $Z(\beta)$ means that the tension between volume independence and Hagedorn instabilities would be relieved if the exponentially-growing parts of ρ_B and ρ_F were the same, leading to sufficient cancellations in (9.1) to avoid Hagedorn instabilities. Supersymmetry would of course be sufficient to drive such cancellations, since in flat space the twisted partition function of a supersymmetric QFT is the Witten index, which is trivially volume-independent.

However, adjoint QCD is not supersymmetric for generic N_f, so it is not a priori obvious why one should expect sufficient cancellations in the twisted partition function to avoid Hagedorn instabilities. In this chapter we show that the necessary cancellations do indeed happen in adjoint QCD on $S^3 \times S^1$ for any $N_f \geq 1$. Since our results involve degeneracies between the energies of an infinite number of bosonic and fermionic states, it appears to call for the presence of emergent fermionic symmetries in large N adjoint QCD.

9.2.4 Utility of $S^3 \times S^1$ compactifications

Both volume independence and Hagedorn instabilities are usually strong coupling phenomena, which makes their interplay difficult to explore analytically. In this chapter we discuss volume independence and Hagedorn instabilities in adjoint QCD on $S^3_R \times S^1_\beta$ and $S^3_R \times S^1_L$, using methods developed in [120, 118, 147]. The reason this setting is interesting is that if $N_f < 5.5$, then the 't Hooft coupling $\lambda(R) \to 0$ as $\Lambda R \to 0$, where Λ is the strong scale. Hence the theory becomes weakly coupled and analytically calculable for *any* L or β.[4] At the same time, the $\Lambda R \ll 1$ theory is confining with a

[4] Our results also apply if $N_f > 5.5$, when the theory becomes IR-free, with a Landau pole Λ for the coupling in the UV. In this regime we can maintain weak coupling by setting $R\Lambda \gg 1$.

mass gap of order $1/R$, with the realization of center symmetry serving as an order parameter for confinement. As we will verify using the techniques of [120, 118], the presence of a Hagedorn density of states in adjoint QCD can be shown by direct calculation so long as $\Lambda R \ll 1$. Consequently, the $R\Lambda \ll 1$ limit gives us a regime where Hagedorn phenomena, center symmetry realizations and large N volume independence can all be explored simultaneously at weak coupling.

The presence of S^3 curvature couplings explicitly breaks the flat-space supersymmetry of the $N_f = 1$ $SU(N)$ theory, while $N_f > 1$ adjoint QCD is not supersymmetric even in flat space. So one might worry that on $S^3_R \times S^1_L$, volume independence would be doomed both with $N_f = 1$ and $N_f > 1$. However, some time ago, it was shown by Ünsal[147] that in adjoint QCD center symmetry is always unbroken on $S^3_R \times S^1_L$ for any $N_f \geq 1$, and hence large N volume independence must hold for any $N_f \geq 1$.[5] We illuminate the physics of this result by explicitly showing that there are no Hagedorn instabilities any $N_f \geq 1$ for any $L \sim N^0$ in the spatially-compactified theory. On the other hand, we show that there *are* Hagedorn instabilities for thermal compactification with $\beta \sim 1/R$. The spatially-compactified theory with $N_f \geq 1$ avoids Hagedorn instabilities due to large cancellations between bosonic and fermionic densities of states, as was advocated on general grounds in [148].

Before diving into the analysis, we make a remark on the global symmetries of adjoint QCD. Since the N_f Majorana fermions are in a real representation of the gauge group, the theory has a classical $U(N_f)$ flavor symmetry. The overall $U(1) \subset U(N_f)$ is anomalous, and on $\mathbb{R}^3 \times S^1$ it is believed that $SU(N_f)$ is spontaneously broken to $SO(N_f)$ by a chiral condensate when the S^1 is large.[6] The situation is quite different on $S^3_R \times S^1_L$, since the chiral symmetry realization depends on $R\Lambda$. For small $R\Lambda$, where the theory is weakly coupled for any $L \sim N^0$, the $SU(N_f)$ chiral symmetry is not spontaneously broken, and the curvature couplings induce a chirally-symmetric mass gap for the fermions[147]. The small $R\Lambda$ regime is an example of a setting where confinement and chiral symmetry breaking are not entangled with each other. These remarks will be important in Section 9.5.

9.3 Large N partition functions on $S^3 \times S^1$

When $R\Lambda \ll 1$, large N adjoint QCD is a nearly free quantum theory with an infinite number of degrees of freedom. Since all of the fields in the theory transform in the adjoint of the gauge group, in the $\lambda \to 0$ limit, each one of these degrees of freedom can be represented by $N \times N$ matrix harmonic

[5] See also [176] for a discussion of the fate of volume independence in this setting when a quark mass is turned on.

[6] See e.g. [177, 178, 179, 180, 181, 182, 183] for studies of confinement and chiral symmetry breaking in adjoint QCD in the volume-*dependent* weakly coupled regime which opens up for spatial circle compactification if $NL\Lambda \ll 1$. See also [184] for a recent overview of some properties of adjoint QCD.

oscillators, which transform as color-adjoints. The frequency of each oscillator is of order $1/R$. On a compact space, the Gauss law constraint, which applies no matter how small $R\Lambda$ becomes, implies that the only states which can contribute to a partition function must be color singlets.[7] Hence all the matrix oscillators have to occur inside color traces, and a typical state looks something like

$$\mathrm{Tr}[B_{43}^{\dagger}B_{2}^{\dagger}B_{2}^{\dagger}B_{17}^{\dagger}F_{9}^{\dagger}]|0\rangle \tag{9.4}$$

where $B_i^{\dagger}, F_i^{\dagger}$ are bosonic and fermionic oscillator creation operators, respectively, with spin and flavor indices suppressed for simplicity.

We will confine our attention to the behavior of adjoint QCD in the 't Hooft large N limit. This means sending N to infinity while fixing (i) N_f, (ii) 't Hooft coupling $\lambda = g^2N$, (iii) S^3 radius R, and (iv) the circle sizes L or β. Thanks to Boltzmann suppression factors, the last condition means that the only states that can contribute significantly to the partition function have energies of order N^0. When $R\Lambda \ll 1$, the energy of a state created by an a single-trace operator is directly proportional to the number of oscillators entering the trace. Thus by working in the 't Hooft large N limit defined by the conditions (i)-(iv) we are justified in only considering states created by N^0 oscillators. This is a major simplification, because it means that the space of multi-trace states is the Fock space of single-trace states.[8]

Combinatorially, the partition function of a system is a generating function which counts the number of states of each energy. In the rest of this section, we review the technology[118, 119, 120] that lets one directly count the states in the large N limit provided that $R\Lambda \ll 1$. First, we recall how to count the independent B_i and F_i operators, taking into account gauge freedom and the equations of motion. Then we count the single-trace and multi-trace color-singlet states. All this is already known from [118, 119, 120], but we repeat it here to keep the presentation self-contained. At the end of the section we obtain exact expressions for the thermal and twisted partition function of adjoint QCD at large N in the weakly coupled small R limit.

[7] The heuristic reason for this is that if one tries to put a source for color charge on a three-sphere there is no place for the color-flux lines to end. In flat space, in contrast, the flux lines have the option of 'ending' at the boundary at infinity.

[8] If the number of oscillators entering a single-trace operator scales with N there are algebraic relations between the single-trace operator and linear combinations of multi-trace operators, making the state counting much more complicated. These relations can be thought of as representing interactions between hadrons, which are $1/N$ suppressed for light states but may be unsuppressed for heavy states, as is well known from studies of large N baryons[185]. These subtleties become important at finite N, and also become important if we consider non-'t Hooft large limits where we allow L to scale as $1/N$.

9.3.1 Single particle partition functions

Adjoint QCD has a gauge field A_μ and fermion fields $\psi_a, a = 1, \lambda \cdots, N_f$. To build up a single-trace state, one can put together states composed of (a) various combinations of derivatives acting on A_μ, as well as (b) various combinations of derivatives acting on ψ. It is convenient to define generating functions z_V and z_F which count the number of independent color-adjoint states of type (a) and type (b) respectively. Following tradition we will call z_V and z_F "single particle" partition functions, though we emphasize that they are not the generating functions for the *physical* single-particle states of a non-Abelian gauge theory. The state-operator correspondence maps the energies associated with these states, $E_{V,F}$, to their classical scaling dimensions, $\Delta_{E,F}$, as $E_{V,F} = \Delta_{F,V}/R$ on $S^3_R \times S^1_{L\,\mathrm{or}\,\beta}$ in the $R\Lambda \ll 1$ limit, and provides an easy way to calculate the single particle partition functions as

$$z_F(q) \quad = \sum\nolimits_{\Delta_F} d_{\Delta_F} q^{\Delta_F} \tag{9.5}$$

$$z_V(q) \quad = \sum\nolimits_{\Delta_V} d_{\Delta_V} q^{\Delta_V}. \tag{9.6}$$

Here $d_{\Delta_{F,V}}$ denotes the degeneracy of the operator with dimension $\Delta_{F,V}$ and $q = e^{-\beta/R}$ or $q = e^{-L/R}$ depending on whether we consider thermal or spatial compactification respectively. Explicitly counting the operators by taking into account the equations of motion and gauge constraints, one obtains [118, 120, 119]

$$\begin{aligned} z_F(q) &= \frac{4q^{\frac{3}{2}}}{(1-q)^3} \\ z_V(q) &= \frac{2q^3 - 6q^2}{(1-q)^3}. \end{aligned} \tag{9.7}$$

See subsection 9.3.1 for a review of the derivations of these functions. Notably, these single particle partition functions have simple properties under the T-reflection symmetry $\beta \to -\beta$ introduced in [6]:

$$\begin{aligned} z_F(1/q) &= -z_F(q) \\ 1 - z_V(1/q) &= -\big(1 - z_V(q)\big). \end{aligned} \tag{9.8}$$

These T-reflection properties are very useful for obtaining analytic expressions for the Hagedorn temperatures of the theory, as well as for being able to write the full partition functions in terms of elliptic functions.

Explicit derivation of single particle partition functions

Here, we pause to recap the details needed to derive the standard expressions[118, 119, 120] for the free single particle partition functions for scalar, fermion, and Maxwell fields[9] on $S^3 \times S^1$, which is included to make the chapter as self-contained as possible.

The idea of the derivation is to use the conformal symmetry of the $\lambda = 0$ theory to map a state with energy E on $S^3 \times \mathbb{R}$ to a local operator on $\mathbb{R}^4$ with dimension $\Delta = E$. With this state-operator mapping, the problem boils down to counting operators with a given dimension Δ. These operators are the conformal descendants $Y_{(n)}$ of a given primary field Y satisfying the condition

$$Y_{(n)} = \partial_{\alpha_1} \partial_{\alpha_2} \cdots \partial_{\alpha_n} Y \,. \tag{9.9}$$

For a primary $Y \equiv Y_{(0)}$ with dimension Δ_Y, the scaling dimension of the descendant $Y_{(n)}$ in (9.9) is $\Delta_n = \Delta_Y + n$. Then the single particle partition function associated with Y can be written as

$$z_Y(q) = \sum_{\Delta} d_\Delta q^\Delta = q^{\Delta_Y} \sum_{n=0}^{\infty} d_n q^n \tag{9.10}$$

where $q = e^{-\beta/R}$. We now need to compute d_n to determine $z_Y(q)$. In doing this, it is important that the contributions of operators that include the equation of motion, $\mathcal{D}Y = 0$, be subtracted from the partition function since

$$Y_{(n)}^{\mathrm{EOM}} = \partial_{\alpha_1} \partial_{\alpha_2} \cdots \partial_{\alpha_n} \big(\mathcal{D}Y\big) = 0 \,. \tag{9.11}$$

For conformally-coupled scalars and fermions, this is the only constraint that must be taken into account in computing the single-particle partition functions, while for Maxwell fields there are additional constraints from gauge invariance, which we discuss separately.

Taking the equation of motion subtraction is easy to do after observing that the degeneracy of the level-n descendant of $\mathcal{D}Y$ is identical to the degeneracy of the level-n descendants of Y with a shift in the dimension by the mass dimension of the operator $\mathcal{D}$ which defines the equation of motion, $[\mathcal{D}]$. Or, in short, $\Delta(Y_{(n)}^{\mathrm{EOM}}) = [\mathcal{D}] + \Delta_Y + n$. Then, the single particle partition function becomes

$$z_Y(q) = q^{\Delta_Y}(1 - q^{[\mathcal{D}]}) \sum_{n=0}^{\infty} \hat{d}_n q^n \tag{9.12}$$

[9] There are some typos in the vector partition function in [118].

where $\hat{d}_n$ counts the number of different operators of the form (9.9) without any restriction. The number of different combinations of $\partial_{\alpha_1} \ldots \partial_{\alpha_n}$ is $\frac{(n+3)!}{n!\,3!}$.[10] Labeling the number of internal degrees of freedom of Y as $\mathcal{N}_Y$ we obtain

$$\hat{d}_n = \mathcal{N}_Y \frac{(n+3)!}{n!\,3!}\,. \tag{9.13}$$

Consequently we arrive at the result

$$z_Y(q) = \mathcal{N}_Y \frac{q^{\Delta_Y}(1 - q^{[\mathcal{D}]})}{(1-q)^4} \tag{9.14}$$

This expression holds for fermions and scalars. Specializing to a conformally-coupled free real scalar ϕ, we have $\Delta_\phi = 1$, $\mathcal{N}_\phi = 1$, and the operator defining the equation of motion is the Laplacian with $[\nabla^2] = 2$. Hence

$$z_\phi(q) = \frac{q + q^2}{(1-q)^3}\,. \tag{9.15}$$

For a free Majorana fermion, we set $\Delta_\psi = (d-1)/2 = 3/2$, and use $\mathcal{N}_\psi = 2^{d/2} = 4$. With $[\not{\mathcal{D}}] = 1$ for the Dirac operator, we obtain

$$z_\psi(q) = \frac{4\,q^{3/2}}{(1-q)^3}\,. \tag{9.16}$$

For a Maxwell gauge field, in addition to the constraint that follows from equation of motion, an additional constraint from gauge fixing has to be imposed on the operators. Let us again start with the most general descendant of the gauge field A_μ which has dimension $n+1$,

$$\partial_{\alpha_1} \cdots \partial_{\alpha_n} A_\mu\,. \tag{9.17}$$

There are $4\frac{(n+3)!}{n!\,3!}$ such operators, where $\mathcal{N}_{A_\mu} = 4$ since there are 4 components of the gauge field. We now fix the gauge and project out the non-gauge-invariant operators. It is convenient to work in the so-called "radial gauge" where

$$A_{\alpha_1} = 0\,,\; \partial_{\alpha_1} A_{\alpha_2} + \partial_{\alpha_2} A_{\alpha_1} = 0\,,\; \cdots\,,\quad \sum_{\text{permutations}} \partial_{\alpha_1} \cdots \partial_{\alpha_n} A_{\alpha_{n+1}} = 0\,. \tag{9.18}$$

[10] In d dimensions one gets $\frac{(n+d-1)!}{n!\,(d-1)!}$.

It is easy to see that these constraints project-out all non-invariant states, at levels $n = 0$ and $n = 1$. For $n = 0$, A_μ is not invariant, and should be projected out. For $n = 1$, there would naively be 16 descendants of A_μ. But the only single-derivative gauge-invariant object is the field-strength tensor, $F_{\mu\nu}$. Subtracting the symmetric combination of derivatives and vector indices in (9.18) from those appearing in (9.17) leaves only the antisymmetric combination, $F_{\mu\nu}$.

The number of symmetric combinations given in Eq. (9.18) with dimension $n + 1$ is simply $\frac{(n+4)!}{(n+1)!\,3!}$. Therefore, the off-shell vector partition function is

$$z_V^{\mathrm{off-shell}}(q) = \sum_{n=0}^{\infty} \left(4\frac{(n+3)!}{n!\,3!} - \frac{(n+4)!}{(n+1)!\,3!} \right) q^{n+1} = \frac{4q-1}{(1-q)^4} + 1\,. \tag{9.19}$$

We still have to project out the operators nullified by the equation of motion, $\partial_\mu F_{\mu\nu} = 0$, from Eq. (9.19). This procedure can be carried on in two steps. First, we identify the family of gauge fixed descendants that are nullified by the equation of motion:

$$\partial_\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0\,,\ \cdots\,,\ \partial_{\alpha_1}\partial_{\alpha_2}\cdots\partial_{\alpha_n}\partial_\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0\,,\ \cdots\,. \tag{9.20}$$

The number of such operators with dimension $n + 1$ is $4\frac{(n+1)!}{(n-2)!\,3!}$, which follows from counting the number of symmetric combinations of $n - 2$ derivatives and multiplying it by the number of components of A_μ. However, not all of the constraints in Eq. (9.20) are independent. Because $F_{\mu\nu}$ is antisymmetric in its two indices, any symmetric contraction one of the extra derivatives hitting the equation of motion in a descendant will identically vanish *independently* of the equation of motion, i.e.

$$\partial_\mu\partial_\nu(F_{\mu\nu}) = 0\,,\ \cdots\,,\ \partial_{\alpha_1}\partial_{\alpha_2}\cdots\partial_{\alpha_n}\partial_\mu\partial_\nu(F_{\mu\nu}) = 0\,.\ \cdots\,. \tag{9.21}$$

The second step is to *add* these terms back to correct for the double counting. The number of these descendants at level $n+1$ is $\frac{n!}{(n-3)!\,3!}$ which is the number of symmetric combinations of $n-3$ derivatives that hit $\partial_\mu\partial_\nu F_{\mu\nu}$. We then find that

$$z_V^{\mathrm{EOM}}(q) = \sum_{n=1}^{\infty} \left(4\frac{(n+1)!}{(n-2)!\,3!} - \frac{n!}{(n-3)!\,3!} \right) q^{n+1} = \frac{(4-q)q^3}{(1-q)^4}\,. \tag{9.22}$$

Putting everything together, the vector single particle partition function is obtained as

$$z_V(q) = z_V^{\text{off-shell}}(q) - z_V^{\text{EOM}}(q) = \sum_{n=1}^{\infty} 2n(n+2)q^{n+1} = \frac{2(3-q)q^2}{(1-q)^3} \tag{9.23}$$

9.3.2 Twisted and thermal partition functions of adjoint QCD

We now write down the twisted and thermal partition functions. To get some intuition on the physics, note that at large N we expect single-trace states to make the dominant contribution in the confined phase. A rough estimate of the contribution to the partition function from e.g. the gauge fields is

$$Z_{\text{ST, naive}} = \sum_{k=1}^{\infty} \frac{1}{k} \left[z_V(q)\right]^k = -\log[1 - z_V(q)] \tag{9.24}$$

This naive estimate counts single-trace operators made with k oscillators with a factor of $1/k$ to account for the cyclicity of the trace. The counting entering this estimate does not correctly deal with the combinatorics of repetitions of oscillators inside a single-trace, and multi-particle contributions are neglected. Both of these omissions lead to an undercounting of the states. Nevertheless, the naive estimate above manages to capture the leading asymptotics of the state degeneracies, which control e.g. the Hagedorn temperature, so it is useful to keep it in mind in what follows.

As shown in [118, 119, 120] the proper way to count the single-trace states with the correct weight for repetitions involves the use of Polya theory. The result is

$$\begin{aligned} Z_{\text{ST}}[q] &= -\sum_{m=1}^{\infty} \frac{\varphi(m)}{m} \log\left[1 - z_V(q^m) + (-1)^m N_f z_F(q^m)\right], & (9.25) \\ \tilde{Z}_{\text{ST}}[q] &= -\sum_{m=1}^{\infty} \frac{\varphi(m)}{m} \log\left[1 - z_V(q^m) + N_f z_F(q^m)\right]. & (9.26) \end{aligned}$$

Here, $\varphi(m)$, the Euler totient function, is the number of positive integers less than or equal to, and relatively prime to m. In the 't Hooft large N limit, the full confining-phase partition function can be obtained from the one above by including contributions from states involving an arbitrary number of particles. The full large N partition function can be written as[120] [11]

$$\log Z[q] = \sum_{k=1}^{\infty} \frac{Z_{\text{ST}}[q^k]}{k}. \tag{9.27}$$

[11] This construction, and its generalizations to finite N, is sometimes referred to as the 'plethystic exponential', popularized in the physics literature in [186, 187].

Euler's formula, $\sum_{k|n} \varphi(k) = n$, then implies

$$\log Z[q] = -\sum_{k=1}^{\infty} \log\left(1 - z_V(q^k) + (-1)^k N_f z_F(q^k)\right) \tag{9.28}$$

$$\log \tilde{Z}[q] = -\sum_{k=1}^{\infty} \log\left(1 - z_V(q^k) + N_f z_F(q^k)\right) \tag{9.29}$$

Note that these expressions are *only* correct at large N. At finite N (or in non-'t Hooft large N limits) there are relations between e.g. single-traces with $\gtrsim N$ oscillators and multi-trace states, and such relations are ignored in the derivation leading to the above result.

Before giving more explicit expressions for the partition functions, we make an important observation regarding the fermionic contributions to the single-trace and full partition functions. Due to the $q^{3/2}$ term in the fermionic single particle partition function, the fermions contribute to the expansions of the single-trace and full partition functions as half integer powers of q. Furthermore from Eqs. (9.7), (9.28) and (9.29) we see that going from the thermal to the twisted compactification amounts to flipping the sign of the coefficients of the half integer powers of q, so that

$$Z = \sum_{n=0}^{\infty} c_n q^n + \sum_{n=0}^{\infty} c_{n+1/2}\, q^{n+1/2} \tag{9.30}$$

$$\tilde{Z} = \sum_{n=0}^{\infty} c_n q^n - \sum_{n=0}^{\infty} c_{n+1/2}\, q^{n+1/2}. \tag{9.31}$$

So as expected, the difference between the twisted and the thermal partition functions is that all the fermionic degeneracy factors (i.e. coefficients of the half integer powers of q) enter with a negative sign to the twisted partition function. It is convenient to make the substitution $Q \equiv q^{1/2}$, so that the partition functions are power series expansion in Q with the even and odd powers of corresponding to bosons and fermions, respectively.

We now give give the expressions for the full partition functions in a more useful form. With the explicit single particle partition functions in Eq. (9.7), the large N pure YM partition function is

$$Z_{YM}(q) = \tilde{Z}_{YM}(q) = \prod_{k=1}^{\infty} \frac{(1-q^k)^3}{(1+q^k)(c-q^k)(c^{-1}-q^k)} \tag{9.32}$$

where $c = 2 + \sqrt{3}$ [12]. For pure YM, there is no difference between twisted and thermal partition

[12] The constant $c = 2 + \sqrt{3}$ appearing in the pure YM expression is a solution of (9.35) for the variable $q = Q^2$ with $N_f = 0$, along with -1 and $1/c$.

functions by definition, since there are no fermionic states. Defining

$$1 - z_V(Q^2) - N_F z_F(Q^2) = \frac{Q^6 - 3Q^4 - 4N_f Q^3 - 3Q^2 + 1}{(1-Q^2)^3} =: \frac{P(Q)}{(1-Q^2)^3} \tag{9.33}$$

with N_f massless adjoint fermions, the thermal partition function is

$$Z_{\mathrm{QCD[Adj]}}(Q) = \prod_{k=1}^{\infty} \frac{(1-Q^{2k})^3}{\prod_{i=1}^{6}(r_i + (-Q)^k)}, \tag{9.34}$$

where $Q = q^{1/2} = e^{-\beta/2R}$ and r_i with $i = 1, 2, \lambda \cdots, 6$ are the six solutions of the equation

$$P(Q) = Q^6 - 3Q^4 - 4N_f Q^3 - 3Q^2 + 1 = 0 \tag{9.35}$$

Note that, due to the $Q \to 1/Q$ T-reflection symmetry of the equation (9.35), the roots of $P(Q)$ come in reciprocal pairs. Organizing the roots as $r_{4,5,6} \equiv 1/r_{1,2,3}$, we obtain

$$Z_{\mathrm{QCD[Adj]}}(Q) = \prod_{k=1}^{\infty}\prod_{i=1}^{3} \frac{(1-Q^{2k})}{(1 + r_i(-Q)^k)\left(1 + r_i^{-1}(-Q)^k\right)} \tag{9.36}$$

The exact expressions for the roots r_i are given in subsection 9.3.4.

As discussed above, the twisted partition function can be obtained by taking $Q \to -Q$ in the thermal partition function, and it is given as

$$\tilde{Z}_{\mathrm{QCD[Adj]}}(Q) = \prod_{k=1}^{\infty}\prod_{i=1}^{3} \frac{(1-Q^{2k})}{(1 + r_i Q^k)(1 + r_i^{-1} Q^k)} \tag{9.37}$$

For completeness, note that the twisted partition function can also be written in terms of elliptic functions as

$$\tilde{Z}_{\mathrm{QCD[Adj]}}(L) \quad = \quad \eta^3\left(\frac{iL}{4\pi R}\right)\eta^3\left(\frac{iL}{2\pi R}\right)\prod_{i=1}^{3}\left[\frac{r_i^{1/2} + r_i^{-1/2}}{\vartheta_2\left(\nu_i | e^{-\frac{L}{4R}}\right)}\right]. \tag{9.38}$$

where $e^{2i\nu_i} \equiv r_i$, and the derivation is given in more detail in the following sub-section 9.3.3. Here $\eta(\tau) = e^{\frac{i\pi\tau}{12}} \prod_{n=1}^{\infty}(1 - e^{2i\pi\tau n})$ is the Dedekind eta function and

$$\vartheta_2(u|e^{i\pi\tau}) = \sum_{n=-\infty}^{\infty} e^{i(n+1/2)^2\,\pi\tau} e^{(2n+1)iu}\,, \tag{9.39}$$

with $Q = e^{-\frac{L}{2R}} =: e^{2i\pi\tau}$.

9.3.3 The representation of the twisted partition function in terms of elliptic functions

The T- reflection symmetry of the twisted partition function allows one to express it in terms of elliptic functions. To see this, let us start with the infinite product form given in Eq. (9.37). For such a representation, it is convenient to use the variables $\xi \equiv Q^{1/2} = e^{-\frac{L}{4R}} \equiv e^{i\pi\tau}$, $r_i \equiv e^{2i\nu}$, where r_i and r_i^{-1} are the roots of $P(Q)$ given in (9.35). The denominator of the twisted partition function can be written as

$$\begin{aligned}
\prod_{i=1}^{3}\prod_{k=1}^{\infty}(1+r_i\xi^{2k})(1+r_i^{-1}\xi^{2k}) &= \left(\prod_{m=1}^{\infty}\frac{1}{(1-Q^m)^3}\right)\prod_{i=1}^{3}\prod_{k=1}^{\infty}\left[(1+2\cos(2\nu_i)\xi^{2k}+\xi^{4k})(1-\xi^{2k})\right] \\
&= \left(\frac{Q^{\frac{1}{8}}}{\eta^3(\tau)}\right)\prod_{j=1}^{3}\left[\frac{\vartheta_2(\nu_j|e^{i\pi\tau})}{2\cos(\nu_j)\xi^{1/4}}\right] = \frac{Q^{-\frac{1}{4}}}{\eta^3(\tau)}\prod_{j=1}^{3}\frac{\vartheta_2(\nu_j|e^{i\pi\tau})}{r_j^{1/2}+r_j^{-1/2}}\,, \qquad (9.40)
\end{aligned}$$

where we have used the Jacobi triple product to obtain the theta function. The numerator can also be expressed in terms of the Dedekind eta function,

$$\prod_{k=1}^{\infty}(1-Q^{2k})^3 = Q^{-1/4}\eta^3(2\tau)\,. \qquad (9.41)$$

Putting everything together, we obtain our final result

$$\tilde{Z}_{\mathrm{QCD[Adj]}} = \eta^3(2\tau)\eta^3(\tau)\prod_{j=1}^{3}\left[\frac{r_j^{1/2}+r_j^{-1/2}}{\vartheta_2(\nu_j|Q^{1/2})}\right]. \qquad (9.42)$$

Note that the above expression can be simplified further when $N_f = 1$. In fact, due to the double root $Q = -1$ for $N_f = 1$, the formula (9.42) should be used with care. Let us analyze this case explicitly. Using the expressions for the roots give in (9.51), we can write

$$\begin{aligned}
\tilde{Z}_{N_f=1} &= \prod_{m=1}^{\infty}(1-Q^{2m})^3\prod_{i=1}^{2}\prod_{k=1}^{\infty}\frac{1}{(1-Q^m)^2(1+r_iQ^k)(1+r_i^{-1}Q^k)} = \eta^3(2\tau)\prod_{j=1}^{2}\left[\frac{r_j^{1/2}+r_j^{-1/2}}{\vartheta_2(\nu_j|Q^{1/2})}\right] \\
&= \frac{\sqrt{6}\,\eta^3(2\tau)}{\vartheta_2\left(\frac{i}{2}\log\left(\frac{1}{2}-\frac{\sqrt[4]{3}}{\sqrt{2}}+\frac{\sqrt{3}}{2}\right)|Q^{1/2}\right)\vartheta_1\left(\frac{1}{2}\sin^{-1}\left(\frac{\sqrt[4]{3}}{\sqrt{2}}\right)|Q^{1/2}\right)}, \quad (N_f=1) \qquad (9.43)
\end{aligned}$$

where we used the identity $\vartheta_1(z|e^{i\pi\tau}) = -\vartheta_2(z+\frac{\pi}{2}|e^{i\pi\tau})$.

9.3.4 Analytic expressions for Hagedorn temperatures

The T-reflection symmetry of the single particle partition functions allows us to analytically determine the roots of $1 - z_V(Q) + N_f z_F(Q)$ for arbitrary N_f. Without T-reflection, this would be impossible, as these equations are sixth order polynomials. Here, we give these analytical expression for the roots, which encode the singularities of the thermal and twisted partition functions given in Eqs. (9.36) and (9.37). As discussed below, in Section 9.4.1, the closest root to the origin along the real axis controls the Hagedorn growth of the density of states. In the thermal compactification, this closest root, r^*, also controls the Hagedorn temperature, T_H, via the relation $T_H = -\frac{1}{2R \log r_*}$.

For the thermal compactification, the relevant polynomial whose roots encode the singularities for the thermal compactification, given in Eq. (9.35), is $P(Q) = Q^6 - 3Q^4 - 4N_f Q^3 - 3Q^2 + 1$. The $Q \leftrightarrow Q^{-1}$, T-reflection symmetry forces the roots to come in reciprocal pairs, which we label as $\{r_i, r_i^{-1}\}$ with $i = 1, 2, 3$. It is also useful to define

$$R_i = r_i + \frac{1}{r_i}, \quad i = 1, 2, 3\,. \tag{9.44}$$

Writing the equation for the roots as

$$0 = P(Q) = \prod_{i=1}^{3}(Q^2 - R_i Q - 1) \tag{9.45}$$

leads to the set of equations

$$\sum_{i=1}^{3} R_i = 0, \qquad \prod_{1 \le i < j \le 3} R_i R_j = -6, \qquad \prod_{i=1}^{3} R_i = 4N_f\,. \tag{9.46}$$

Solving Eqs. (9.46) simultaneously, for $N_f \geq 2$, we arrive at the expressions

$$\begin{aligned} r_1 &= \tfrac{\kappa^2+2-\sqrt{\kappa^4+4}}{2\kappa} \\ r_2 &= -\tfrac{1}{16\kappa^2}\left[\kappa^3 + 2\kappa - 2\sqrt{\eta} + \left((\kappa^3+2\kappa-2\sqrt{\eta})^2 - 16\kappa^4\right)^{1/2}\right] \\ r_3 &= -\tfrac{1}{16\kappa^2}\left[\kappa^3 + 2\kappa + 2\sqrt{\eta} - \left((\kappa^3+2\kappa+2\sqrt{\eta})^2 - 16\kappa^4\right)^{1/2}\right] \end{aligned} \tag{9.47}$$

where

$$\kappa \equiv \left(2N_f + 2\sqrt{N_f^2 - 2}\right)^{1/3} \tag{9.48}$$

$$\eta \equiv 3\left(\kappa^4 - N_f \kappa^3 - \kappa^2 + 2\right)\,. \tag{9.49}$$

Among these roots and their reciprocals, the one closest to origin along the real axis is r_1.

For $N_f = 1$, there is a further simplification. The polynomial $P(Q)$ can be factored as

$$P(Q) = (1+Q)^2(1 - 2Q - 2Q^3 + Q^4) \qquad (N_f = 1), \tag{9.50}$$

and has roots

$$\begin{aligned} r_1 &= \frac{1}{2} - \frac{\sqrt{2}\sqrt[4]{3}}{2} + \frac{\sqrt{3}}{2} \\ r_2 &= \frac{1}{2} - \frac{i\sqrt[4]{3}}{\sqrt{2}} - \frac{\sqrt{3}}{2} = -e^{i \sin^{-1}(3^{1/4}/\sqrt{2})} \qquad (N_f = 1) \\ r_3 &= -1 \end{aligned} \tag{9.51}$$

with their reciprocals. For spatial compactification, the leading singularity is $-r_1$ and it is on the negative real axis. The rest of them can be obtained by substituting $N_f \to -N_f$ in (9.47) and their reciprocals.

9.4 Instabilities and their disappearance

Equipped with the exact formulas for the partition functions, we now discuss Hagedorn instabilities. In this section we show that the bosonic and fermionic states have identical asymptotics for $N_f \geq 1$. As a consequence spatially-compactified adjoint QCD with $N_f \geq 1$ *does not* have a Hagedorn instability. In contrast, the thermal theory has a Hagedorn instability, as expected.

9.4.1 Thermal compactification and the Hagedorn instability

The Hagedorn instability shows up as a singularity in the partition function at $\beta = \beta_H$, where β_H is the first singularity encountered as β is lowered from infinity. The presence of the Hagedorn instability signals that the system goes through a phase transition at a temperature $T \leq T_H \equiv \beta_H^{-1}$. This phase transition is believed to be the deconfinement transition of the gauge theory. On $S^3 \times S^1$ it was first explored in [118, 120], and was discussed in the specific context of large N volume independence in [147].

The Hagedorn singularity arises when one of the roots r_i is in the unit interval $[0, 1)$ and we hit a pole in (9.36) as we vary β. As the circle size is decreased from $\beta = \infty$ (or $Q = 0$), the first singularity occurs when $Q = r_*$, where r_* is the root closest to the the origin on the unit interval. For the thermal compactification, we are guaranteed to have such a root for any $N_f \geq 0$, since

Number of flavors	$N_f = 0$	$N_f = 1$	$N_f = 2$	$N_f = 3$	$N_f = 4$	$N_f = 5$
RT_H	0.759	0.601	0.532	0.490	0.461	0.440

Table 9.1: Hagedorn temperatures (rounded to three digits) for the large N limit of on $S^3_R \times S^1_\beta$ with N_f massless fermion flavors in the limit $R\Lambda \to 0$ with anti-periodic boundary conditions for fermions, so that S^1 is a *thermal* circle.

$P(0) = 1$ and $P(1) = -4(1 + N_f)$ so that there is at least one root $r_* \in [0, 1)$. Furthermore the first singularity of (9.36), r_*, is determined solely by the $k = 1$ factor in the infinite product since for $k > 1$ the singularity is at $(r_*)^{1/k} > r_*$. The Hagedorn temperature is thus

$$\beta_H = -2R \log r_*, \tag{9.52}$$

and the asymptotic behavior of the thermal density of states is

$$\rho(E) \sim \left(\frac{1}{r_*}\right)^{E/R}. \tag{9.53}$$

This asymptotic behavior follows from the fact that the coefficient of a given term, say Q^n, in (9.36) is generated by an finite product of geometric series with $k = 1, \ldots, n$ and is of the form

$$\rho_n = \sum_{\{-n \le k_{1,2,3} \le n\}} c_{k_1,k_2,k_3}\, r_1^{k_1} r_2^{k_2} r_3^{k_3} \tag{9.54}$$

with some constants c_{k_1,k_2,k_3}, and the set of allowed k_i's is determined by a combinatorial constraint. Then we see that asymptotically $\rho_n \sim (1/r_*)^n$. In fact, this leading asymptotic is simply generated by the geometric series $(1 - r_* Q)^{-1}$ in the infinite product (9.36), which is consistent with the statement that the Hagedorn singularity is encoded in the $k = 1$ factor in (9.36).

As explained in subsection 9.3.4, the roots r_* can be expressed analytically and they are given in closed form as

$$N_f = 0: \qquad r_* = \sqrt{2 - \sqrt{3}} \tag{9.55}$$

$$N_f = 1: \qquad r_* = \left(\frac{1}{2} - \frac{\sqrt{2}\sqrt[4]{3}}{2} + \frac{\sqrt{3}}{2}\right) \tag{9.56}$$

$$N_f \ge 2: \qquad r_* = \frac{\kappa^2 + 2 - \sqrt{\kappa^4 + 4}}{2\kappa}, \quad \kappa \equiv \left(2\,N_f + 2\sqrt{N_f^2 - 2}\right)^{1/3}. \tag{9.57}$$

The corresponding Hagedorn temperatures are given in Table 9.1. Notice that with increasing N_f, the Hagedorn temperature decreases, as expected, since adding more degrees of freedom to the theory leads to a faster growth of density of states.

9.4.2 Spatial compactification and the disappearance of the Hagedorn instability

We now discuss the theory on a spatial circle, with periodic boundary conditions for the fermions. The Euclidean path integral now computes the twisted partition function, $\tilde{Z}$, given in (9.37). This is the setting in which we expect large N volume independence to apply[147], so the Hagedorn instability should disappear. But getting rid of the Hagedorn instability is hard. It is not enough for the leading exponential behavior of the bosonic and fermionic density of states to be identical to get a twisted partition function without singularities. There are an infinite number of subleading exponentially-growing terms in the asymptotics of the bosonic and fermionic densities of states, and if *any* of them differ there will still be a Hagedorn instability. We now show that the degeneracies between the bosonic and fermionic states are sufficiently strong that this does not happen, and there are no Hagedorn instabilities in the twisted partition function. The absence of instabilities as a function of $L \in \mathbb{R}^+$ in the twisted partition function is illustrated in Fig. 9.1, which shows the locations of the poles in the twisted and thermal partition function as a function of $Q \in \mathbb{C}$.

With a spatial S^1, the polynomials that appear in the denominator of $\tilde{Z}$ are $P\big[(-Q)^k\big]$, and the singularities of $\tilde{Z}$ are determined by the roots of $\tilde{P}(Q) \equiv P(-Q)$,

$$\tilde{P}(Q) = Q^6 - 3Q^4 + 4N_f Q^3 - 3Q^2 + 1 = 0\,. \tag{9.58}$$

Given that the polynomial $Q^6 - 3Q^4 - 3Q^2 + 1 = (Q^2+1)(Q^4 - 4Q^2 + 1)$ has only one root in $[0, 1)$, and $\tilde{P}(0) = 1$ and $\tilde{P}(1) = 4(N_f - 1)$ are both non-negative, we see that none of roots of $\tilde{P}(Q)$ can be in $[0, 1)$. In fact, due to the $Q \to Q^{-1}$ symmetry of (9.58), the only roots of $P(Q)$ along the positive real axis can be at $Q = 1$. This is the case for $N_f = 1$. For $N_f > 1$, $P(Q)$ has no roots in the positive real axis at all. Furthermore, none of the factors with $k > 1$ can produce singularities in $[0, 1)$ either, since those singularities are given by the $1/k^{\text{th}}$ powers of roots of $\tilde{P}(Q)$, none of which are in $[0, 1)$. Therefore we conclude that the twisted partition function is singularity free for any L and reach our main conclusion:

Adjoint QCD on $S^3_R \times S^1_L$ with $N_f \geq 1$ and periodic boundary conditions on S^1_L *does not* have a Hagedorn instability and stays in the confined phase for *any* L at $N = \infty$.

We now give a physical explanation for this result by taking a closer look at the the twisted and thermal partition functions. The coefficients of Q^n in $\tilde{Z}$ count the number of bosonic states minus the number of fermionic states at energy $E_n = n/(2R)$, while in Z they count the number of bosonic states plus fermion states. The states counted by even powers of Q are purely bosonic, while states counted by odd powers of Q are purely fermionic.[13] Expanding the partition functions in Q with e.g. $N_f = 1$ yields

$$\tilde{Z}_{N_f=1}(Q) = 1 - 4Q^3 + 6Q^4 - 12Q^5 + 28Q^6 - 72Q^7 + 168Q^8 - 364Q^9 + 828Q^{10} + \cdots \quad (9.59)$$

$$Z_{N_f=1}(Q) = 1 + 4Q^3 + 6Q^4 + 12Q^5 + 28Q^6 + 72Q^7 + 168Q^8 + 364Q^9 + 828Q^{10} + \cdots \quad (9.60)$$

The coefficients ρ_n of Q^n grow rapidly with n and reach their asymptotic behavior $\rho_n \sim (1/r_*)^n$ quickly.

As illustrated in Fig. 9.3, where we plot the logarithms of d_n for $N_f = 2$, the asymptotic behavior of bosonic and fermionic density of states is identical. The sole difference between the thermal and the twisted case is that

$$d_n^{\text{twisted}} = (-1)^n d_n^{\text{thermal}} \quad (9.61)$$

where $d_n^{\text{twisted/thermal}}$ are the coefficients of Q^n. This is of course an obvious consequence of the definitions. What is far less obvious a priori is that as illustrated in Fig. 9.3, it appears that both the bosonic and fermionic degeneracy factors in the thermal partition function can be thought as coming from the *same* smooth function of n, which becomes monotonic past some $n = n_*$ (in the figure $n_* = 4$). This apparent underlying function gets sampled at even integers to give the bosonic degeneracies, and gets sampled at the odd integers to give the fermionic degeneracies. If an analytic continuation of d_n to a function $f(n)$ of $n \in \mathbb{C}$ were to be found explicitly and could be shown to be monotonic, it would be one way to demonstrate that the bosonic and fermionic hadronic states are entirely degenerate up to an offset due to the curvature for any N_f. We leave this challenging task to future work, since in our view understanding the degeneracy pattern in terms of symmetries may be more directly illuminating.

From Fig. 9.2 and Fig. 9.3 it is clear that the d_n^{twisted} coefficients form an alternating sequence with a symmetric envelope around zero. These oscillations, illustrated in Fig. 9.2, are behind the disappearance of the Hagedorn instability for the spatial compactification.

[13] The same result also follows from the fact that in the $R\Lambda \to 0$ limit, the energy of a given bosonic/fermionic state is simply given by the radial quantum number of the vector/spinor S^3 spherical harmonic function, i.e.

$$\omega_{B,n} = \frac{n+1}{R}, \qquad \omega_{F,n} = \frac{n+\frac{1}{2}}{R}$$

Since $Q^n = e^{-2L\omega_n}$, even/odd powers of Q^n correspond to bosonic/fermionic states respectively.

We note that this type of cancellation mechanism of bosonic and fermionic contributions to the twisted partition function is rather different than the more familiar "supersymmetry-like" fermion-boson cancellations, which occur *within* each given energy level. The cancellations we see in adjoint QCD on $S^3 \times S^1$ instead involve repeated cancellations neighboring levels of bosons and fermions. The same effect was seen in work on misaligned supersymmetry [138, 188, 139], and we discuss the connection between adjoint QCD and misaligned supersymmetry in Section 9.4.5. Note however that the offset between the bosonic and fermionic degeneracies which leads to the oscillations is due to the S^3 curvature. If $R\Lambda \gtrsim 1$ the curvature should become unimportant, and the boson-fermion cancellations should start taking place within each level if the theory still lacks a Hagedorn instability, as discussed in [148].

9.4.3 Twisted Casimir energy in adjoint QCD

In this section we compute the twisted vacuum energy

$$\tilde{C} = C_B - C_F \tag{9.62}$$

where C_B, C_F are the vacuum energies due to the bosonic states and C_F, which can be computed from the behavior of the twisted partition function. Since we are working on $S^3 \times S^1$, these vacuum energies can be thought of as Casimir energies on S^3, motivating the notation. The computation of Casimir energies $C = C_B + C_F$ in large N gauge theories on $S^3 \times S^1$ with thermal boundary conditions involves similar techniques but is more involved, and is discussed in the previous chapter (see also Ref. [7]).

To begin, recall that the physical states of this large N theory are single-trace operators, and their energies and degeneracies are counted by the twisted single-trace partition function from (9.26)

$$\begin{aligned}\tilde{Z}_{\mathrm{ST}}[q] &= -\sum_{m=1}^{\infty} \frac{\varphi(m)}{m} \log\left[1 - z_V(q^m) + N_f z_F(q^m)\right] &(9.63)\\ &\equiv \sum_{n=1}^{\infty} D_n e^{-L\omega_n} &(9.64)\end{aligned}$$

and $\omega_n = n/(2R)$ is the energy of the n-th mode with degeneracy D_n. Let us define

$$\tilde{C}(L) \equiv -\frac{1}{2}\frac{\partial \tilde{Z}_{\mathrm{ST}}}{\partial L} \quad = \tfrac{1}{2}\textstyle\sum_{n=1}^{\infty} D_n \omega_n e^{-L\omega_n}\,. \tag{9.65}$$

Then the twisted Casimir energy[14] can be formally written as

$$\tilde{C} = \frac{1}{2}\sum_{n=1}^{\infty} D_n \omega_n = -\frac{1}{2}\frac{\partial \tilde{Z}_{\rm ST}}{\partial L}\Big|_{L=0} = \tilde{C}(0)\,. \tag{9.66}$$

Of course this formal expression is divergent and has to be regularized and renormalized to extract the physical quantity $\tilde{C}$. Thanks to the absence of any phase transitions as L is varied, $\tilde{L}(C)$ is well-defined for any $L \neq 0$, and can be viewed as defining as a spectral regularization of the divergent sum in $\tilde{C}$. The structure of the singularities in the twisted single-trace partition function is illustrated in Fig. 9.1 for $N_f = 1$ and $N_f = 2$. The absence of any singularities on the positive real axis makes it easy to take the $L \to 0$ limit above. The situation is more subtle for thermal compactifications, see [7] for a full discussion.

Our renormalization prescription amounts to isolating the divergent part of $\tilde{C}(L \to 0)$ and extracting the L independent, finite part. The divergent part of $\tilde{C}(L)$, which scales with the UV cutoff μ as μ^2/R^2 [15] is absorbed by a $\mu^2 \int d^4x \sqrt{g}\,\mathcal{R}$ counter-term, and since the only divergence is a power law there are no issues with cutoff scheme dependence.

We now evaluate the twisted Casimir energy in two different ways. First, we use a hybrid zeta function and heat-kernel-like regularization procedure to extract the finite part of $\tilde{C}(L \to 0)$ analytically. Second, we directly evaluate $\tilde{C}(L)$ numerically, and confirm the findings of the analytical manipulations. The details of the numerical computation are explained in subsection 9.4.4. In both cases we find that the finite, L-independent part of $\tilde{C}(L \to 0)$ vanishes and conclude that the twisted Casimir energy of adjoint QCD on $S^3_R \times S^1_L$ at $N = \infty$ and small R is *zero* for any $N_f \geq 1$.

To compute $\tilde{C}$ we need to understand the $L \to 0$ limit in (9.66), and to this end we first isolate the part of the sum from (9.63) in $\partial \tilde{Z}/\partial L$ which is divergent:

$$\begin{aligned}\frac{1}{4R} Q \frac{\partial}{\partial Q} \log\left[1 - z_V(Q^{2m}) + N_f z_F(Q^{2m})\right] \;&=\; \frac{1}{2R}\left(\frac{3mQ^{2m}\left(2N_f Q^m - 2Q^{2m} + Q^{4m} - 1\right)}{Q^{2m}\left(4N_f Q^m - 3Q^{2m} + Q^{4m} - 3\right) + 1}\right) \\ &+\; \frac{3}{2R}\frac{mQ^{2m}}{Q^{2m} - 1}\end{aligned} \tag{9.67}$$

We can take $Q = 1$ in the first term since the divergent part is isolated in the second term.[16] Doing

[14] We emphasize that this definition relies on using the N independent spectrum obtained after large N limit being taken first. We thank O. Aharony, C. P. Herzog, and M. Yamazaki for discussions on this point.

[15] The absence of a μ^4 divergence is itself quite interesting. See [146] for a related recent discussion in the context of supersymmetric QFTs.

[16] For $N_f = 1$, the separation of the divergent and finite part in (9.67) is different. However the N_f dependence drops out in the final answer for the twisted Casimir energy for arbitrary N_f. So, taking $N_f = 1$ at the end of the calculation, as presented above, is safe.

so, we arrive at the expression

$$\tilde{C}(L \to 0) = -\frac{3}{4R}\sum_{m=1}^{\infty}\varphi(m) - \frac{3}{2R}\lim_{L\to 0}\sum_{m=1}^{\infty}\varphi(m)\frac{Q^{2m}}{1-Q^{2m}}\,. \tag{9.68}$$

Both of these expressions are formally divergent. Regulating the first term using the zeta-function identity $\sum_{m=1}^{\infty}\varphi(m)m^{-s} = \zeta(s-1)/\zeta(s)$, and using a Lambert series identity $\sum_{m=1}^{\infty}\varphi(m)q^m/(1-q^m) = q/(1-q)^2$ for the second term, leads to the result

$$\begin{aligned}\tilde{C}(L \to 0) &= -\frac{3}{4R}\frac{\zeta(-1)}{\zeta(0)} - \frac{3}{2R}\lim_{L\to 0}\frac{Q^2}{(1-Q^2)^2} \\ &= -\frac{3\,\zeta(-1)}{4\,\zeta(0)R} + \frac{1}{8R} - \frac{3R}{2L^2} = -\frac{3R}{2L^2} + 0 \times L\end{aligned} \tag{9.69}$$

The fact that the L-independent term vanishes yield the conclusion that the twisted Casimir energy vanishes.

Two remarks about the calculation above are in order. First, in principle, one might be worried about the algebraic manipulations such as splitting terms in formally divergent sums and regularizing them individually. This is not an issue because $\tilde{C}(L)$ is finite for any finite L. Moreover, even if (9.67) is not viewed in the context of being embedded in the regularized expression $\tilde{C}(L)$, note that both of the regularizations leading to (9.69) involve cutoff functions which only depend on the energy spectrum, justifying the manipulations. Second, one might be concerned that the L^{-2} terms in the analytical calculation above and in the numerical computation in subsection 9.4.4 are different. This is not issue, because only the finite L-independent terms are physical and regulator independent. The divergent pieces do not have to agree if different regulators are used. The numerical calculation extracts $\tilde{C}$ directly from the scaling of $\tilde{C}(L)$ at small L, while the analytic calculation brings in a zeta function along the way, which amounts to a modification of the regularization scheme and a corresponding difference in the coefficients of the divergent pieces in the two computations.

The underlying physical reason for the remarkable result that the twisted Casimir energy is zero is not known to us, but presumably it is a consequence of previously unrecognized symmetries of large N adjoint QCD, as are the rest of our results. We note that it is actually expected from the fact that the twisted partition function of (9.29) has a T-reflection symmetry with a zero vacuum energy, as noted in [6]. A more detailed exploration of the very interesting interplay between T-reflection symmetry and the vacuum energy of confining large N gauge theories on $S^3 \times S^1$ is discussed in [7].

$\mathbf{N_f = 1}$				
M	CR	C_1	C_2R	σ
5.00×10^2	2.74×10^{-2}	1.94×10^{-3}	1.12	2.53×10^{-7}
8.00×10^2	4.64×10^{-3}	2.78×10^{-4}	1.12	6.34×10^{-8}
1.50×10^3	1.51×10^{-3}	6.95×10^{-5}	1.12	2.07×10^{-8}
8.00×10^3	1.18×10^{-4}	1.60×10^{-6}	1.12	3.45×10^{-9}
$\mathbf{N_f = 2}$				
M	CR	C_1	C_2R	σ
5.00×10^2	5.45×10^{-2}	3.88×10^{-3}	3.40	1.68×10^{-7}
8.00×10^2	8.76×10^{-3}	5.38×10^{-4}	3.40	4.24×10^{-8}
1.50×10^3	2.66×10^{-3}	1.31×10^{-4}	3.40	1.46×10^{-8}
8.00×10^3	1.61×10^{-5}	2.59×10^{-7}	3.40	5.89×10^{-10}

Table 9.2: Best-fit parameters for the low-L behavior of $C(L)$ as a function of the cutoff M. Note that in both the $N_f = 1$ and $N_f = 2$ theories the twisted Casimir energy C goes to zero as the cutoff M is removed.

9.4.4 Numerical computation of the twisted Casimir energy

We compute C numerically. If we cut off the infinite sum in (9.63) at some high $n = M$, then $C(L)$ rapidly becomes insensitive to M except at low L. Accessing lower L requires increasing M. In Fig. 9.4 we illustrate the dependence of the low-β behavior on the cutoff M in $N_f = 2$ adjoint QCD. From the figure it is clear that the leading small L divergence in $C(L)$ is $\sim 1/L^2$. One can then verify that the M-independent small-L regions of $C(\mu)$ can be modeled to a very high accuracy by a polynomial fit function $F(L)$:

$$F(L) = C + \frac{C_1}{L} + \frac{C_2 R}{L^2} \tag{9.70}$$

The parameters C, C_1, C_2 are read off from a least-squares fit of the $F(L)$ to $C(L)$ at low L for a variety of values of M. We then take $M \to \infty$ limit. Our results for $N_f = 1$ and $N_f = 2$ are summarized in Table 9.2. To characterize the quality of the fits to the function (9.70), the tables also show the value of

$$\sigma = \frac{1}{n} \sqrt{\sum_{L_i} \left(\frac{C(L_i) - F(L_i)}{C(L_i)} \right)^2} \tag{9.71}$$

where $L_i, i = 1, 2, \lambda \cdots, n$ are the set of values of L used to do the fit. A good fit is characterized by $\sigma \ll 1$, which is true for all the cases we show. Our results for higher N_f are similar. We find that the best-fit values of C decrease rapidly toward zero with increasing M, and an extrapolation to $M = \infty$ results in $C = 0$ for all $N_f \geq 1$. The same is true for C_1, while C_2 has a non-zero limit which depends on N_f.

9.4.5 Relation to misaligned supersymmetry

We have seen that the way spatially compactified adjoint QCD on $S^3 \times S^1$ escapes the Hagedorn instability involves cancellations between the bosonic and fermionic densities of states, both of which grow exponentially, and the cancellations arise due to an oscillation between the number of bosonic and fermionic states at successive excitation levels.

These cancellations fit the framework of 'misaligned supersymmetry' developed in [138, 188, 139]. These papers explored the structure of the partition functions of perturbative fundamental closed string theories. Consistent closed string theories are always modular-invariant, but may or may not have spacetime supersymmetry. Refs. [138, 188, 139] pointed out that modular invariance along with the absence of tachyons implies certain intricate patterns of relations between the degeneracies of bosonic and fermionic states. These relations imply that the leading exponentially-growing parts of the bosonic and fermionic densities of states in the closed string theories cancel against each other in the twisted partition function. With spacetime supersymmetry, the cancellations occur within each level. More generally, however, for modular-invariant string partition functions without spacetime supersymmetry, these cancellations are due to sign-oscillating mismatches between bosonic and fermionic state degeneracies[138, 188, 139]. Misaligned supersymmmetry can also imply the vanishing of super-traces which contribute to the cosmological constant and its divergences[141].

Such oscillating cancellations between bosonic and fermionic states are exactly what we have seen in our analysis. In this sense, large N adjoint QCD on $S^3 \times S^1$ with $N_f \geq 1$ appears to give the first known field-theoretic realization of the string-theoretic idea of misaligned supersymmetry. This raises many interesting questions. For instance, in the analysis of [138, 188, 139] the modular invariance of the partition functions of string theories played a starring role. Large N gauge theories are believed to be describable as some kind of weakly-coupled string theories, so if adjoint QCD enjoys a realization of misaligned supersymmetry, one might wonder whether its partition function enjoys some form of modular invariance. If the partition function were to be modular invariant, the would yield an underlying reason for the pattern of cancellations. We now explore this possibility.

Modular invariance of a partition function Z for a theory on a spatial circle implies

$$Z(\tau) = Z(\tau + 1) = Z(-1/\tau) \tag{9.72}$$

where τ is defined through $Q = e^{2\pi i \tau} = e^{-\frac{L}{2R}}$. Hence modular invariance implies $Z(L) = Z[(4\pi R)^2/L]$, which is a manifestation of T-duality. However, the twisted partition function (9.37) does not have modular invariance. The simplest way to see this is to observe that $\tilde{Z}(L)$ does not have the right shape for modular invariance, as is illustrated in Fig. 9.5, since it has different limits for $L \to 0$ and $L \to \infty$, approaching 0 and 1 respectively. We can also see the lack of modular invariance algebraically. By using the modular properties of the Dedekind function, and Jacobi's transformation identities for the theta functions, it can be shown that under the two generators of $SL(2, Z)$ modular transformations

$$T:\ \tau \to \tau + 1 \qquad S:\ \tau \to -1/\tau \tag{9.73}$$

where τ is assumed to be in the upper half-plane, the full partition function transforms as

$$\tilde{Z}_{\mathrm{QCD[Adj]}}(\tau + 1) = \tilde{Z}_{\mathrm{QCD[Adj]}}(\tau) \tag{9.74}$$

$$\tilde{Z}_{\mathrm{QCD[Adj]}}(-1/\tau) = (-i\tau)^{3/2}\,\frac{\eta^3(\tau/2)}{\eta^3(2\tau)}\left(\prod_{i=1}^{3}\frac{e^{i\tau\nu_i^2/\pi}\vartheta_2(\nu_i|e^{i\pi\tau})}{\vartheta_4(\tau\nu_i|e^{i\pi\tau})}\right)\tilde{Z}_{\mathrm{QCD[Adj]}}(\tau) \tag{9.75}$$

where $e^{2i\nu_i} = r_i$. This means that the partition function of large N adjoint QCD on $S^3 \times S^1$ is *not* invariant under the $SL(2, \mathbb{Z})$ modular group, nor does it transform as a modular form. However, as discussed extensively in e.g. [138], closed string partition functions are made from special combinations of both holomorphic and antiholomorphic (in τ) modular functions. As a result the modular invariance of closed string theories is intimately related to the fact that string partition functions include contributions from 'off-shell' states with $m \neq n$ where (m, n) are the world-sheet energies of (left, right) moving states. Such states do not appear in field theory, so one should not normally expect that modular invariance would show up in any simple way in a field theory partition function, even if the field theory has a dual description as a string theory with modular invariance.[17] Nevertheless, it would be interesting to explore whether our results are some sort of field-theoretic remnant of misaligned supersymmetry in the string dual of adjoint QCD.

Of course, we are dealing with a weakly-coupled limit of adjoint QCD, so the phenomena we are seeing should have a description directly within field theory in any case. While it would be wonderful to understand the string theory dual of the adjoint QCD, there should be no need to do this to understand the pattern of degeneracies between bosonic and fermionic states that we have seen. In the next section we make some remarks on how our results may be understood directly in field theory through emergent fermionic symmetries.

[17] We thank K. Dienes for explaining this to us.

9.5 Emergent fermionic symmetries in adjoint QCD on $S^3 \times S^1$

In this section we comment on the relation between our results and the notion of emergent fermionic symmetries in the large N limit.

9.5.1 $N_f = 1$

$SU(N)$ massless adjoint QCD in flat space with $N_f = 1$ has $\mathcal{N} = 1$ supersymmetry, since it is just $\mathcal{N} = 1$ super-Yang-Mills theory. However, the supersymmetry is broken on $S^3 \times S^1$ due to the curvature couplings. On a curved generic manifold there are no covariantly constant spinors, so there is no way to define conserved supercharges. The exception is when the compactification manifold has enough isometries *and* the field theory has a non-anomalous continuous $\mathcal{R}$ symmetry.[18] In general, 4D $\mathcal{N} = 1$ SUSY QFTs have a classical $U(1)_{\mathcal{R}}$ global symmetry. When a 4D $\mathcal{N} = 1$ theory is compactified on $S^3_R \times \mathbb{R}$, the SUSY algebra is modified from its flat-space form to (see e.g. [189]):

$$\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} \quad = 2i\sigma^\mu_{\alpha\dot{\alpha}}\partial_\mu - \tfrac{2}{R}\sigma^0 n_{\mathcal{R}} \tag{9.76}$$

Here $n_{\mathcal{R}} = \int d^3x\, j^0_{\mathcal{R}}$ is the charge operator associated with the $U(1)$ $\mathcal{R}$-current $j^\mu_{\mathcal{R}}$. Under the $\mathcal{R}$ symmetry, gauge fields have charge zero, while the Weyl fermions have charge 1. Hence when there is an unbroken continuous $\mathcal{R}$ symmetry in the full quantum theory, supersymmetry is preserved on $S^3 \times \mathbb{R}$ and on $S^3 \times S^1$ with periodic boundary conditions.

This setup does not work for $\mathcal{N} = 1$ $SU(N)$ SYM, since it suffers from a chiral anomaly that breaks $U(1)_{\mathcal{R}} \to Z_{2N}$. So there is no continuous $\mathcal{R}$ symmetry.[19] As a result the classical supersymmetry of $N_f = 1$ $SU(N)$ adjoint QCD on $S^3_R \times \mathbb{R}$ or $S^3_R \times S^1_L$ suffers from an anomaly, and the theory has no fermionic symmetries except in the $\mathbb{R}^4$ limit.

This raises a puzzle, because $N_f = 1$ adjoint QCD on $S^3_R \times S^1_L$ with $R\Lambda \ll 1$ has unbroken center symmetry for any $L \sim N^0$, enjoys large N volume independence, and has no Hagedorn instabilities for any L. The absence of Hagedorn instabilities is due to conspiracies between the bosonic and fermionic densities of states which amount to relations between degeneracies and energies of an

[18] Then one can define a 'twisted' subgroup of the Lorentz symmetry which lives in a diagonal subgroup of isometry transformations and $\mathcal{R}$ symmetry rotations, and at least some fraction of the original supersymmetry can be preserved in the compactified theory. For discussions of how this works for theories with $\mathcal{N} \geq 1$ supersymmetry on $S^3 \times \mathbb{R}$ and $S^3_R \times S^1_L$ see [189, 123, 190, 191].

[19] On $\mathbb{R}^4$, there is a further spontaneous breaking of the non-anomalous part of the $\mathcal{R}$-symmetry down to $\mathbb{Z}_2$.

infinite number of bosonic and fermionic states. As argued in the introduction and in [148], this seems to call for a symmetry. And yet we have just said that the $SU(N)$ $N_f = 1$ theory definitely has no fermionic symmetries. What is going on? We now argue that the resolution of the puzzle is that there is an emergent large N fermionic symmetry.

Recall that the chiral anomaly for the would-be conserved current $j_\mu^{\mathcal{R}}$ is

$$\partial^\mu j_\mu^R = \frac{\lambda}{16\pi^2} \mathrm{Tr} F_{\mu\nu} \tilde{F}^{\mu\nu}, \tag{9.77}$$

This anomaly equation has no manifest $1/N$ suppression factors, and $U(1)_{\mathcal{R}}$ breaking appears to be unsuppressed at large N. While this is true, there are some important subtleties on $S_R^3 \times S_L^1$ with $R\Lambda \ll 1$, the regime in which we are working.

It is useful to recall the reason for the anomaly breaking pattern $U(1)_{\mathcal{R}} \to \mathbb{Z}_{2N}$. The origin of the unbroken $\mathbb{Z}_{2N}$ factor lies in the fact that the right-hand side of the anomaly equation is a total derivative, and is only non-zero on instanton field configurations with non-zero topological charge Q. But in the $N_f = 1$ theory the instantons carry $2N|Q|$ fermion zero modes, and generate effective 't Hooft vertex interactions for the fermions which break $U(1)_{\mathcal{R}}$ but are invariant under its $\mathbb{Z}_{2N}$ subgroup. So the interacting theory only enjoys the $\mathbb{Z}_{2N}$ symmetry. On the one hand, at large N, a $\mathbb{Z}_{2N}$ symmetry ought to have the same power as a $U(1)$ symmetry, up to $1/N$ corrections. This makes it appear that the anomaly is suppressed at large N. On the other hand, the anomaly cannot be suppressed, because the RHS of Eq. (9.77) is unsuppressed relative to the LHS.

Despite first appearances, these observations are not in conflict with each other. To get a non-vanishing contribution from the right-hand side of Eq. (9.77) one must consider correlation functions with enough fermion operators to saturate the $2N|Q|$ zero modes. Let us call color-singlet operators with $\gtrsim N^1$ fermionic operators inside the color trace 'heavy', and call operators which have $\sim N^0$ fermions 'light'. The fact that this distinction can be made relies on the fact that in the regime we are considering, $R\Lambda \ll 1$, there is no chiral condensate, so there is no spontaneous breaking $\mathbb{Z}_{2N} \to \mathbb{Z}_2$. So it makes sense to classify operators by their $\mathbb{Z}_{2N}$ charge when $R\Lambda \ll 1$.[20] It does not make sense to do so if $R\Lambda \gtrsim 1$, because then the $\mathbb{Z}_{2N}$ symmetry becomes spontaneously broken due to the formation of a gluino condensate.

[20]To see this recall that the fermions have a effective curvature-induced $\mathcal{R}$-symmetry-preserving mass $1/(2R)$. This implies that e.g. the two-point correlation function falls off exponentially:

$$\langle \lambda\lambda(t)\ \lambda\lambda(0)\rangle \sim e^{-t/(2R)} \tag{9.78}$$

So there is no long-range order, meaning that there is no spontaneous breaking of the discrete remnant of the $\mathcal{R}$ symmetry. Note as well that the absence of spontaneous symmetry breaking is not happening for trivial Coleman-Mermin-Wagner reasons, since we are working at large N.

The N-independence of the right-hand side of Eq. (9.77) means that for heavy states the $\mathcal{R}$-symmetry is irredeemably broken. There is no reason to expect their energies and degeneracies to be related to each other by any fermionic symmetry. But consider states whose interpolating operators are light. Correlators of light operators cannot saturate the instanton zero modes, so for these states the $\mathbb{Z}_{2N}$ symmetry gives non-trivial relations. At large N, as far as these light states are concerned, the theory enjoys a $U(1)_{\mathcal{R}}$ symmetry. These light states are precisely the ones that are important throughout our analysis of partition functions with $L \sim N^0$.[21] So *when acting on states that remain light at large* N, the SUSY algebra in Eq. (9.76) is anomaly-free up to $1/N$ corrections.

The punchline should now be clear: $N_f = 1$ adjoint QCD on $S^3 \times S^1$ has an emergent fermionic symmetry in the large N limit. Not coincidentally, it also enjoys large N volume independence, with no Hagedorn instabilities in the twisted partition function thanks to massive cancellations between bosonic and fermionic densities of states.

9.5.2 $N_f > 1$

The relations we saw between the spectrum of bosonic and fermionic hadronic excitations in adjoint QCD on $S^3 \times S^1$ are very similar for $N_f = 1$ and $N_f > 1$. Here we comment on the symmetries of adjoint QCD for $N_f > 1$. First, note that at the microscopic level, adjoint QCD has $2(N^2-1)$ bosonic degrees of freedom (from the gluons) and $2N_f(N^2-1)$ fermionic ones (from the quarks). Once $N_f > 1$, something more exotic than the story in Sec. 9.5.1 is necessary due to the mismatch in the number of microscopic degrees of freedom. It seems that any emergent fermionic symmetry could not be a standard supersymmetry. What could it look like?[22]

At the moment we can only make a suggestive observation in this direction. We have seen above that the $\lambda \to 0$ limit of adjoint QCD on $S^3 \times S^1$ is already very interesting, with many of the features of the $\lambda > 0$ theory (such as confinement) remaining qualitatively preserved. With this as an inspiration we examine the $\lambda = 0$ limit of adjoint QCD in *flat space* and show that it has a fermionic symmetry for any $N_f \geq 1$. The Lagrangian density of the theory is

$$\mathcal{L} = \frac{1}{g^2}\mathrm{Tr}\left[-\frac{1}{2}F^2 + 2i\sum_{a=1}^{N_f}\left(\bar{\psi}_{a\dot{\alpha}}\bar{\sigma}^{\mu\dot{\alpha}\alpha}D_\mu\psi_{a\alpha}\right)\right] \tag{9.79}$$

where $\psi_{a\alpha}, a = 1, \lambda \cdots, N_f$, α is a spinor index is an adjoint Weyl fermion, and $F_{\mu\nu} = \partial_\mu A_\nu -$

[21] If $L \sim N^{-1}$, states with energies of order N start to participate in the partition function, and volume independence is expected to be lost on very general grounds. This fits nicely with our discussion here: the emergent symmetry should stop being effective once L becomes of order $1/N$.

[22] We are very grateful to D. Dorigoni for collaboration on the material in this section at an early stage.

$\partial_\nu A_\mu - i[A_\mu, A_\nu], D_\mu\psi^a = \partial_\mu\psi^a - i[A_\mu, \psi^a]$. The equations of motion are

$$D_\mu F^{\mu\nu} = [\bar{\psi}^a_{\dot{\alpha}}\bar{\sigma}^{\nu\dot{\alpha}\alpha}, \psi_{a,\alpha}], \qquad \bar{\sigma}^{\mu\dot{\alpha}\alpha} D_\mu \psi_{a\alpha} = 0. \tag{9.80}$$

We now exhibit field variations that lead to a fermionic symmetry in the $\lambda = 0$ limit for any $N_f \geq 1$. The variations are proportional to N_f infinitesimal Weyl fermion parameters $\bar{\epsilon}^a, \epsilon_a$:

$$\delta A^\mu(x) = -\frac{1}{\sqrt{2}}\left[\bar{\epsilon}^a_{\dot{\alpha}}\bar{\sigma}^{\mu\dot{\alpha}\alpha}\psi_{a\alpha}(x) + \bar{\psi}^a_{\dot{\alpha}}(x)\bar{\sigma}^{\dot{\alpha}\alpha}_\mu \epsilon_{a\alpha}\right] \tag{9.81}$$

$$\delta\psi_{a\alpha}(x) = \frac{-i}{2\sqrt{2}}\sigma^\mu_{\ \alpha\dot{\beta}}\bar{\sigma}^{\nu\dot{\beta}\beta}\epsilon_{a\beta}\, F_{\mu\nu}(x) \tag{9.82}$$

$$\delta\bar{\psi}^a_{\dot{\alpha}}(x) = \frac{+i}{2\sqrt{2}}\bar{\epsilon}^a_{\dot{\beta}}\bar{\sigma}^{\nu\dot{\beta}\alpha}\sigma^\mu_{\alpha\dot{\alpha}}\, F_{\mu\nu}(x) \tag{9.83}$$

Note that for $N_f = 1$ these are simply the $\lambda = 0$ limit of the standard on-shell $\mathcal{N} = 1$ SUSY transformations. To check the variation of the action, we write

$$\delta\mathcal{L} = \delta\mathcal{L}|_{\text{gauge}} + \delta\mathcal{L}|_{\text{fermion}} \tag{9.84}$$

where

$$\delta\mathcal{L}|_{\text{gauge}} = \frac{-1}{2}\text{Tr}\left[2F_{\mu\nu}\delta\left(F^{\mu\nu}\right)\right] \tag{9.85}$$

and

$$\delta\mathcal{L}|_{\text{fermion}} = 2i\delta\left(\text{Tr}\left[\bar{\psi}^a\bar{\sigma}^\mu D_\mu\psi_a\right]\right) \tag{9.86}$$

$$= 2i\text{Tr}\left[\delta\left(\bar{\psi}^a\right)\bar{\sigma}^\mu D_\mu\psi_a\right] + 2i\text{Tr}\left[\bar{\psi}^a\bar{\sigma}^\mu D_\mu\delta\left(\psi_a\right)\right] + 2i\text{Tr}\left[\bar{\psi}^a\bar{\sigma}^\mu i[\delta\left(A_\mu\right), \psi_a]\right]$$

$$= 2i\text{Tr}\left[\delta\left(\bar{\psi}^a\right)\bar{\sigma}^\mu\partial_\mu\psi_a\right] + 2i\text{Tr}\left[\bar{\psi}^a\bar{\sigma}^\mu\partial_\mu\delta\left(\psi_a\right)\right] \tag{9.87}$$

and in the last line we passed to the $\lambda \to 0$ limit.

One can next verify that

$$\delta\mathcal{L}|_{\text{gauge}} = \partial^\mu\text{Tr}\left[\frac{2}{2\sqrt{2}}F_{\mu\nu}\bar{\epsilon}^a\bar{\sigma}^\nu\psi_a\right] - \text{Tr}\left[\frac{2}{2\sqrt{2}}\left(D^\mu F_{\mu\nu}\right)\bar{\epsilon}^a\bar{\sigma}^\nu\psi_a\right] + \text{h.c.}, \tag{9.88}$$

while

$$\begin{aligned} \delta\mathcal{L}|_{\text{fermion}} &= \partial_\mu \mathrm{Tr}\left[\frac{-1}{2\sqrt{2}}\bar{\epsilon}^a\bar{\sigma}^\nu\sigma^\alpha\, F_{\alpha\nu}\bar{\sigma}^\mu\psi_a + \text{h.c.}\right] \\ &+ \mathrm{Tr}\left[\frac{2}{2\sqrt{2}}\bar{\epsilon}^a\bar{\sigma}^\nu\partial^\alpha F_{\alpha\nu}\psi_a\right] + \mathrm{Tr}\left[\frac{2}{2\sqrt{2}}\bar{\psi}^a\bar{\sigma}^\nu\partial^\alpha F_{\alpha\nu}\epsilon_a\right] \end{aligned} \tag{9.89}$$

where we used

$$\bar{\sigma}^\nu\sigma^\alpha\bar{\sigma}^\mu = -\eta^{\nu\mu}\bar{\sigma}^\alpha + \eta^{\alpha\mu}\bar{\sigma}^\nu + \eta^{\nu\alpha}\bar{\sigma}^\mu + i\epsilon^{\nu\alpha\mu\kappa}\bar{\sigma}_\kappa. \tag{9.90}$$

twice.

So acting on $\mathcal{L}$, the field variations above lead to

$$\delta\mathcal{L} = \partial_\mu G^\mu \tag{9.91}$$

where

$$G^\mu = \mathrm{Tr}\left[\frac{2}{2\sqrt{2}}F^{\mu\nu}\bar{\epsilon}^a\bar{\sigma}_\nu\psi_a\right] + \mathrm{Tr}\left[\frac{-1}{2\sqrt{2}}\bar{\epsilon}^a\bar{\sigma}^\nu\sigma^\alpha\, F_{\alpha\nu}\bar{\sigma}^\mu\psi_a\right] + \text{h.c.}, \tag{9.92}$$

so that the variation of the action is a total derivative. This means that these field variations are associated with a fermionic symmetry, with N_f spin-3/2 Noether currents

$$J_a^{\mu\dot{\kappa}} = \frac{1}{\sqrt{2}}\mathrm{Tr}\left[F_{\rho\nu}(\bar{\sigma}^\nu\sigma^\rho\bar{\sigma}^\mu\psi_a)^{\dot{\kappa}}\right]\,. \tag{9.93}$$

One can easily verify that these Noether currents are conserved at $\lambda = 0$: $\partial_\mu J_a^\mu$ vanishes on-shell by using (9.80). Hence there are $4N_f$ conserved fermionic charges in the $\lambda = 0$ limit of adjoint QCD, in flat space.

While it is amusing that there is a fermionic symmetry in flat-space adjoint QCD at $\lambda = 0$, this observation raises two obvious questions. First, it would be very interesting to work out how this fermionic symmetry behaves on $S^3 \times S^1$ in the limit $R\Lambda \ll 1$. To answer this question one would first need to understand the full symmetry algebra generated by the combination of the fermionic charges, the bosonic flavor charges, and the Poincare charges. An answer to this question with $\lambda = 0$ should already be quite interesting since it seems quite unlikely that such a symmetry algebra could be a standard superalgebra. Indeed, there are reasons to suspect that the full symmetry algebra may end up being infinite-dimensional.[23] Second, it would be even more interesting to understand whether a generalization of this kind of symmetry can survive at $\lambda \neq 0$ in the large N limit of adjoint QCD. An exploration of some of these issues is now in progress[192].

9.6 Conclusions

We have studied adjoint QCD in the large N limit on $S^3_R \times S^1_L$ in the weakly coupled limit $R\Lambda \ll 1$. Despite being weakly coupled, these theories have all of the features one would expect from any well-to-do confining large N theory, with a Hagedorn spectrum of stable hadrons created by single-trace operators. We have found that the bosonic and fermionic density of states have a Hagedorn growth. Nevertheless, the bosonic and fermionic states appear to be essentially degenerate up to a curvature-driven misalignment for any $N_f \geq 1$ as discussed in Sec. 9.4. The spatially compactified theory was explicitly shown to have no Hagedorn instabilities due to enormous cancellations between bosons and fermions. Our analysis shows that adjoint QCD stays in the confining phase persists of any L, and hence enjoys large N volume independence for any L. We also found that the difference of bosonic and fermionic Casimir energies vanishes.[24] As discussed in Sec. 9.4.5 large N adjoint QCD on $S^3 \times S^1$ appears to provide a field theoretic example of the idea of misaligned supersymmetry from string theory.

Our results involve conspiracies between the energies and degeneracies of all of the bosonic and fermionic hadronic excitations. This is quite surprising, since the family of theories we consider is not supersymmetric at any finite N, and so cannot have any fermionic symmetries at finite N. Since our results are obtained in the large N limit, rather than at finite N, they cry out for an explanation in terms of an emergent fermionic symmetry large N, as advocated in [148]. We have shown that such a symmetry emerges for $N_f = 1$, as discussed in Sec. 9.5, while for $N_f > 1$ we were only able to make some preliminary observations.

The analysis we have done takes essential advantage of the $R\Lambda \ll 1$ weak-coupling limit, and it is not clear how to generalize it to study the decompactified regime $R\Lambda \gtrsim 1$, where the theory becomes strongly coupled. Understanding what happens with large N volume independence once $R\Lambda \gtrsim 1$ presumably requires different techniques, such as numerical lattice calculations, or a refined understanding of the large N symmetries of adjoint QCD.

Indeed, the most pressing direction for future work is understanding whether (and if so, how)

[23]We are very grateful to S. Dubovsky for alerting us to this possibility and for related discussions.

[24]In [7] it is shown that the sum of the Casimir energies also vanishes at $N = \infty$ in the confined phase, and these two observations taken together imply that these Casimir energies are actually separately zero.

fermionic symmetries emerge at large N in $N_f > 1$ adjoint QCD, either on $S^3 \times S^1$ or directly on $\mathbb{R}^4$. The stakes are high: historical experience with supersymmetry shows that fermionic symmetries can be very powerful, and given the very close relationship between adjoint QCD and a sensible large N limit of real-world QCD[193, 194], finding such symmetries in adjoint QCD could be very useful both theoretically and phenomenologically.

9.7 Acknowledgements.

We are grateful to A. Armoni, T. D. Cohen, S. Dubovsky, V. Gobernko, B. Lucini, M. Shifman, and M. Yamazaki for fun and inspirational discussions at various stages of the long gestation of this chapter. We owe a special thanks to M. Ünsal and D. Dorigoni for encouragement, extensive discussions, and collaboration on related topics. We are also deeply grateful to K. Dienes for very inspirational discussions and comments on the manuscript. This work is supported by the U.S. Department of Energy under the grants DE-FG-88ER40388 (G. B.), DE-FG02-93ER-40762 (G. B.), and DE-SC0011842 (A. C.).

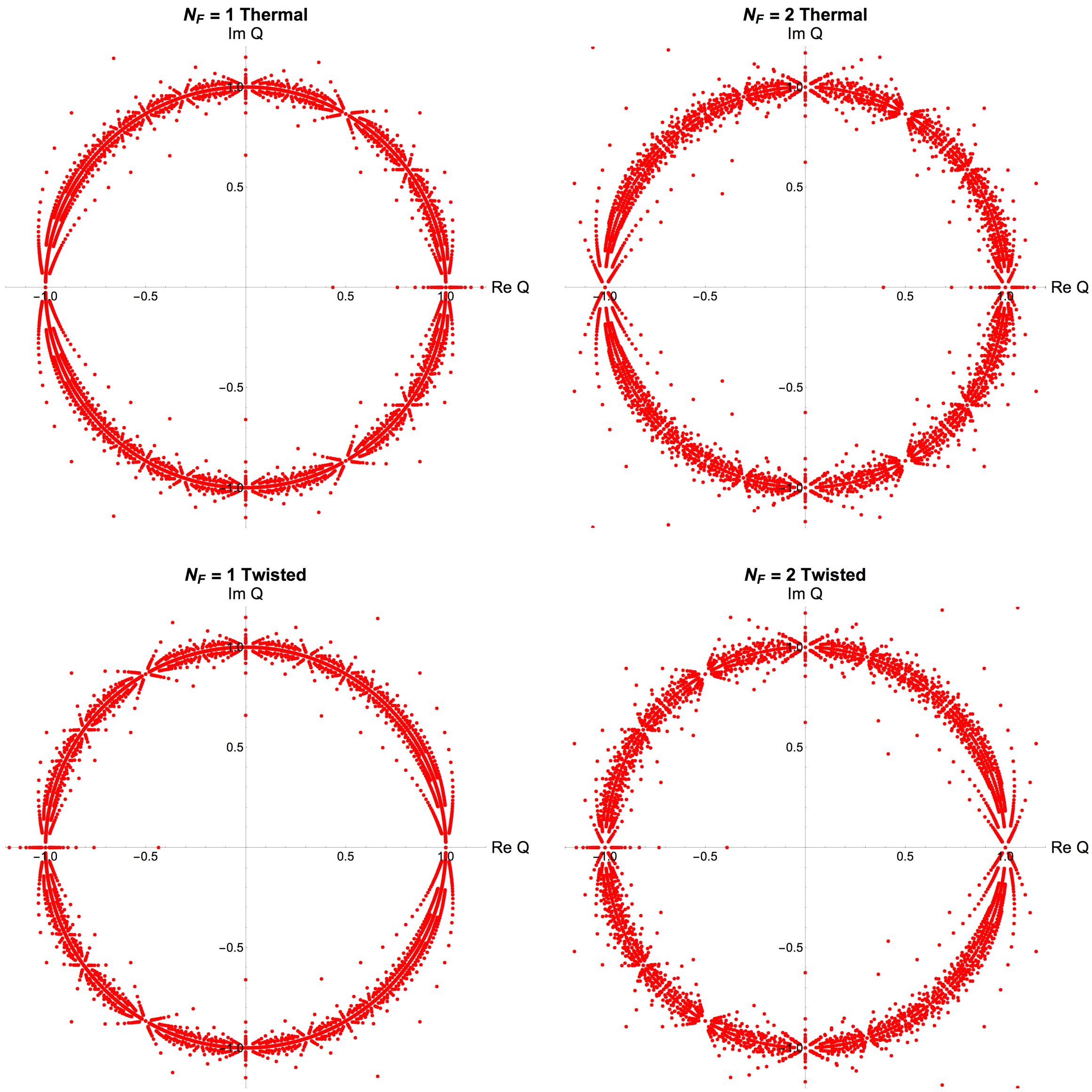


Figure 9.1: (Color Online.) This plot summarizes much of the chapter. The red dots are singularities of the thermal (top row) and twisted (bottom row) partition functions of adjoint QCD as a function of complex temperature $Q = e^{-L/2R}$ for $N_f = 1$ (left column) and $N_f = 2$ (right column). The absence of singularities on the positive real axis (except at $Q = 1$, corresponding to $L = 0$) is tied to the absence of Hagedorn instabilities in the twisted partition function. The evident $Q \to -Q$ symmetry relating the singularity structure of the twisted and thermal partition follows from (9.36) and (9.37). For visual clarity we only show singularities arising from the first 30 terms in (9.36) and the first 45 terms in (9.37).

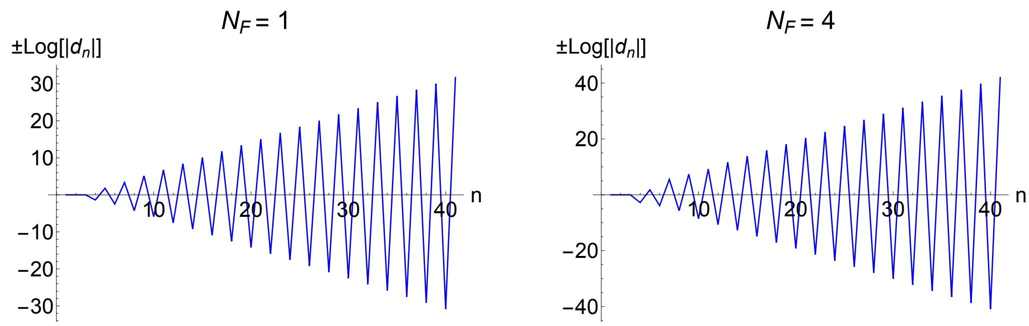


Figure 9.2: Logarithms of the coefficients of Q^n of the series expansion of the twisted partition function $\tilde{Z}(Q)$, with $+/-$ signs for bosons/fermions. The coefficients of even/odd powers of Q are boson/fermion degeneracy factors. We draw lines between successive data points as a visual aid to make the oscillations easier to follow. The linearity of the envelope function means that the bosonic and fermionic densities of states both have Hagedorn growth, while the symmetry of the envelope function around zero is responsible for the elimination of Hagedorn instabilities in the twisted partition function.

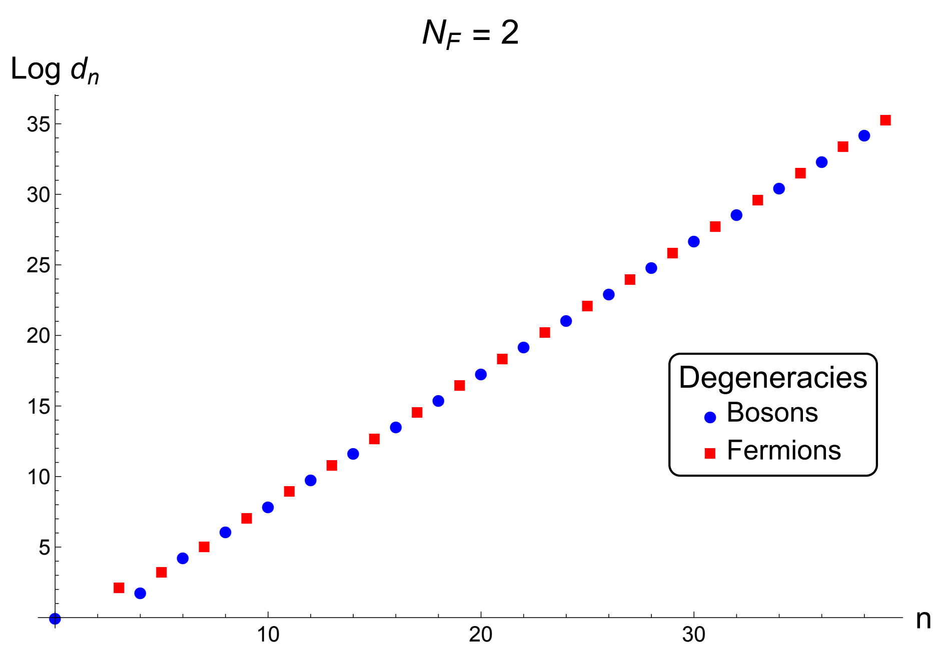


Figure 9.3: Logarithms of the coefficients of Q^n of the series expansion of the thermal partition function $Z(Q)$ for $N_F = 2$. The bosonic and fermionic state degeneracy factors have identical asymptotic scaling with n.

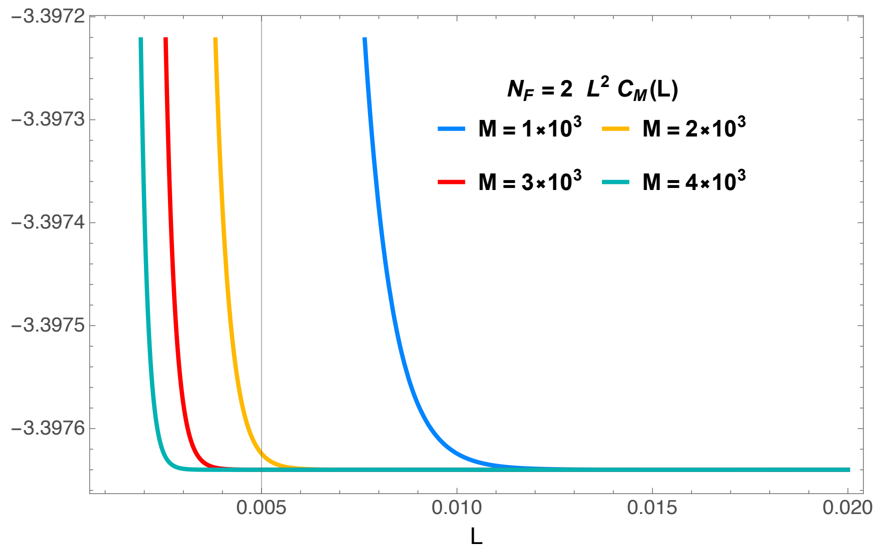


Figure 9.4: Behavior of $L^2C(L)$ at small L for $N_f = 2$ (as an example) as a function of a cutoff M on the upper end of the sum in (9.63).

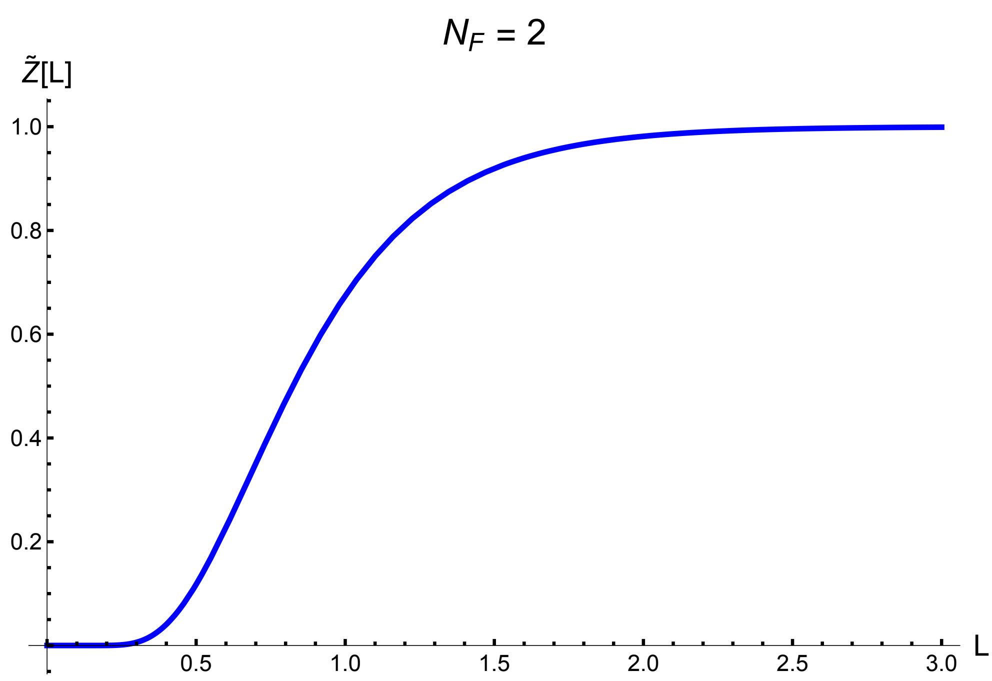


Figure 9.5: Plot of the $N_f = 2$ partition function on $S^3_R \times S^1_L$ when $R\Lambda \ll 1$, which illustrates the lack of invariance under $L \to \frac{c}{L}$ for any $c > 0$. The fact that the confined-phase twisted partition function is well-defined and continuous for any $L \sim N^0$ is a consequence of massive cancellations between bosons and fermions.

Bibliography

[1] Y. t. Huang, D. A. McGady and C. Peng, Phys. Rev. D **87**, no. 8, 085028 (2013) [arXiv:1205.5606 [hep-th]].

[2] Y. t. Huang and D. McGady, Phys. Rev. Lett. **112**, no. 24, 241601 (2014) [arXiv:1307.4065 [hep-th]].

[3] W. M. Chen, Y. t. Huang and D. A. McGady, arXiv:1402.7062 [hep-th].

[4] D. A. McGady and L. Rodina, Phys. Rev. D **90**, no. 8, 084048 (2014) [arXiv:1311.2938 [hep-th]].

[5] D. A. McGady and L. Rodina, arXiv:1408.5125 [hep-th].

[6] G. Basar, A. Cherman, D. A. McGady and M. Yamazaki, arXiv:1406.6329 [hep-th].

[7] G. Basar, A. Cherman, D. A. McGady and M. Yamazaki, arXiv:1408.3120 [hep-th].

[8] G. Basar, A. Cherman and D. A. McGady, arXiv:1409.1617 [hep-th].

[9] L.M. Brown and R.P. Feynman, Phys. Rev. 85 (1952) 231;
G. Passarino and M. Veltman, Nucl. Phys. B160 (1979) 151;
G. 't Hooft and M. Veltman, Nucl. Phys. B 153:365 (1979);
R. G. Stuart, Comp. Phys. Comm. 48:367 (1988);
R. G. Stuart and A. Gongora, Comp. Phys. Comm. 56:337 (1990).

[10] D. B. Melrose, Il Nuovo Cimento 40A:181 (1965);
W. van Neerven and J. A. M. Vermaseren, Phys. Lett. 137B:241 (1984);
G. J. van Oldenborgh and J. A. M. Vermaseren, Z. Phys. C46:425 (1990);
G. J. van Oldenborgh, PhD thesis, University of Amsterdam (1990);
A. Aeppli, PhD thesis, University of Zurich (1992).

[11] Z. Bern, L. Dixon and D. A. Kosower, Phys. Lett. B302:299 (1993); erratum B318:649 (1993); Nucl. Phys. B412:751 (1994)

[12] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, Nucl. Phys. B **425**, 217 (1994) [hep-ph/9403226].

[13] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, Nucl. Phys. B **435**, 59 (1995) [hep-ph/9409265].

[14] W. T. Giele, E. W. N. Glover, Phys. Rev. **D46**, 1980-2010 (1992).

W. T. Giele, E. W. N. Glover, D. A. Kosower, Nucl. Phys. **B403**, 633-670 (1993). [hep-ph/9302225].

Z. Kunszt, A. Signer, Z. Trocsanyi, Nucl. Phys. **B420**, 550-564 (1994). [hep-ph/9401294].

[15] L. Dixon, private communication, 2002.

[16] N. Arkani-Hamed, F. Cachazo and J. Kaplan, JHEP **1009**, 016 (2010) [arXiv:0808.1446 [hep-th]].

[17] H. Elvang, Y. -t. Huang, C. Peng, JHEP **1109**, 031 (2011). [arXiv:1102.4843 [hep-th]].

[18] S. Lal and S. Raju, Phys. Rev. D **81**, 105002 (2010) [arXiv:0910.0930 [hep-th]].

[19] D. Forde, Phys. Rev. D **75**, 125019 (2007) [arXiv:0704.1835 [hep-ph]].

[20] R. Britto, F. Cachazo and B. Feng, Nucl. Phys. B **715**, 499 (2005) [hep-th/0412308];
R. Britto, F. Cachazo, B. Feng and E. Witten, Phys. Rev. Lett. **94**, 181602 (2005) [hep-th/0501052].

[21] F. Cachazo, P. Svrcek and E. Witten, JHEP **0409**, 006 (2004) [hep-th/0403047].

[22] B. Feng, J. Wang, Y. Wang and Z. Zhang, JHEP **1001**, 019 (2010) [arXiv:0911.0301 [hep-th]].

[23] S. J. Parke, and T. R. Taylor, Phys. Rev. Lett. **56**, 2459 (1986)

[24] A. Ferber, Nucl. Phys. B **132**, 55 (1978).

[25] N. Arkani-Hamed, and J. Kaplan, JHEP **0804**, 076 (2008) [arXiv:0801.2385 [hep-th]].

[26] L. J. Dixon, J. M. Henn, J. Plefka and T. Schuster, JHEP **1101**, 035 (2011) [arXiv:1010.3991 [hep-ph]].

[27] L. J. Dixon, In *Boulder 1995, QCD and beyond* 539-582 [hep-ph/9601359].

[28] K. Risager, JHEP **0512**, 003 (2005) [hep-th/0508206].

[29] H. Elvang, D. Z. Freedman and M. Kiermaier, JHEP **0906**, 068 (2009) [arXiv:0811.3624 [hep-th]].

[30] R. Britto, B. Feng and P. Mastrolia, Phys. Rev. D **73**, 105004 (2006) [hep-ph/0602178].

[31] G. 't Hooft and M. J. G. Veltman, Annales Poincare Phys. Theor. A **20**, 69 (1974).

[32] H. Kawai, D. C. Lewellen and S. H. H. Tye, Nucl. Phys. B **269**, 1 (1986).

[33] Z. Bern, J. J. M. Carrasco and H. Johansson, Phys. Rev. Lett. **105**, 061602 (2010) [arXiv:1004.0476 [hep-th]].

[34] Z. Bern, T. Dennen, Y. -t. Huang and M. Kiermaier, Phys. Rev. D **82**, 065003 (2010) [arXiv:1004.0693 [hep-th]].

[35] Z. Bern, J. J. Carrasco, D. Forde, H. Ita and H. Johansson, Phys. Rev. D **77**, 025010 (2008) [arXiv:0707.1035 [hep-th]].

[36] R. Britto, F. Cachazo and B. Feng, Nucl. Phys. B **725**, 275 (2005) [hep-th/0412103].

[37] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, Nucl. Phys. B **435**, 59 (1995) [hep-ph/9409265].

[38] Z. Bern, J. J. M. Carrasco, H. Johansson and R. Roiban, Phys. Rev. Lett. **109**, 241602 (2012) [arXiv:1207.6666 [hep-th]].

[39] Z. Bern, J. J. Carrasco, L. J. Dixon, H. Johansson and R. Roiban, Phys. Rev. Lett. **103**, 081301 (2009) [arXiv:0905.2326 [hep-th]].

[40] Z. Bern, J. J. Carrasco, L. J. Dixon, M. R. Douglas, M. von Hippel and H. Johansson, Phys. Rev. D **87**, 025018 (2013) [arXiv:1210.7709 [hep-th]].

[41] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, S. Caron-Huot and J. Trnka, JHEP **1101**, 041 (2011) [arXiv:1008.2958 [hep-th]].

[42] A. Gustavsson, Nucl. Phys. B **811**, 66 (2009) [arXiv:0709.1260 [hep-th]]; J. Bagger and N. Lambert, Phys. Rev. D **77**, 065008 (2008) [arXiv:0711.0955 [hep-th]].

[43] O. Aharony, O. Bergman, D. L. Jafferis and J. Maldacena, JHEP **0810**, 091 (2008) [arXiv:0806.1218 [hep-th]].

[44] T. Bargheer, F. Loebbert and C. Meneghelli, Phys. Rev. D **82**, 045016 (2010) [arXiv:1003.6120 [hep-th]].

[45] Y. -t. Huang and A. E. Lipstein, JHEP **1010**, 007 (2010) [arXiv:1004.4735 [hep-th]].

[46] M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 117 (1984)

[47] Z. Bern, L. J. Dixon and D. A. Kosower, Nucl. Phys. B **437**, 259 (1995) [hep-ph/9409393].

[48] C. Cheung and D. O'Connell, JHEP **0907**, 075 (2009) [arXiv:0902.0981 [hep-th]].

[49] Z. Bern, J. J. Carrasco, T. Dennen, Y. -t. Huang and H. Ita, Phys. Rev. D **83**, 085022 (2011) [arXiv:1010.0494 [hep-th]].

[50] J. J. M. Carrasco, R. Kallosh, R. Roiban and A. A. Tseytlin, arXiv:1303.6219 [hep-th].

[51] S. R. Coleman and B. Grossman, Nucl. Phys. B **203**, 205 (1982).

[52] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis and R. Rattazzi, JHEP **0610**, 014 (2006) [hep-th/0602178].

[53] Z. Bern and A. G. Morgan, Nucl. Phys. B **467**, 479 (1996) [hep-ph/9511336];
Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, Phys. Lett. B **394**, 105 (1997) [hep-th/9611127];
C. Anastasiou, R. Britto, B. Feng, Z. Kunszt and P. Mastrolia, Phys. Lett. B **645**, 213 (2007) [hep-ph/0609191];
C. Anastasiou, R. Britto, B. Feng, Z. Kunszt and P. Mastrolia, JHEP **0703**, 111 (2007) [hep-ph/0612277];
P. Draggiotis, M. V. Garzelli, C. G. Papadopoulos and R. Pittau, JHEP **0904**, 072 (2009) [arXiv:0903.0356 [hep-ph]].

[54] Z. Bern, L. J. Dixon and D. A. Kosower, Annals Phys. **322**, 1587 (2007) [arXiv:0704.2798 [hep-ph]].

[55] M. B. Green, J. H. Schwarz and P. C. West, Nucl. Phys. B **254**, 327 (1985).

[56] L. Alvarez-Gaume and E. Witten, Nucl. Phys. B **234**, 269 (1984).

[57] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, H. Ita, D. A. Kosower and D. Maitre, Phys. Rev. D **78**, 036003 (2008) [arXiv:0803.4180 [hep-ph]].

[58] Z. Bern, L. J. Dixon and D. A. Kosower, Phys. Rev. D **71**, 105013 (2005) [hep-th/0501240];
Z. Bern, L. J. Dixon and D. A. Kosower, Phys. Rev. D **73**, 065013 (2006) [hep-ph/0507005];
D. Forde and D. A. Kosower, Phys. Rev. D **73**, 065007 (2006) [hep-th/0507292];
D. Forde and D. A. Kosower, Phys. Rev. D **73**, 061701 (2006) [hep-ph/0509358];
C. F. Berger, Z. Bern, L. J. Dixon, D. Forde and D. A. Kosower, Phys. Rev. D **75**, 016006 (2007) [hep-ph/0607014].

[59] Z. Bern, L. J. Dixon, D. A. Kosower, Phys. Rev. D **72**, 125003 (2005). [hep-ph/0505055].

[60] S. Davies, Phys. Rev. D **84**, 094016 (2011) [arXiv:1108.0398 [hep-ph]].

[61] B. Czech, Y. -t. Huang and M. Rozali, JHEP **1210**, 143 (2012) [arXiv:1110.2791 [hep-th]].

[62] B. Zumino, Y. -S. Wu and A. Zee, Nucl. Phys. B **239**, 477 (1984).

[63] J. M. Henn, Phys. Rev. Lett. **110**, no. 25, 251601 (2013) [arXiv:1304.1806 [hep-th]].

[64] S. L. Adler and W. A. Bardeen, Phys. Rev. **182**, 1517 (1969).

[65] S. Caron-Huot, JHEP **1105**, 080 (2011) [arXiv:1007.3224 [hep-ph]].

[66] R. H. Boels, JHEP **1011**, 113 (2010) [arXiv:1008.3101 [hep-th]].

[67] Henrik Johansson, Y-t Huang, work in progress

[68] Z. Bern, S. Davies, T. Dennen, A. V. Smirnov and V. A. Smirnov, Phys. Rev. Lett. **111**, 231302 (2013) [arXiv:1309.2498 [hep-th]].

[69] S. Weinberg, Phys. Rev. **135**, B1049 (1964).

[70] S. Weinberg, Phys. Rev. **138**, B988 (1965).

[71] M.T. Grisaru and H.N. Pendleton, Phys. Lett. B **67**, 323 (1977)

[72] C. Fronsdal, Phys. Rev. D **18**, 3624 (1978)
A. K. H. Bengtsson, I. Bengtsson and L. Brink, Nucl. Phys. B **227**, 31, (1983)
N. Boulanger and S. Leclercq, JHEP **0611**, 034 (2006) [hep-th/0609221]
N. Boulanger, S. Leclercq and P. Sundell, JHEP **0808**, 056 (2008) [arXiv:0805.2764 [hep-th]]
R. R. Metsaev, Nucl. Phys. B **859**, 13 (2012) [arXiv:0712.3526 [hep-th]]
M. Taronna, arXiv:1209.5755 [hep-th].

[73] A. Sagnotti and M. Tsulaia, Nucl. Phys. B **682**, 83 (2004) [hep-th/0311257].
A. Sagnotti and M. Taronna, Nucl. Phys. B **842**, 299 (2011) [arXiv:1006.5242 [hep-th]]
A. Sagnotti, J. Phys. A **46**, 214006 (2013) [arXiv:1112.4285 [hep-th]].

[74] C. Aragone and S. Deser, "Constraints on gravitationally coupled tensor fields," Nuovo Cim. A **3**, 709 (1971).
C. Aragone and S. Deser, Phys. Lett. B **86**, 161 (1979)
M. Porrati, Phys. Rev. D **78**, 065016 (2008) [arXiv:0804.4672 [hep-th]]
M. Porrati, [arXiv:1209.4876 [hep-th]]
M. Henneaux, G. Lucena Gmez and R. Rahman, JHEP **1208**, 093 (2012) [arXiv:1206.1048 [hep-th]]
M. Henneaux, G. Lucena Gmez and R. Rahman, JHEP **1401**, 087 (2014) [arXiv:1310.5152 [hep-th]].

[75] S. Weinberg and E. Witten, Phys. Lett. B **96**, 59 (1980).

[76] J. Maldacena and A. Zhiboedov, J. Phys. A **46**, 214011 (2013) [arXiv:1112.1016 [hep-th]].

[77] V. Alba and K. Diab, arXiv:1307.8092 [hep-th].

[78] P. Benincasa and F. Cachazo, arXiv:0705.4305 [hep-th].

[79] P. Schuster and N. Toro, JHEP **0906**, 079 (2009) [arXiv:0811.3207 [hep-th]].

[80] S. He and H. Zhang, JHEP **1007**, 015 (2010) [arXiv:0811.3210 [hep-th]].

[81] G. 't Hooft, Ed. by G. 't Hooft et al., (Plenum Press, New York NY 1980).

[82] P. Benincasa and E. Conde, Phys. Rev. D **86**, 025007 (2012) [arXiv:1108.3078 [hep-th]].

[83] A. Fotopoulos and M. Tsulaia, JHEP **1011**, 086 (2010) [arXiv:1009.0727 [hep-th]].

[84] P. Dempster and M. Tsulaia, Nucl. Phys. B **865**, 353 (2012) [arXiv:1203.5597 [hep-th]].

[85] D. Francia, J. Mourad and A. Sagnotti, Nucl. Phys. B **773**, 203 (2007) [hep-th/0701163]
A. Sagnotti PoS CORFU **2011**, 106 (2011) [arXiv:1002.3388 [hep-th]].

[86] N. Arkani-Hamed, PITP 2011 lectures ($http://www.sns.ias.edu/pitp$).

[87] C. Aragone and S. Deser, Nucl. Phys. B **170**, 329 (1980)
N. Boulanger, T. Damour, L. Gualtieri and M. Henneaux, Nucl. Phys. B **597**, 127 (2001) [hep-th/0007220].

[88] S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976)
M. T. Grisaru, H. N. Pendleton and P. van Nieuwenhuizen, Phys. Rev. D **15**, 996 (1977).

[89] T. Cohen, H. Elvang, and M. Kiermaier, JHEP **1104**, 053 (2011) [arXiv:1010.0257 [hep-th]].

[90] P. Benincasa, C. Boucher-Veronneau, and F. Cachazo, JHEP**0711**, 057 (2007) [arXiv:hep-th/0702032].

[91] Steven Weinberg, Cambridge University Press, 2005, ISBN-0521670535.

[92] K. Kampf, J. Novotny, and J. Trnka, Phys. Rev. D **87**, 081701 (2013) [arXiv:1212.5224 [hep-th]].

[93] K. Kampf, J. Novotny, and J. Trnka, arXiv:1304.3048 [hep-th].

[94] S. Coleman and J. Mandula, Phys. Rev. **159**, 1251 (1967).

[95] R. Haag, J. T. Lopuszanski, and M. Sohnius, Nucl. Phys. **B88**, 257 (1975).

[96] Z. Bern, J. J. M. Carrasco and H. Johansson, Phys. Rev. D **78**, 085011 (2008) [arXiv:0805.3993 [hep-ph]].

[97] N. Beisert, H. Elvang, D. Z. Freedman, M. Kiermaier, A. Morales and S. Stieberger, Phys. Lett. B **694**, 265 (2010) [arXiv:1009.1643 [hep-th]].

[98] R. H. Boels and R. S. Isermann, JHEP **1306**, 017 (2013) [arXiv:1212.3473].
Z. Bern, L. J. Dixon and R. Roiban, Phys. Lett. B **644**, 265 (2007) [hep-th/0611086].
N. E. J. Bjerrum-Bohr and P. Vanhove, Fortsch. Phys. **56**, 824 (2008) [arXiv:0806.1726 [hep-th]].

[99] N. Arkani-Hamed and J. Trnka, arXiv:1312.2007 [hep-th].

[100] F. Cachazo and P. Svrcek, hep-th/0502160.

[101] N. E. J. Bjerrum-Bohr, D. C. Dunbar, H. Ita, W. B. Perkins and K. Risager, JHEP **0612**, 072 (2006) [hep-th/0610043].

[102] C. Cheung, JHEP **1003**, 098 (2010) [arXiv:0808.0504 [hep-th]].

[103] R. H. Boels and R. S. Isermann, Phys. Rev. D **85**, 021701 (2012) [arXiv:1109.5888 [hep-th]].
R. H. Boels and R. S. Isermann, JHEP **1203**, 051 (2012) [arXiv:1110.4462 [hep-th]].
Y. -J. Du, B. Feng and C. -H. Fu, Phys. Lett. B **706**, 490 (2012) [arXiv:1110.4683 [hep-th]].
Y. -J. Du, B. Feng and C. -H. Fu, JHEP **1203**, 016 (2012) [arXiv:1111.5691 [hep-th]].

[104] B. Feng, R. Huang and Y. Jia, Phys. Lett. B **695**, 350 (2011) [arXiv:1004.3417 [hep-th]]

[105] M. Spradlin, A. Volovich and C. Wen, Phys. Lett. B **674**, 69 (2009) [arXiv:0812.4767 [hep-th]]. S. He, D. Nandan and C. Wen, JHEP **1102**, 005 (2011) [arXiv:1011.4287 [hep-th]].

[106] D. Gang, Y. -t. Huang, E. Koh, S. Lee and A. E. Lipstein, JHEP **1103**, 116 (2011) [arXiv:1012.5032 [hep-th]].

[107] S. Weinberg, Phys. Rev. **140**, B516 (1965). S. Badger, N. E. J. Bjerrum-Bohr and P. Vanhove, JHEP **0902**, 038 (2009) [arXiv:0811.3405 [hep-th]].

[108] D. Nguyen, M. Spradlin, A. Volovich and C. Wen, JHEP **1007**, 045 (2010) [arXiv:0907.2276 [hep-th]].

[109] D. A. McGady, Nucl. Phys. Proc. Suppl. **216**, 254 (2011).

[110] A. Hodges, arXiv:1204.1930 [hep-th].

[111] A. Hodges, Journal of High Energy Physics **1307** (2013) [arXiv:1108.2227 [hep-th]].

[112] Jin-Yu Liu and En Shih, arXiv:1409.1710 [hep-th].

[113] D. E. Kaplan and R. Sundrum, JHEP **0607**, 042 (2006) [hep-th/0505265].

[114] G. 't Hooft and S. Nobbenhuis, Class. Quant. Grav. **23**, 3819 (2006) [gr-qc/0602076].

[115] A. Berkovich, B. M. McCoy and W. P. Orrick, J. Statist. Phys. **83**, 795 (1996) [hep-th/9507072].

[116] J. L. Cardy, Nucl. Phys. B **270**, 186 (1986).

[117] P. Di Francesco, P. Mathieu and D. Senechal, New York, USA: Springer (1997) 890 p

[118] B. Sundborg, Nucl. Phys. B **573**, 349 (2000) [hep-th/9908001].

[119] A. M. Polyakov, Int. J. Mod. Phys. A **17S1**, 119 (2002) [hep-th/0110196].

[120] O. Aharony, J. Marsano, S. Minwalla, K. Papadodimas and M. Van Raamsdonk, Adv. Theor. Math. Phys. **8**, 603 (2004) [hep-th/0310285].

[121] S. El-Showk, Y. Nakayama and S. Rychkov, Nucl. Phys. B **848**, 578 (2011) [arXiv:1101.5385 [hep-th]].

[122] J. Kinney, J. M. Maldacena, S. Minwalla and S. Raju, Commun. Math. Phys. **275**, 209 (2007) [hep-th/0510251].

[123] C. Romelsberger, Nucl. Phys. B **747**, 329 (2006) [hep-th/0510060].

[124] F. A. Dolan and H. Osborn, Nucl. Phys. B **818**, 137 (2009) [arXiv:0801.4947 [hep-th]].

[125] F. Benini, T. Nishioka and M. Yamazaki, Phys. Rev. D **86**, 065015 (2012) [arXiv:1109.0283 [hep-th]].

[126] Y. Imamura and S. Yokoyama, JHEP **1104**, 007 (2011) [arXiv:1101.0557 [hep-th]].

[127] A. Rocha-Caridi, New York, USA: Springer (1985) p. 451–473

[128] B. L. Feigin and D. B. Fuks, Funct. Anal. Appl. **17**, 241 (1983).

[129] L. Onsager, Phys. Rev. **65**, 117 (1944).

[130] M. Yamazaki, JHEP **1205**, 147 (2012) [arXiv:1203.5784 [hep-th]].

[131] Y. Terashima and M. Yamazaki, Phys. Rev. Lett. **109**, 091602 (2012) [arXiv:1203.5792 [hep-th]].

[132] M. Yamazaki, J. Statist. Phys. **154**, 895 (2014) [arXiv:1307.1128 [hep-th]].

[133] , S. Naya, Prog. Theor. Phys. **48**, 407 (1972)

[134] J. Frieman, M. Turner and D. Huterer, Ann. Rev. Astron. Astrophys. **46**, 385 (2008) [arXiv:0803.0982 [astro-ph]].

[135] V. Khachatryan *et al.* [CMS Collaboration], Phys. Lett. B **698**, 196 (2011) [arXiv:1101.1628 [hep-ex]].

[136] G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. Lett. **106**, 131802 (2011) [arXiv:1102.2357 [hep-ex]].

[137] G. W. Moore, Nucl. Phys. B **293**, 139 (1987) [Erratum-ibid. B **299**, 847 (1988)].

[138] K. R. Dienes, Nucl. Phys. B **429**, 533 (1994) [hep-th/9402006].

[139] K. R. Dienes, M. Moshe and R. C. Myers, Phys. Rev. Lett. **74**, 4767 (1995) [hep-th/9503055].

[140] S. Kachru, J. Kumar and E. Silverstein, Phys. Rev. D **59**, 106004 (1999) [hep-th/9807076].

[141] K. R. Dienes, Nucl. Phys. B **611**, 146 (2001) [hep-ph/0104274].

[142] G. 't Hooft, Nucl. Phys. B **72**, 461 (1974).

[143] C. P. Herzog and K. W. Huang, Phys. Rev. D **87**, 081901 (2013) [arXiv:1301.5002 [hep-th]].

[144] S. Giombi, I. R. Klebanov and A. A. Tseytlin, Phys. Rev. D **90**, 024048 (2014) [arXiv:1402.5396 [hep-th]].

[145] M. Beccaria, X. Bekaert and A. A. Tseytlin, JHEP **1408**, 113 (2014) [arXiv:1406.3542 [hep-th]].

[146] L. Di Pietro and Z. Komargodski, arXiv:1407.6061 [hep-th].

[147] M. Unsal, Phys. Rev. D **76**, 025015 (2007) [hep-th/0703025 [HEP-TH]].

[148] G. Basar, A. Cherman, D. Dorigoni and M. nsal, Phys. Rev. Lett. **111**, no. 12, 121601 (2013) [arXiv:1306.2960 [hep-th]].

[149] R. Hagedorn, Nuovo Cim. Suppl. **3**, 147 (1965).

[150] J. Polchinski, Cambridge, UK: Univ. Pr. (1998) 402 p

[151] T. D. Cohen, JHEP **1006**, 098 (2010) [arXiv:0901.0494 [hep-th]].

[152] T. D. Cohen and V. Krejcirik, JHEP **1108**, 138 (2011) [arXiv:1104.4783 [hep-th]].

[153] T. Eguchi and H. Kawai, Phys. Rev. Lett. **48**, 1063 (1982).

[154] G. Bhanot, U. M. Heller and H. Neuberger, Phys. Lett. B **113**, 47 (1982).

[155] R. Narayanan and H. Neuberger, Phys. Rev. Lett. **91**, 081601 (2003) [hep-lat/0303023].

[156] T. D. Cohen, Phys. Rev. Lett. **93**, 201601 (2004) [hep-ph/0407306].

[157] P. Kovtun, M. Unsal and L. G. Yaffe, JHEP **0706**, 019 (2007) [hep-th/0702021 [HEP-TH]].

[158] M. Unsal and L. G. Yaffe, JHEP **1008**, 030 (2010) [arXiv:1006.2101 [hep-th]].

[159] G. Cossu and M. D'Elia, JHEP **0907**, 048 (2009) [arXiv:0904.1353 [hep-lat]].

[160] P. F. Bedaque, M. I. Buchoff, A. Cherman and R. P. Springer, JHEP **0910**, 070 (2009) [arXiv:0904.0277 [hep-th]].

[161] B. Bringoltz, JHEP **0906**, 091 (2009) [arXiv:0905.2406 [hep-lat]].

[162] B. Bringoltz and S. R. Sharpe, Phys. Rev. D **80**, 065031 (2009) [arXiv:0906.3538 [hep-lat]].

[163] A. Hietanen and R. Narayanan, JHEP **1001**, 079 (2010) [arXiv:0911.2449 [hep-lat]].

[164] E. Poppitz and M. Unsal, JHEP **1001**, 098 (2010) [arXiv:0911.0358 [hep-th]].

[165] T. Azeyanagi, M. Hanada, M. Unsal and R. Yacoby, Phys. Rev. D **82**, 125013 (2010) [arXiv:1006.0717 [hep-th]].

[166] E. Poppitz and M. Unsal, Phys. Rev. D **82**, 066002 (2010) [arXiv:1005.3519 [hep-th]].

[167] A. Hietanen and R. Narayanan, Phys. Lett. B **698**, 171 (2011) [arXiv:1011.2150 [hep-lat]].

[168] D. Dorigoni, G. Veneziano and J. Wosiek, JHEP **1106**, 051 (2011) [arXiv:1011.1200 [hep-th]].

[169] S. Catterall, R. Galvez and M. Unsal, JHEP **1008**, 010 (2010) [arXiv:1006.2469 [hep-lat]].

[170] B. Bringoltz, M. Koren and S. R. Sharpe, Phys. Rev. D **85**, 094504 (2012) [arXiv:1106.5538 [hep-lat]].

[171] A. Armoni, D. Dorigoni and G. Veneziano, JHEP **1110**, 086 (2011) [arXiv:1108.6196 [hep-th]].

[172] A. Gonzalez-Arroyo and M. Okawa, PoS LATTICE **2012**, 046 (2012) [arXiv:1210.7881 [hep-lat]].

[173] A. Gonzlez-Arroyo and M. Okawa, Phys. Rev. D **88**, 014514 (2013) [arXiv:1305.6253 [hep-lat]].

[174] A. Gonzlez-Arroyo and M. Okawa, arXiv:1304.0306 [hep-lat].

[175] M. Shifman, private communications, 2012.

[176] T. J. Hollowood and J. C. Myers, JHEP **0911**, 008 (2009) [arXiv:0907.3665 [hep-th]].

[177] M. Unsal, Phys. Rev. Lett. **100**, 032005 (2008) [arXiv:0708.1772 [hep-th]].

[178] M. Unsal, Phys. Rev. D **80**, 065001 (2009) [arXiv:0709.3269 [hep-th]].

[179] M. Unsal, Phys. Rev. Lett. **102**, 182002 (2009) [arXiv:0807.0466 [hep-th]].

[180] H. Nishimura and M. C. Ogilvie, Phys. Rev. D **81**, 014018 (2010) [arXiv:0911.2696 [hep-lat]].

[181] M. M. Anber, E. Poppitz and M. Unsal, JHEP **1204**, 040 (2012) [arXiv:1112.6389 [hep-th]].

[182] T. Misumi and T. Kanazawa, JHEP **1406**, 181 (2014) [arXiv:1405.3113 [hep-ph]].

[183] T. Misumi, M. Nitta and N. Sakai, JHEP **1406**, 164 (2014) [arXiv:1404.7225 [hep-th]].

[184] M. Shifman, Mod. Phys. Lett. A **28**, 1350179 (2013)

[185] E. Witten, Nucl. Phys. B **160**, 57 (1979).

[186] S. Benvenuti, B. Feng, A. Hanany and Y. H. He, JHEP **0711**, 050 (2007) [hep-th/0608050].

[187] B. Feng, A. Hanany and Y. H. He, JHEP **0703**, 090 (2007) [hep-th/0701063].

[188] K. R. Dienes, In *Syracuse 1994, Proceedings, PASCOS '94* 234-243, and Inst. Adv. Stud. Princeton - IASSNS-HEP-94-071 (94/09,rec.Sep.) 13 p. (417474) [hep-th/9409114].

[189] D. Sen, Nucl. Phys. B **284**, 201 (1987).

[190] G. Festuccia and N. Seiberg, JHEP **1106**, 114 (2011) [arXiv:1105.0689 [hep-th]].

[191] T. T. Dumitrescu, G. Festuccia and N. Seiberg, JHEP **1208**, 141 (2012) [arXiv:1205.1115 [hep-th]].

[192] , G. Basar, A. Cherman, Aleksey and D. A. McGady, to appear.

[193] A. Armoni, M. Shifman and G. Veneziano, Nucl. Phys. B **667**, 170 (2003) [hep-th/0302163].

[194] A. Armoni, M. Shifman and G. Veneziano, In *Shifman, M. (ed.) et al.: From fields to strings, vol. 1* 353-444 [hep-th/0403071].