OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **57** (2024) 333002 (179pp)
https://doi.org/10.1088/1751-8121/ad5fd0

Topical Review

The duality between color and kinematics and its applications

Zvi Bern[1,2], John Joseph Carrasco[3,4], Marco Chiodaroli[5], Henrik Johansson[5,6,*] and Radu Roiban[7]

[1] Mani L. Bhaumik Institute for Theoretical Physics, Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095, United States of America
[2] Theoretical Physics Department, CERN, 1211 Geneva 23, Switzerland
[3] Department of Physics and Astronomy Northwestern University, Evanston, IL 60208, United States of America
[4] Institute of Theoretical Physics (IPhT), CEA/CNRS-Saclay and University of Paris-Saclay, F-91191 Gif-sur-Yvette cedex, France
[5] Department of Physics and Astronomy, Uppsala University, 75108 Uppsala, Sweden
[6] Nordita, Stockholm University and KTH Royal Institute of Technology, Roslagstullsbacken 23, 10691 Stockholm, Sweden
[7] Institute for Gravitation and the Cosmos, Pennsylvania State University, University Park, PA 16802, United States of America

E-mail: henrik.johansson@physics.uu.se, bern@physics.ucla.edu, carrasco@northwestern.edu, marco.chiodaroli@physics.uu.se and radu@phys.psu.edu

Received 15 March 2024
Accepted for publication 5 July 2024
Published 8 August 2024

CrossMark

Abstract
This review describes the duality between color and kinematics and its applications, with the aim of gaining a deeper understanding of the perturbative structure of gauge and gravity theories. We emphasize, in particular, applications to loop-level calculations, the broad web of theories linked by the duality and the associated double-copy structure, and the issue of extending the duality and double copy beyond scattering amplitudes. The review is aimed at doctoral

* Author to whom any correspondence should be addressed.

students and junior researchers both inside and outside the field of amplitudes and is accompanied by various exercises.

Keywords: gauge theory, gravity, scattering amplitudes

Contents

1. Introduction

Gauge and gravity theories play a crucial role in our understanding of physical phenomena. Yet, they appear to be distinct. The weak, strong and electromagnetic interactions are manifestations of gauge theories, while gravity shapes the macroscopic evolution of the Universe and spacetime itself. Finding a unified framework which seamlessly combines these two classes of theories constitutes, arguably, the most important open problem in theoretical physics. It is by now clear that realizing this unification requires a departure from conventional approaches through new principles or novel symmetries. The double-copy perspective reviewed here offers a radically different way to interpret gravity. Its relation to the other forces through color/kinematics duality [1, 2] leads to remarkable new insights and powerful computational tools.

Despite their clear differences, gauge and gravity theories are already known to share many features, supporting the existence of an underlying unified framework, such as string theory. While many of these similarities are not apparent from a standard Lagrangian or Hamiltonian standpoint, the study of objects closely related to observable quantities, such as scattering amplitudes, reveals deep and highly-nontrivial connections. This is most apparent in their perturbative expansions, which make it clear that the dynamics of these two classes of theories are governed by the same kinematical building blocks, even when their physical properties are strikingly different.

The developments which exposed these features were systematized by the introduction of the duality between color and kinematics and of the double-copy construction. The scattering amplitudes of many perturbative quantum field theories (QFTs) exhibit a double-copy structure. It is central to our ability to carry out calculations to very high loop orders and a property of all supergravities whose amplitudes have been analyzed in detail. This leads to the natural question whether all (super)gravity theories are double copies of suitably-chosen matter-coupled gauge theories. Perhaps more importantly, the double copy realizes a unification of gauge and gravity theories in the sense of providing a framework where calculations in both theories can be carried out using an identical set of building blocks, yielding vast simplifications.

The primary purpose of this review is to offer an introduction to the duality between color and kinematics—also referred to as color/kinematics (CK) duality and Bern–Carrasco–Johansson (BCJ) duality—and the associated double-copy relation in the hope of stimulating new progress both inside and outside of the fairly well-understood setting of scattering amplitudes. Beyond gauge and gravity theories, double-copy relations also provide a new perspective on QFT, generating a surprisingly wide web of theories through building blocks obeying the same algebraic relations.

The duality essentially states that scattering amplitudes in gauge theories—and, more generally, in theories with some Lie-algebra symmetry—can be rearranged so that kinematic building blocks obey the same generic algebraic relations as their color factors. Via the duality, we can not only constrain the kinematic dependence of each graph, but we can also convert gauge-theory scattering amplitudes to gravity ones through the simple replacement

$$\text{color} \Rightarrow \text{kinematics}\,. \tag{1}$$

Evidence provided by explicit calculations suggests that CK duality and the double-copy construction hold for a wide class of theories at loop level [2–23]. Formal proofs, using a variety of methods [24–29], have been constructed for only tree-level scattering amplitudes in these theories. The duality also gave novel descriptions for tree-level amplitudes in bosonic and supersymmetric string theories, as well as in various effective field theories related to spontaneous symmetry breaking, and more. It has also been observed that, in the presence of adjoint-representation fermions, the duality implies supersymmetry [30].

The schematic rule (1) has served as a powerful guide for many studies in perturbative gravity and supergravity, especially on their loop-level ultraviolet (UV) properties (see e.g. [2, 6, 31–40]), showing a surprisingly tame behavior. For many supergravity theories, the physical degrees of freedom are obtained by the substitution (1). In others, such as pure Einstein gravity, the desired spectrum can only be obtained after a subset of the double-copy states are projected out. As we describe in some detail in section 5, CK duality and the associated double-copy properties hold for a remarkably large web of theories.

Given the success at exploiting the double-copy structure for scattering amplitudes, it is natural to wonder whether it also carries over to other areas of gravitational physics, especially for understanding and simplifying generic classical solutions. Scattering amplitudes have an important property that makes transparent the duality and double-copy structure: they are independent of the choice of gauge and field-variables. Generic classical solutions, on the other hand, do depend on these choices, making the problem of relating gauge and gravity classical solutions inherently more involved. Nevertheless, the prospect of solving problems in gravity by recycling gauge-theory solutions is especially alluring. While the differences with scattering amplitudes are significant and make it a nontrivial challenge to implement this program, there has been significant progress in unraveling both the underlying principles of CK duality [26, 41–49] and finding explicit examples of classical solutions related by the double-copy property [50–77]. One of the most promising applications of the double copy beyond scattering amplitudes relates to gravitational-wave physics, as highlighted by [57, 69, 78–82].

The origins of the double copy can be traced back to the dawn of string theory, with the observation of a curious connection between the Veneziano scattering amplitude [83], $A(s,t)$, (later identified as an open-string scattering amplitude) and the Virasoro–Shapiro amplitude [84, 85], $M(s,t,u)$, (later identified as a closed-string amplitude). With an appropriate normalization, these two amplitudes are related as [86]

$$M(s,t,u) = \frac{\sin(\pi\alpha' s)}{\pi\alpha'} A(s,t)A(s,u)\,, \tag{2}$$

where α' is the inverse string tension. The arguments are the kinematic (Mandelstam) invariants of a four-point scattering process,

$$s = (p_1+p_2)^2\,, \quad t = (p_2+p_3)^2\,, \quad u = (p_1+p_3)^2\,. \tag{3}$$

Equation (2) carries over to all string states, including the gluons of the open string and the gravitons in the closed string. In the low-energy limit, when string theory reduces to field theory, it yields a relation between scattering amplitudes in Einstein gravity and those of Yang–Mills (YM) theory [87],

$$\mathcal{M}_4^{\text{tree}}(1,2,3,4) = \left(\frac{\kappa}{2}\right)^2 s A_4^{\text{tree}}(1,2,3,4) A_4^{\text{tree}}(1,2,4,3)\,, \tag{4}$$

where $A_4^{\text{tree}}(1,2,3,4)$ is a color-ordered gauge-theory four-gluon partial scattering amplitude, $\mathcal{M}_4^{\text{tree}}(1,2,3,4)$ is a four-graviton tree amplitude and κ is the gravitational coupling to related to Newton's constant via $\kappa^2 = 32\pi G_N$ and, for reasons that will become clear shortly, the polarization vectors of gluons on the right-hand side of equation (4) are taken to be null. We will typically suppress the gravitational coupling by setting $\kappa = 2$ in this review; however, occasionally we display it for clarity. The color-ordered partial tree amplitudes are the coefficients of basis elements once the amplitude's color factors are expressed in the trace color basis, and the coupling g is set to unity. They are gauge invariant—see section 2 and e.g. [88–92] for further details. Equation (4) is rather striking, asserting that tree-level four-graviton scattering is described completely by gauge-theory four-gluon scattering, bypassing the usual machinery of general relativity. Similar relations were later derived for higher-point string-theory tree-level amplitudes [86], and generalized in the field-theory limit to an arbitrary number of external particles [93]. Besides the remarkable implication that the detailed dynamics of the gravitational field can be described in terms of the dynamics of gauge fields, equation (4) has other surprising features not visible in standard Lagrangian formulations. For example, equation (4) implies that the four-graviton amplitude can be re-arranged so that Lorentz indices factorize [94, 95] into 'left' indices belonging to one gauge-theory amplitude and 'right' indices belonging to another gauge theory.

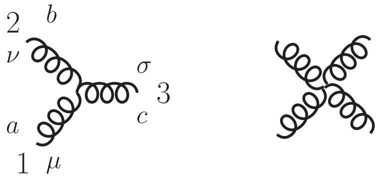


Figure 1. Gauge theories have three- and four-point vertices in a Feynman diagrammatic description.

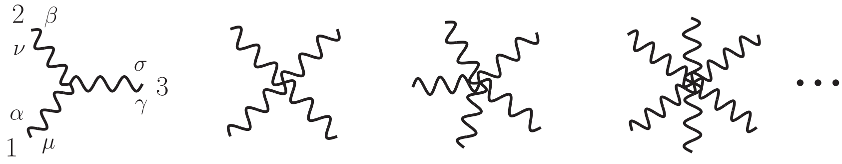


Figure 2. Gravity theories have an infinite number of higher-point contact interactions in a Feynman diagrammatic description.

1.1. Motivation: complexity of gravity versus gauge theory

It is interesting to contrast the remarkable simplicity encoded in the relation (4) with the much more complicated expressions that arise from standard Lagrangian methods. Scattering amplitudes for gauge and gravity theories can be obtained using the Feynman rules derived from their respective Lagrangians

$$\mathcal{L}_{\text{YM}} = -\frac{1}{4} F^a{}_{\mu\nu} F^{a\,\mu\nu}\,, \qquad\qquad \mathcal{L}_{\text{EH}} = \frac{2}{\kappa^2}\sqrt{-g}R\,. \tag{5}$$

Here $F^a_{\mu\nu}$ is the usual YM field strength and R the Ricci scalar.

Following standard Feynman-diagrammatic methods, we gauge-fix and then extract the propagator(s) and the three- and higher-point vertices. For gravity we also expand around flat spacetime, taking the metric to be $g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}$ where $\eta_{\mu\nu}$ is the Minkowski metric and $h_{\mu\nu}$ is the graviton field. As illustrated in figures 1 and 2, with standard gauge choices, gauge theory has only three- and four-point vertices, while gravity has an infinite number of vertices of arbitrary multiplicity. The complexity of each individual interaction term is perhaps more striking than their infinite number. Consider, for example, the three-graviton interaction. In the standard de Donder gauge, $\partial_\nu h^\nu{}_\mu = \frac{1}{2}\partial_\mu h^\nu{}_\nu$, the corresponding vertex is [96, 97],

$$\begin{aligned}
&G_{3\,\mu\rho,\nu\lambda,\sigma\tau}(p_1,p_2,p_3)\\
&\quad = i\,\text{Sym}\Big[-\frac{1}{2}P_3\left(p_1\cdot p_2\eta_{\mu\rho}\eta_{\nu\lambda}\eta_{\sigma\tau}\right) - \frac{1}{2}P_6\left(p_{1\nu}p_{1\lambda}\eta_{\mu\rho}\eta_{\sigma\tau}\right) + \frac{1}{2}P_3\left(p_1\cdot p_2\eta_{\mu\nu}\eta_{\rho\lambda}\eta_{\sigma\tau}\right)\\
&\qquad + P_6\left(p_1\cdot p_2\eta_{\mu\rho}\eta_{\nu\sigma}\eta_{\lambda\tau}\right) + 2P_3\left(p_{1\nu}p_{1\tau}\eta_{\mu\rho}\eta_{\lambda\sigma}\right) - P_3\left(p_{1\lambda}p_{2\mu}\eta_{\rho\nu}\eta_{\sigma\tau}\right)\\
&\qquad + P_3\left(p_{1\sigma}p_{2\tau}\eta_{\mu\nu}\eta_{\rho\lambda}\right) + P_6\left(p_{1\sigma}p_{1\tau}\eta_{\mu\nu}\eta_{\rho\lambda}\right) + 2P_6\left(p_{1\nu}p_{2\tau}\eta_{\lambda\mu}\eta_{\rho\sigma}\right)\\
&\qquad + 2P_3\left(p_{1\nu}p_{2\mu}\eta_{\lambda\sigma}\eta_{\tau\rho}\right) - 2P_3\left(p_1\cdot p_2\eta_{\rho\nu}\eta_{\lambda\sigma}\eta_{\tau\mu}\right)\Big]\,,
\end{aligned} \tag{6}$$

where we set $\kappa = 2$, p_i are the momenta of the three gravitons, $\eta_{\mu\nu}$ is the flat metric, 'Sym' implies a symmetrization in each pair of graviton Lorentz indices $\mu \leftrightarrow \rho$, $\nu \leftrightarrow \lambda$ and $\sigma \leftrightarrow \tau$, and P_3 and P_6 signify a symmetrization over the three graviton legs, generating three or six

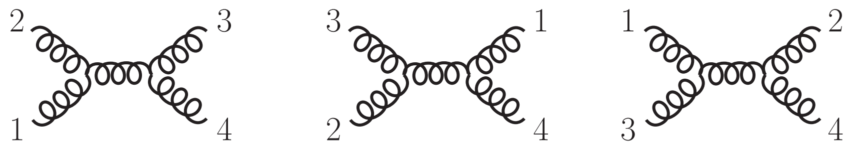


Figure 3. The three Feynman diagrams corresponding to the s, t and u channels.

terms respectively. The symmetrization over the three external legs ensures the Bose symmetry of the vertex. In total, the vertex has of the order of 100 terms. This generally undercounts the number of terms, because within a diagram each vertex momentum is a linear combination of the independent momenta of that diagram.

We may contrast this to the three-gluon vertex in Feynman gauge,

$$V^{abc}_{3\,\mu\nu\sigma}(p_1,p_2,p_3) = g f^{abc}\left[(p_1-p_2)_\sigma\,\eta_{\mu\nu} + \text{cyclic}\right]. \tag{7}$$

which does not appear to bear any obvious relation to the corresponding three-graviton vertex (6). These considerations seemingly suggest that gravity is much more complicated than gauge theory. Moreover, the three-graviton vertex immediately appears to conflict with the simple factorization of Lorentz indices into left and right sets visible in equation (4). The first term in equation (6), for example, contains a factor $\eta_{\mu\rho}$ which explicitly contracts a left graviton index with a right one.

The reason why the three-graviton vertex is so complicated is that it is gauge-dependent[8]. With special gauge choices and appropriate field redefinitions [94, 95, 98, 99], it is possible to considerably simplifying the Feynman rules. Still, direct perturbative gravity calculations in a Feynman diagram approach are rather nontrivial, especially beyond leading order, even with modern computers. To eliminate the gauge dependence we should instead focus on the three-vertex with on-shell conditions imposed on external legs, by demanding that the vertex is contracted into physical states that satisfy,

$$\varepsilon^{\mu\rho} = \varepsilon^{\rho\mu}, \qquad p_\mu\varepsilon^{\mu\rho} = 0, \qquad p_\rho\varepsilon^{\mu\rho} = 0, \qquad \varepsilon_\mu{}^\mu \equiv \eta^{\mu\nu}\varepsilon_{\mu\nu} = 0, \tag{8}$$

where p is a graviton momentum and $\varepsilon^{\mu\nu}$ the associated graviton polarization tensor. This removes all trace and longitudinal terms, reducing the vertex to a simple form,

$$G_{3\,\mu\rho,\nu\lambda,\sigma\tau}(p_1,p_2,p_3) = -i\left[(p_1-p_2)_\sigma\,\eta_{\mu\nu} + \text{cyclic}\right]\left[(p_1-p_2)_\tau\,\eta_{\rho\lambda} + \text{cyclic}\right], \tag{9}$$

exposing its simple relation to the three-gluon vertex of gauge theory. This is a hint that there should be much better ways to organize the perturbative expansion of gravity. We now turn to four-graviton scattering amplitude, which is a better example as it corresponds directly to a physical process.

1.2. Invitation: four-point example

Consider the full four-gluon tree amplitude in YM theory, which can be obtained, for example, by following textbook Feynman rules [100]. We write it as a sum over three channels corresponding to the three diagrams in figure 3

$$i\mathcal{A}_4^{\text{tree}} = g^2\left(\frac{n_s c_s}{s} + \frac{n_t c_t}{t} + \frac{n_u c_u}{u}\right), \tag{10}$$

[8] While somewhat less complicated than the three-graviton vertex, the three-gluon vertex is also gauge-dependent.

where the Mandelstam variables are defined in equation (3). The s-channel color factor, normalized to be compatible with the scattering amplitudes literature [88], is

$$c_s = -2f^{a_1a_2b}f^{ba_3a_4}\,, \tag{11}$$

where the color-group structure constants f_{abc} are the standard textbook ones [100]. With this normalization, the s-channel kinematic numerator, n_s, is

$$\begin{aligned} n_s = -\frac{1}{2}\{&[(\varepsilon_1\cdot\varepsilon_2)\,p_1^\mu + 2\,(\varepsilon_1\cdot p_2)\,\varepsilon_2^\mu - (1\leftrightarrow 2)]\,[(\varepsilon_3\cdot\varepsilon_4)\,p_{3\mu} + 2\,(\varepsilon_3\cdot p_4)\,\varepsilon_{4\mu} - (3\leftrightarrow 4)] \\ &+s\,[(\varepsilon_1\cdot\varepsilon_3)\,(\varepsilon_2\cdot\varepsilon_4) - (\varepsilon_1\cdot\varepsilon_4)\,(\varepsilon_2\cdot\varepsilon_3)]\}\,, \end{aligned} \tag{12}$$

where the momenta and polarization vectors satisfy on-shell conditions $p_i^2 = \varepsilon_i\cdot p_i = 0$. The other color factors and numerators are obtained by cyclic permutations of the particle labels $(1,2,3)$:

$$c_t n_t = c_s n_s\big|_{1\to2\to3\to1}\,, \qquad\qquad c_u n_u = c_s n_s\big|_{1\to3\to2\to1}\,. \tag{13}$$

Feynman rules for gluons contain a four-gluon vertex, as in figure 1. Here we have absorbed its contribution into the three diagrams in figure 3 according to the color factors, by multiplying and dividing by an appropriate propagator. This is the origin of the term on the second line of equation (12).

A key property of the gauge-theory scattering amplitude (10) is its linearized gauge invariance. To check this, we need to verify that the amplitude vanishes with the replacement $\varepsilon_4 \to p_4$. Upon doing this replacement for the s-channel numerator we get, after some algebra, the nonzero result

$$n_s\big|_{\varepsilon_4\to p_4} = -\frac{s}{2}\,[(\varepsilon_1\cdot\varepsilon_2)\,((\varepsilon_3\cdot p_2) - (\varepsilon_3\cdot p_1)) + \text{cyclic}\,(1,2,3)] \equiv s\,\alpha\,(\varepsilon,p)\,, \tag{14}$$

which is no surprise since individual diagrams are, in general, gauge dependent. The function $\alpha(\varepsilon,p)$ is clearly invariant under cyclic permutations of the labels $(1,2,3)$. For the full amplitude we get therefore

$$\frac{n_s c_s}{s} + \frac{n_t c_t}{t} + \frac{n_u c_u}{u}\bigg|_{\varepsilon_4\to p_4} = (c_s + c_t + c_u)\,\alpha\,(\varepsilon,p)\,, \tag{15}$$

where $\alpha(\varepsilon,p)$ is the expression in equation (14). Hence the amplitude is gauge invariant if $c_s + c_t + c_u$ vanish, i.e.

$$c_s + c_t + c_u = -2\left(f^{a_1a_2b}f^{ba_3a_4} + f^{a_2a_3b}f^{ba_1a_4} + f^{a_3a_1b}f^{ba_2a_4}\right) = 0\,. \tag{16}$$

This is the standard Jacobi identity, which indeed is satisfied by the group-theory structure constants in a gauge theory.

Consider the three-term sum over kinematic numerators in equations (12) and (13), $n_s + n_t + n_u$, analogous to the sum over color factors on the left-hand side of equation (16). Remarkably, this combination vanishes when the on-shell conditions are applied,

$$n_s + n_t + n_u = 0\,. \tag{17}$$

We will refer to this relation as a *kinematic Jacobi identity*. This was originally noticed some time ago for four-point amplitudes, as a curiosity related to radiation zeros in four-point amplitudes [101–103]. Generic representations of four-point amplitudes in terms of diagrams with only cubic vertices obey these identities, but at higher points nontrivial rearrangements are needed. The significance of the identity equation (17) and its generality was understood later [1, 2]. We refer to kinematic identities that are analogous to generic color-factor identities as a duality between color and kinematics. It turns out that they constitute an ubiquitous, yet hidden, structure not only of gauge theories, but also of an ever-increasing web of theories, as described in section 5.

Exercise 1.1. Use equations (12) and (13) to verify the numerator Jacobi identity (17). Redefine the numerators by eliminating c_u in favor of c_s and c_t, defining new numerators n_s' and n_t' as the coefficient of c_s/s and c_t/t. The numerator n_u' vanishes by construction. Show that the kinematic Jacobi identity still holds for these redefined numerators.

The fact that the kinematic factors satisfy the same relations as the color factors suggests that they are mutually exchangeable. Indeed, we can swap color factors for kinematic factors in the YM four-point amplitude (10), which gives a new gauge-invariant object that, as we will discuss momentarily, is a four-graviton amplitude,

$$i\mathcal{A}_4^{\text{tree}}\Big|_{\substack{c_i\to\tilde{n}_i\\ g\to\kappa/2}} \equiv i\mathcal{M}_4^{\text{tree}} = \left(\frac{\kappa}{2}\right)^2\left(\frac{n_s^2}{s}+\frac{n_t^2}{t}+\frac{n_u^2}{u}\right). \tag{18}$$

The new amplitude $\mathcal{M}_4^{\text{tree}}$ doubles up the kinematic numerators, and so we refer to it as a double copy. (The i in front of the $\mathcal{M}_4^{\text{tree}}$ is a phase convention.) The expression in equation (18) has the following properties: the external states are captured by symmetric polarization tensors $\varepsilon^{\mu\nu}=\varepsilon^{\mu}\varepsilon^{\nu}$, the interactions are of the two-derivative type, and the amplitude is invariant under linearized diffeomorphism transformations. By choosing the polarization vectors to be null $\varepsilon^2=0$ (corresponding to circular polarization), implying that $\varepsilon^{\mu\nu}$ is traceless, this amplitude should describe the scattering of four gravitons in Einstein's general relativity, up to an overall normalization. There are a number of ways to prove that this is the case, including using on-shell recursion relations [41] and ordinary gravity Feynman rules [94]; here we will show that equation (18) reproduces the Kawai–Lewellen–Tye (KLT) form of gravity amplitudes [86], derived using the low-energy limit of string theory.

The diffeomorphism invariance of the amplitude requires some elaboration. Consider a linearized diffeomorphism of the asymptotic (weak) graviton field $h_{\mu\nu}$. The diffeomorphism is parametrized by the function ξ_μ and take the simple form

$$\delta h_{\mu\nu}=\partial_\mu\xi_\nu+\partial_\nu\xi_\mu\,. \tag{19}$$

Translating this to momentum space implies that a diffeomorphism-invariant amplitude should vanish upon replacing a polarization tensor as: $\varepsilon^{\mu\nu}\to p^\mu\varepsilon^\nu+p^\nu\varepsilon^\mu$. Applying this to leg 4 of the amplitude, we find

$$\frac{n_s^2}{s}+\frac{n_t^2}{t}+\frac{n_u^2}{u}\Big|_{\varepsilon_4^{\mu\nu}\to p_4^\mu\varepsilon_4^\nu+p_4^\nu\varepsilon_4^\mu}=2\left(n_s+n_t+n_u\right)\alpha\left(\varepsilon,p\right)=0\,. \tag{20}$$

Thus, we see that the kinematic Jacobi identity needs to be satisfied for the amplitude to be invariant under linearized diffeomorphism transformations, in complete analogy to the color Jacobi identity in the gauge-theory amplitude.

Returning to the YM amplitude, we note that the amplitude can be written in a manifestly gauge-invariant form if we solve the Jacobi relation by choosing $c_t=-c_u-c_s$,

$$\begin{aligned} i\mathcal{A}_4^{\text{tree}} &= g^2\left(\frac{n_sc_s}{s}+\frac{n_tc_t}{t}+\frac{n_uc_u}{u}\right)\\ &= g^2\left(\left(\frac{n_s}{s}-\frac{n_t}{t}\right)c_s-\left(\frac{n_t}{t}-\frac{n_u}{u}\right)c_u\right)\\ &\equiv i\,g^2A_4^{\text{tree}}\left(1,2,3,4\right)c_s-i\,g^2A_4^{\text{tree}}\left(1,3,2,4\right)c_u\,. \end{aligned} \tag{21}$$

The partial amplitudes $A_4^{\text{tree}}(1,2,3,4)$, which we define to be stripped of color and the coupling, are gauge invariant because the color-dressed amplitude $\mathcal{A}_4^{\text{tree}}$ is now decomposed in a basis of independent color factors, with elements c_s and c_u, and thus the gauge invariance of $\mathcal{A}_4^{\text{tree}}$ implies the gauge invariance of the individual terms of this decomposition.

It is not difficult to show that the partial amplitude can be written as

$$A_4^{\text{tree}}(1,2,3,4) = -i\frac{t_8F^4}{st}\,, \tag{22}$$

where

$$t_8F^4 \equiv 4\text{Tr}(F_1F_2F_3F_4) - \text{Tr}(F_1F_2)\text{Tr}(F_3F_4) + \text{cyclic}(1,2,3) \tag{23}$$

contains various Lorentz traces over four linearized Fourier transformed field strengths,

$$F_i^{\mu\nu} \equiv p_i^\mu \varepsilon_i^\nu - \varepsilon_i^\mu p_i^\nu\,, \tag{24}$$

where the fields are replaced with polarization vectors. These are manifestly invariant under linearized gauge transformations.

We can also solve the kinematic Jacobi relation (17) by choosing $n_t = -n_u - n_s$. The partial amplitudes then become

$$\begin{aligned} iA_4^{\text{tree}}(1,2,3,4) &= \frac{n_s}{s} - \frac{n_t}{t} = n_s\left(\frac{1}{s}+\frac{1}{t}\right) + \frac{n_u}{t}\,,\\ iA_4^{\text{tree}}(1,3,2,4) &= \frac{n_t}{t} - \frac{n_u}{u} = -n_u\left(\frac{1}{u}+\frac{1}{t}\right) - \frac{n_s}{t}\,, \end{aligned} \tag{25}$$

which may also be organized as a matrix relation

$$i\begin{pmatrix} A_4^{\text{tree}}(1,2,3,4) \\ A_4^{\text{tree}}(1,3,2,4)\end{pmatrix} = \begin{pmatrix} \frac{1}{s}+\frac{1}{t} & \frac{1}{t} \\ -\frac{1}{t} & -\frac{1}{u}-\frac{1}{t}\end{pmatrix}\begin{pmatrix} n_s \\ n_u\end{pmatrix}. \tag{26}$$

It might seem that it is possible to solve for the numerators in terms of the partial amplitudes by inverting the two-by-two matrix of propagators. Existence of a solution would contradict, however, the fact that on the one hand numerators are gauge-dependent and on the other partial amplitudes are gauge-invariant. Indeed, the matrix of propagators has no inverse as its determinant is proportional to $s+t+u=0$. At best, we can solve for one of the numerators, say, n_u,

$$n_u = itA_4^{\text{tree}}(1,2,3,4) + u\frac{n_s}{s}\,. \tag{27}$$

Replacing this into $A_4^{\text{tree}}(1,3,2,4)$ in equation (25), the dependence on the undetermined kinematic numerator n_s cancels out, and we obtain the gauge-invariant relation

$$A_4^{\text{tree}}(1,3,2,4) = \frac{s}{u}A_4^{\text{tree}}(1,2,3,4)\,. \tag{28}$$

Given the vanishing of the determinant of the above matrix of propagators, it is not surprising to find that the two partial amplitudes are linearly dependent. In fact, one may phrase equation (28) as the orthogonality condition of the left-hand side of equation (26) onto the null eigenvector of the matrix of propagators.

The existence of relations between partial amplitudes is a general feature. Such *BCJ amplitude relations* exist whenever the duality between color and kinematics and gauge invariance conspire to prevent the relation between partial amplitudes and numerators to be inverted. These relations have been demonstrated in a variety of ways, including using both string theory [104–112] and field theory methods [113–118].

In string theory, one finds similar identities that follow from world-sheet monodromy relations. For massless vector amplitudes of the open string, from world-sheet monodromy relations [104, 105] one finds

$$A_4^{\text{tree}}(1,3,2,4) = \frac{\sin(\pi\alpha' s)}{\sin(\pi\alpha' u)} A_4^{\text{tree}}(1,2,3,4)\,. \tag{29}$$

where α' is the inverse string tension.

We can also use the two relations, $n_t = -n_u - n_s$ and $n_u = tA_4^{\text{tree}}(1,2,3,4) + u\,n_s/s$, in equation (18). The result is

$$\mathcal{M}_4^{\text{tree}}(1,2,3,4) = -i\left[\frac{n_s^2}{s} + \frac{n_t^2}{t} + \frac{n_u^2}{u}\right] = -i\frac{st}{u}\left[A_4^{\text{tree}}(1,2,3,4)\right]^2, \tag{30}$$

where, as usual, we have suppressed the gravitational coupling setting $\kappa = 2$. As for the gauge-theory case, n_s drops out; as in that case, this is to be expected as it would otherwise lead to a relation between gauge invariant and gauge-dependent quantities, $\mathcal{M}_4^{\text{tree}}$ and n_s respectively. We can put this equation into a more standard form using a relabeling identity (28),

$$\mathcal{M}_4^{\text{tree}}(1,2,3,4) = -i\,s A_4^{\text{tree}}(1,2,3,4) A_4^{\text{tree}}(1,2,4,3)\,, \tag{31}$$

which is the simplest of the *KLT relations* between gravity and gauge-theory amplitudes. We derived it here as a consequence of CK duality and gauge-invariance constraints, but the original derivation [86] comes from string theory. It is worth noting that these relations are not unique given amplitude relations such as equation (28).

Replacing the four-point YM amplitude in the from equation (22) into the KLT relation (31), we obtain an explicit form for the four-graviton amplitude

$$\mathcal{M}_4^{\text{tree}}(1,2,3,4) = -i\,\frac{t_{16}R^4}{stu}\,, \tag{32}$$

where we define $t_{16}R^4$ in terms of t_8F^4 in equation (23) as

$$t_{16}R^4 \equiv \left(t_8F^4\right)^2\,. \tag{33}$$

As the notation suggest, $t_{16}R^4$ can also be written as a contraction between a rank-16 tensor t_{16} and four linearized Riemann tensors, using the relationship to linearized gauge-theory field strengths in equation (24),

$$R_i^{\mu\nu\rho\sigma} = F_i^{\mu\nu}F_i^{\rho\sigma} = (p_i^\mu\,\varepsilon_i^\nu - \varepsilon_i^\mu p_i^\nu)(p_i^\rho\,\varepsilon_i^\sigma - \varepsilon_i^\rho p_i^\sigma)\,. \tag{34}$$

1.3. Outline of topics

In this review, we will describe the duality between color and kinematics and the double copy, as proposed in the original work [1, 2], and later refined through various extensions and applications. As indicated in figure 4, CK duality and double copy are intertwined with the topics of several vigorous research fields. The areas that the review will mainly focus on include the web of theories, loop amplitudes and the classical double copy. The web of theories allude to the large classes of known double-copy constructions and their underlying single-copy theories, whose existence became clear after important theories, such as Chern–Simons [119], Yang–Mills–Einstein (YME) [120], Maxwell-Einstein [120, 121], spontaneously-broken theories [122] and gauged supergravities [123] were observed to fit into the general framework. In addition, from the Cachazo, He and Yuan (CHY) formulation [124], it was observed [125] that also effective field theories such as the non-linear-sigma-model (NLSM) [126], (Dirac)-Born-Infeld (DBI) and special-Galileon theory played a central role.

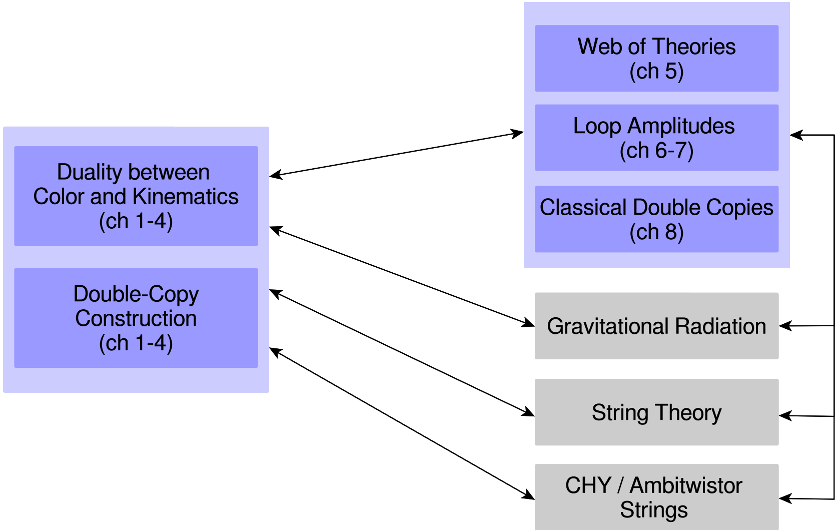


Figure 4. Connections of CK duality to various topics. This review will discuss in some detail the connection of CK duality to the topics in the upper right (with the main chapters indicated) and less so to the topics on the lower right. The various topics are intertwined with each other as well.

The usefulness of the duality and the double copy for loop amplitudes became clear once the framework was applied to obtain compact integrands for the three- [2] and four-loop [6] amplitudes in $\mathcal{N} = 4$ SYM and in $\mathcal{N} = 8$ supergravity. By now it is clear that loop amplitudes in many other theories can be obtained using the duality and double copy.

When the double copy was shown to be applicable to problems of classical gravity, such as the Schwarzschild and Kerr metrics [51] as well as other perturbatively constructable metrics [58], it opened the door to further applications relevant to gravitational physics. With the discovery of gravitational waves from merging binary black holes and neutron stars [127, 128], it is becoming increasingly important to find better ways to accurately calculate classical observables in general relativity. The double-copy approach is still in its infancy, but it bears the promise of drastically changing the way we think of carrying out computations in gravity.

In order to keep the discussion manageable, we will not discuss in much detail the challenges of understanding gravitational radiation and potentials (see e.g. [129–132] for reviews and [78, 80, 82] for a state-of-the-art application of the double copy). Nor will we be thorough in describing the connections to string theory (see e.g. [104, 105, 133, 134]), the CHY construction [124, 135–137] and ambitwistor strings [138–148], all of which have interesting connections to CK duality and the double-copy construction.

The outline of topics in each section is as follows: in section 2, we describe the duality in some detail and give various examples, and show how the double copy implies diffeomorphism invariance of gravity. In section 3, we give a way to visualize how the duality can be thought of as specifying amplitudes in terms of boundary data on a graph of graphs and on making use of relabeling invariance. Then, in section 4, we discuss the inheritance of symmetries in the double-copy theories from their component theories. Section 5 gives a detailed description of the web of double-copy constructible theories, emphasizing the widespread applicability of these ideas. In section 6, we give loop-level examples of the duality between color and kinematics. In section 7, we explain a generalized double-copy procedure that does not require loop integrands to manifest the duality. Section 8 discusses the important issue of extending the double-copy procedure to solutions of the classical Equations of motion. Conclusions and prospects for the future are given in section 9. In appendix A, we collect acronyms and notation used throughout the review. Appendix B summarizes spinor helicity and on-shell supersymmetry, which will be useful in various sections. Finally, appendix C briefly describes generalized unitarity, used in sections 6 and 7.

2. The duality between color and kinematics

The duality between color and kinematics is by now an extensive topic with a variety of perspectives and applications. However, it is not always clear from the literature what rules govern this framework. In this section, the central aspects of CK duality will be described, with the aim of clarifying the reason for imposing various requirements as well as providing an understanding of when they can be relaxed.

2.1. What is the duality between color and kinematics?

CK duality in its original formulation states that it is possible to reorganize the perturbative expansion of tree-level amplitudes in D-dimensional pure YM theory with a general gauge group G in terms of cubic diagrams where the kinematic numerators obey the same Jacobi relations and symmetry properties as their color factors [1, 2]. While it is not *a priori* obvious why such a reorganization is possible or even desirable, from a Lagrangian perspective this is a highly nontrivial statement about YM theory. The associated double-copy construction however, does make it clear that the duality is worth understanding because of the way it connects gravity to gauge theory. While there are tree-level proofs of the duality from the amplitudes perspective [24, 25, 149], at present, only a partial Lagrangian-level understanding has been achieved [41, 42, 150, 151].

More generally, CK duality refers to the statement that in many gauge theories, extending well beyond YM theories with or without matter, it should be possible to reorganize the perturbative expansion so that there is a one-to-one map between the Lie-algebra identities of the color factors carried by certain diagrams (with cubic or higher-point vertices) and the identities of the kinematic numerators of the same diagrams. In the broad class of general gauge theories, one can think of CK duality as a constraint that can be imposed on fields, gauge-group representations, interactions and operators, such that the theories give amplitudes that exhibit the duality structure. These constraints often result in theories with properties that are interesting for reasons not directly related to the duality [120–122, 152, 153].

In generalizing beyond gauge theories, one can consider matter theories that are comprised of spin <1 states that transform nontrivially under a semi-simple global group. In this case, CK duality refers to the one-to-one map between the Lie-algebra relations of this global group and the relations satisfied by the corresponding kinematic numerators of the diagrams. It is convenient to still refer to the global group as the color group since such theories can often be regarded as the matter sector of a gauge theory. Such matter theories can have amplitudes that nontrivially obey the duality (as in the case of the NLSM [126] discussed in sections 3.2 and 5.3.11), thus mimicking the intricate kinematic structure of gauge theories, or they can be completely trivial manifestations of the duality (e.g. bi-adjoint ϕ^3 theory [44, 154]). The most remarkable aspect of CK duality is that it naturally leads to scattering amplitudes in double-copy theories. Section 5 describes a remarkable web of theories that are connected by the duality and the double copy.

Finally, for amplitudes that are not obtained from the standard QFT framework involving Feynman diagrams, such as string-theory amplitudes, it is convenient to define CK duality to mean that these amplitudes obey the same relations as if they were generated by a duality-satisfying diagrammatic expansion of the gauge-theory type. For example, the single-trace vector-amplitude sector of the heterotic string obeys the same relations as that of YM theory [155]. Hence, we can write heterotic string amplitudes as a sum over cubic diagrams with duality-satisfying kinematic numerators, even if this might not seem completely natural from a string-theory perspective.

2.2. General statement of the duality and the double copy for gauge theories

Consider scattering amplitudes in a nonabelian gauge theory with the following properties: there is a gauge-group G under which all fields transform nontrivially; particles of different mass are assigned to various representations of the gauge group; the interactions are controlled by a gauge coupling constant g and a set of elementary color tensors $\mathcal{C} = \left\{f^{abc}, (t^a)_i^{\ j}, \dots\right\}$. The set of elementary color tensors may include higher-rank tensors as indicated by the ellipsis.

An L-loop m-point scattering amplitude in this D-dimensional gauge theory can then be organized as[9]

$$\mathcal{A}_m^{(L)} = i^{L-1} g^{m-2+2L} \sum_i \int \frac{d^{LD}\ell}{(2\pi)^{LD}} \frac{1}{S_i} \frac{c_i n_i}{D_i}\,, \tag{35}$$

where the sum runs over the distinct L-loop m-point diagrams that can be constructed by contracting the elements of $\mathcal{C}$ in various allowed ways (consistent with the choice of external particle representations, and where the valency of each vertex is determined by the tensor rank). We take each such diagram to correspond to a unique color factor c_i. Each diagram has an associated denominator factor D_i which is constructed by taking a product of the denominators of the Feynman propagators $\sim 1/(p^2 - m_j^2)$ of each internal line of the diagram. For simplicity of notation, we assume that the color representation of the line uniquely specifies the mass m_j of the propagator. Cases with differing masses, but the same color representation, are easily taken into account by setting appropriate masses and representations equal at the end. The adjoint representation is by default massless and is associated to gluons (and, in some cases, additional fields). The remaining nontrivial kinematic dependence is collected in the kinematic numerator n_i associated with each diagram. The numerators n_i are in general gauge-dependent functions that depend on external momenta p_j, loop momenta ℓ_l, polarizations ε_j, spinors, flavor, etc everything except for the color degrees of freedom. The integral measure is defined as $d^{LD}\ell = \prod_{l=1}^{L} d^D\ell_l$. Finally, S_i are standard symmetry factors that remove internal overcount of loop diagrams; they can be computed by counting the number of discrete symmetries of each diagram with fixed external legs.

The color factors c_i are in general not independent. They satisfy linear relations that are inherited from the Lie algebra structure, such as the Jacobi identity and the defining commutation relation,

$$\begin{gathered} f^{dae} f^{ebc} - f^{dbe} f^{eac} = f^{abe} f^{ecd}\,, \\ (t^a)_i^{\ k} \left(t^b\right)_k^{\ j} - \left(t^b\right)_i^{\ k} (t^a)_k^{\ j} = i f^{abc} \left(t^c\right)_i^{\ j}\,, \end{gathered} \tag{36}$$

[9] Our conventions for the overall phase of gauge-theory and gravity amplitudes follow the one in [156] rather than the original BCJ papers [1, 2].

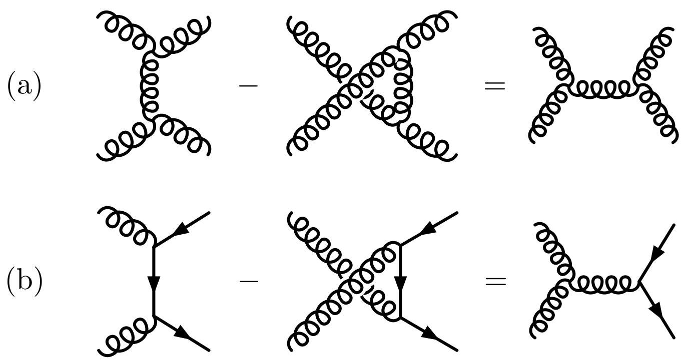


Figure 5. Color-algebra relations in the adjoint (a) and fundamental representation (b). The curly lines show adjoint representation states and the straight lines fundamental representation. The vertices correspond to the color matrices in equation (38).

and similar identities for other color tensors that might appear in the theory. In equation (36) we follow the standard textbook normalization of color generators [100],

$$\mathrm{Tr}\left(t^a t^b\right) = \frac{\delta^{ab}}{2} . \tag{37}$$

Such Lie-algebra relations are directly tied to gauge invariance of amplitudes.

In the amplitudes community, color generators differ from the textbook definition by a $\sqrt{2}$ factor absorbed into each generator [88]. It is also useful to rescale the group-theory structure constants and make them imaginary,

$$T^a \equiv \sqrt{2} t^a , \quad \tilde{f}^{abc} \equiv i \sqrt{2} f^{abc} . \tag{38}$$

After these changes, the new trace normalization is

$$\mathrm{Tr}\left(T^a T^b\right) = \delta^{ab} , \tag{39}$$

and the defining commutation relations are

$$\begin{aligned} \tilde{f}^{dae}\tilde{f}^{ebc} - \tilde{f}^{dbe}\tilde{f}^{eac} &= \tilde{f}^{abe}\tilde{f}^{ecd} , \\ \left(T^a\right)_i^{\ k} \left(T^b\right)_k^{\ j} - \left(T^b\right)_i^{\ k} \left(T^a\right)_k^{\ j} &= \tilde{f}^{abc} \left(T^c\right)_i^{\ j} , \end{aligned} \tag{40}$$

as illustrated in figure 5. These identities imply that there exist relations between triplets of color factors $\{c_i, c_j, c_k\}$ which take, for example, the form $c_i - c_j = c_k$.

The scattering amplitude (35) is said to obey CK duality if the kinematic numerator factors obey the same general algebraic relations as the color factors do, e.g.

$$n_i - n_j = n_k \quad \Leftrightarrow \quad c_i - c_j = c_k , \tag{41}$$

which is a generalization of the kinematic Jacobi identity in equation (17). The relative signs between the terms depend on choices in defining the color factors for each diagram. The essential point regarding the signs is that whatever choice is made for the color factors are inherited by the corresponding numerator factors. Another form of the duality in terms of color traces has also been found [48, 157–161], but the most natural form is in terms of color factors of diagrams as described above.

It is a nontrivial task to find duality-satisfying numerators since standard methods such as Feynman rules, on-shell recursion [162], or generalized unitarity [163–166], generally do not automatically gives such numerators. A straightforward but somewhat tedious way to find such representations is to use an ansatz constrained to match the amplitude and manifest the duality [4, 6]. Constructive ways to obtain numerators have also been devised [24–29, 167–169]. Aside from amplitudes, the duality has also been found to hold for currents with one off-shell leg [9, 17, 21, 22, 114, 170–172]. A natural way for making the duality valid for general off-shell quantities would be to find a Lagrangian that generates Feynman rules whose diagrams manifest the duality. At present, such Lagrangian is only known to a few orders in perturbation theory [41, 42, 150, 151]; an important problem is to find a closed form of such a Lagrangian valid to all orders.

The color relations (40) have important implications for kinematic numerators of diagrams. If we start with a set of numerators that satisfy the duality (41), and shift the numerators,

$$n_i = n_i' - \Delta_i \, . \tag{42}$$

subject to the constraint,

$$\sum_i \int \frac{d^{LD}\ell}{(2\pi)^{LD}} \frac{1}{S_i} \frac{c_i \Delta_i}{D_i} = 0 \, , \tag{43}$$

the amplitude is unchanged. Because the color factors are not independent, nontrivial shifts of the kinematic numerators can be carried out. In this way, without changing the amplitude, we can rewrite the amplitude in terms of a set of numerators n_i' not obeying the duality relations (41) starting from ones that do obey it. The Δ_i are pure gauge functions, i.e. they drop out of the amplitude.

When we have numerators n_i that obey the same algebraic relations as the color factors c_i, we can obtain sensible objects by formally replacing color factors by kinematic numerators as

$$c_i \to n_i \, , \tag{44}$$

in any given formula or amplitude. Given the algebraic properties are the same, this replacement is consistent with gauge-invariance properties inherited from the gauge theory. As we discuss below, this color-to-kinematics replacement—or double-copy construction—gives us gravity amplitudes with remarkable ease.

Consider two amplitudes $\mathcal{A}_m^{(L)}$ and $\widetilde{\mathcal{A}}_m^{(L)}$, and organize them as in equation (35). Furthermore, take the color factors to be the same in the two amplitudes, and label the two sets of numerators as n_i and $\tilde{n}_i$, respectively. If at least one of the amplitudes, say $\widetilde{\mathcal{A}}_m^{(L)}$, manifests CK duality, we may now replace the color factors of the first amplitude with the duality-satisfying numerators $\tilde{n}_i$ of the second one. This gives the double-copy formula for gravitational scattering amplitudes [1, 2],

$$\mathcal{M}_m^{(L)} = \mathcal{A}_m^{(L)}\Big|_{\substack{c_i \to \tilde{n}_i \\ g \to \kappa/2}} = i^{L-1} \left(\frac{\kappa}{2}\right)^{m-2+2L} \sum_i \int \frac{d^{LD}\ell}{(2\pi)^{LD}} \frac{1}{S_i} \frac{n_i \tilde{n}_i}{D_i} \, , \tag{45}$$

where the gravitational coupling $\kappa/2$ which compensates for the change of engineering dimension when replacing color factors with kinematic numerators. In general we will omit the factors of $\kappa/2$ by taking $\kappa = 2$.

For the replacement $c_i \to \tilde{n}_i$ to be valid under the integration symbol, it is important that the color factors are not explicitly evaluated by summing over the contracted indices. At least one contracted index per loop should not be explicitly summed over; this is required so that the duality is not spoiled by treating color and kinematics differently. The numerators depend on loop momenta $n_i = n_i(\ell)$ that is not yet integrated over, thus analogously the color factors should be thought of as depending on the unevaluated internal indices. If this subtlety is ignored, it may happen that color factors explicitly vanish when combining the color sum with symmetries of particular color factors, and this vanishing behavior should not be imposed on the un-integrated numerators. Stated differently, we do not wish to impose any specific color-factor properties on the numerator factors, only generic ones.

As the notation suggests, the two sets of numerators n_i, $\tilde{n}_i$ can differ in several ways: (1) they can describe different gauge choices for the same scattering process, (2) they can describe different external states in the same theory, and (3) they can originate from two different gauge theories. The first case allows us to work with numerators where only one set obeys the duality manifestly. The second case allows us to describe gravitational states that are not built out of a symmetric-tensor product

$$(\text{gravity state}) = (\text{gauge state}) \otimes (\widetilde{\text{gauge state}})\,. \tag{46}$$

The third case allows us to describe gravitational theories that are not left-right symmetric double copies of gauge theories

$$(\text{gravity theory}) = (\text{gauge theory}) \otimes (\widetilde{\text{gauge theory}})\,. \tag{47}$$

In section 5, we will see that this latter case is crucial for probing the web of double-copy-constructible theories.

When two different gauge theories are considered in the double-copy formula, it is important that both, in principle, can be put into a form displaying CK duality, even if this property needs only to be explicit in one of the amplitudes. This ensures that the generalized unitarity cuts of the loop-level double-copy formula will be unique and gauge invariant. The link between gauge invariance and BCJ amplitude relations has been explored in [173–177]. The amplitude relations can also be understood in terms of a symmetry that act as momentum-dependent shifts on the color factors [178, 179]. Note that the precise form of the BCJ amplitude relations depends on the details of the gauge-group representations and elementary color tensors. The standard BCJ amplitude relations [1], for example, follow from considering theories with only adjoint particles that interact via f^{abc} color tensors.

We will come back to the double-copy constructions of different theories in later sections, but for now we will focus on illustrating the details of CK duality on some familiar gauge theories.

2.3. Example 1: tree level amplitudes with adjoint-only particles

Consider pure YM theory in D spacetime dimensions, consisting of gluons transforming in the adjoint representation of a gauge group G, with Lagrangian

$$\mathcal{L}_{\text{YM}} = -\frac{1}{4}\left(F^a_{\mu\nu}\right)^2, \quad \text{where} \quad F^{\;a}_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu\,. \tag{48}$$

Next consider m-point tree-level amplitudes. We know that the only color structure that appears are contractions of f^{abc} structure constants, thus the color factors must be in one-to-one correspondence with all possible cubic diagrams with m external legs.

Cubic diagrams at multiplicity $m = j+1$ can be built recursively by attaching a new leg to every possible edge of a multiplicity-j diagram. There are $(2j-3)$ edges of a given j-point diagram, hence the recursion gives:

$$\text{number of cubic diagrams} = 1\times 3\times 5\times 7\times\cdots\times(2j-3) = (2m-5)!!\,. \tag{49}$$

We organize the tree amplitude in terms of all such propagator-distinct diagrams with only cubic vertices,

$$\mathcal{A}_m^{\text{tree}} \equiv \mathcal{A}_m^{(0)} = -i\, g^{m-2} \sum_{i=1}^{(2m-5)!!} \frac{c_i\, n_i}{D_i}\,, \tag{50}$$

where c_i are the color factors that are straightforwardly obtained from the ith diagram. Similarly, the D_i denote the denominators of the propagators that correspond to the diagram lines. The n_i are the corresponding kinematic numerators. Depending on the context we will alternate between using the diagram factors n_i, c_i, D_i with subscripts indexed by a diagram-id number, as well as a functional maps from graph to their respective factors: $n_i \equiv n(g_i)$, $c_i \equiv c(g_i)$, and $D_i \equiv D(g_i)$ where g_i is the graph corresponding to the index i.

It is useful to first clarify what we mean by independent diagrams. The least redundancy occurs when we insist on only one instance of a diagram with the same propagator contribution. This is distinct from the number of unique diagram topologies. Let us take a concrete example at four-points. We have discussed in section 1 that we need s, t, and u diagrams at four point. They have same graphical topology, but different external labels, which results in different generic propagator contributions.

As a trivial example, at four points for each distinct propagator structure we can relabel the external legs without altering the propagators but flipping the signs of the color. For example, consider the s-channel diagram in figure 3 which we can label as $g_{s:1}$. Taking the graph $g_{s:2}$ to be $g_{s:1}$ but with legs 1 and 2 swapped, we obtain the same propagator but the color factors are different:

$$\begin{aligned} c\left(g_{s:1}\right) &= \tilde{f}^{a_1 a_2 b}\tilde{f}^{b a_3 a_4}\,, \\ c\left(g_{s:2}\right) &= \tilde{f}^{a_2 a_1 b}\tilde{f}^{b a_3 a_4}\,, \end{aligned} \tag{51}$$

where we use the normalization in equation (38). The color factors, while distinct, are related by a negative sign inherited by the antisymmetry of the structure generators: $c(g_{s:2}) = -c(g_{s:1})$. For the purpose of describing scattering amplitudes in terms of functions of diagrams, we will always take the kinematic factors of the diagrams to obey the same antisymmetry: $n(g_{s:2}) = -n(g_{s:1})$, whether or not we are discussing a CK-satisfying representation[10]. This means that for any multiplicity and loop order we will have in mind a canonical layout of distinct diagrams which determine the color factor and numerator signs. These signs cancel from color-dressed amplitudes because the numerator sign are correlated with the color signs. However, they will affect the signs appearing in the relation between color-ordered partial amplitudes and kinematic numerators, as well as the relative signs between terms in the Jacobi identities.

To be more explicit, as illustrated in figure 6, triplets of diagrams (i,j,k) satisfy Jacobi relations of the form

$$c_i - c_j + c_k = \left(\tilde{f}^{dae}\tilde{f}^{ebc} - \tilde{f}^{abe}\tilde{f}^{ecd} + \tilde{f}^{dbe}\tilde{f}^{eca}\right) C^{abcd} = 0\,. \tag{52}$$

where the last factor C^{abcd} is a color tensor that is common to the diagrams in the triplet (external adjoint indices $a_1, \ldots, a_m$ are suppressed). As noted above, the relative signs are simply due to choices in the ordering of the color indices in the $\tilde{f}^{abc}$s. While these relative sign choices are arbitrary, these signs are the same as for the corresponding kinematic Jacobi identities.

[10] The word *representation* here refers to the specific functional form used to describe the amplitude.

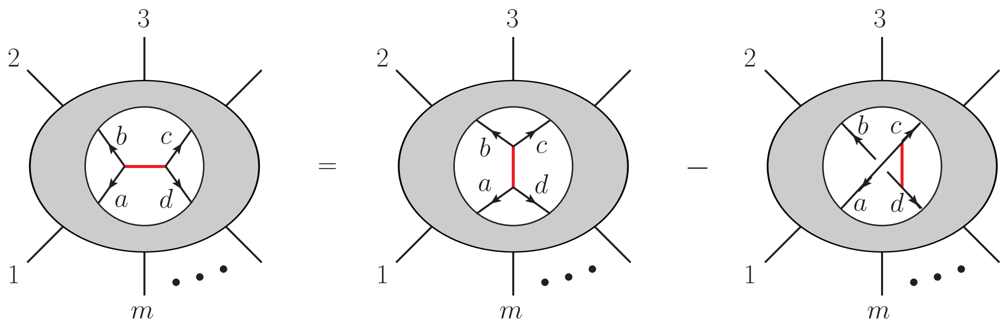
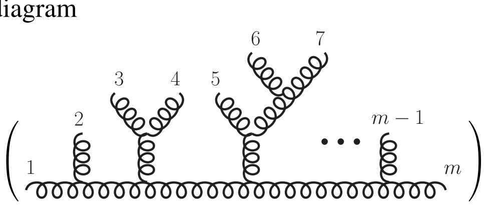


Figure 6. A Jacobi identity embedded in a generic diagram. The diagram can be either at tree level or at loop level. The arrows indicate that the lines are oriented the same way.

More generally, the $(2m-5)!!$ color factors in (50) are related by Jacobi identities. In total, at every multiplicity m there are $\frac{1}{3}(m-3)(2m-5)!!$ such Jacobi relations; however, only $(2m-5)!!-(m-2)!$ of them are independent equations because we can formally solve all Jacobi relations by mapping the color factors to a $(m-2)!$ basis.

Writing the adjoint generator matrices as $(\tilde{f}^a)_{bc} \equiv \tilde{f}^{bac}$, defined in equation (38), we can write any color factor as products of $\tilde{f}^{a_i}$'s, possibly involving commutators of the adjoint generators. For example, pick a cubic tree diagram and find the unique path through the diagram that connect leg 1 and leg m. For each cubic vertex along this path, write down the corresponding commutator of $\tilde{f}^{a_i}$'s that describes the subdiagram that attaches this vertex. The product of these factors give c_i for the full diagram. For example, consider the color factor of the following diagram

$c\Bigg($  $\Bigg) = (\tilde{f}^{a_2}[\tilde{f}^{a_3},\tilde{f}^{a_4}][\tilde{f}^{a_5},[\tilde{f}^{a_6},\tilde{f}^{a_7}]]\cdots\tilde{f}^{a_{m-1}})_{a_1 a_m}\,,$ (53)

where the adjoint indices of leg 1 and m correspond to the external matrix indices of the adjoint representation. The commutators arise from systematically eliminating subdiagrams involving $\tilde{f}^{bac}$ using the standard Lie-algebra identity $\tilde{f}^{abc}\tilde{f}^c = [\tilde{f}^a, \tilde{f}^b]$. Once only commutators of $\tilde{f}_i^a$'s remain, they can of be written out as differences and sums of generators in different orders.

In summary, any color factor can in general be written as

$$c_i = \sum_{\sigma \in S_{m-2}} b_{i\sigma} \left(\tilde{f}^{a_{\sigma(2)}}\tilde{f}^{a_{\sigma(3)}}\tilde{f}^{a_{\sigma(4)}}\cdots\tilde{f}^{a_{\sigma(m-1)}}\right)_{a_1 a_m}\,, \tag{54}$$

where $b_{i\sigma} \in \{0, \pm 1\}$ are coefficients that depend on the permutation and on the specific color factor. They can be evaluated case by case, but their explicit values are not important here for our purposes. The main result is that color factors c_i in equation (50) can be eliminated in favor of expressing the gauge-theory tree amplitude in terms of a sum over the possible products of adjoint generators $\tilde{f}^{a_i}$, where the first and m-th leg is kept fixed. This gives a so-called Del Duca–Dixon–Maltoni (DDM) color decomposition [180] of the gauge-theory tree amplitude,

$$\mathcal{A}_m^{\mathrm{tree}} = g^{m-2} \sum_{\sigma \in S_{m-2}} A_m^{\mathrm{tree}}\left(1, \sigma\left(2\right), \sigma\left(3\right), \ldots, \sigma\left(m-1\right), m\right) \left(\tilde{f}^{a_{\sigma(2)}} \tilde{f}^{a_{\sigma(3)}} \cdots \tilde{f}^{a_{\sigma(m-1)}}\right)_{a_1 a_m}, \tag{55}$$

where the sum runs over $(m-2)!$ permutations. The kinematic coefficients multiplying the color factors define a basis of $(m-2)!$ partial amplitudes, which we indicate as $A_m^{\mathrm{tree}}\big(1, \sigma(2), \sigma(3), \ldots, \sigma(m-1), m\big)$. This is usually called the Kleiss-Kuijf (KK) basis [181].

The partial tree amplitudes in YM theory, $A_m^{\mathrm{tree}}\big(1, 2, \ldots, m\big)$, have a number of useful properties [88]:

- They are functions of kinematic variables only, (ε_i, p_i); the color dependence is only reflected by the ordering of the external particle labels.
- They at most have poles in planar channels, i.e. when consecutive momenta add up to a null momentum $\big(\sum_{j \leq i \leqslant k} p_i\big)^2 = 0 \; (\mathrm{mod}\; m)$.
- The amplitudes are invariant under cyclic permutations:

$$A_m^{\mathrm{tree}}\left(1, 2, \ldots, m\right) = A_m^{\mathrm{tree}}\left(2, \ldots, m, 1\right). \tag{56}$$

- Under reversal of the ordering, they at most change by a sign flip:

$$A_m^{\mathrm{tree}}\left(m, \ldots, 2, 1\right) = \left(-1\right)^m A_m^{\mathrm{tree}}\left(1, 2, \ldots, m\right). \tag{57}$$

- They satisfy a photon-decoupling identity:

$$\sum_{\sigma \in \mathrm{cyclic}} A_m^{\mathrm{tree}}\left(1, \sigma\left(2\right), \ldots, \sigma\left(m\right)\right) = 0, \tag{58}$$

  where cyclic permutations of all but one leg are summed over.
- They satisfy KK relations [181]:

$$A_m^{\mathrm{tree}}\left(1, \alpha, m, \beta\right) = \left(-1\right)^{|\beta|} \sum_{\sigma \in \alpha \,\sqcup\!\sqcup\, \beta^T} A_m^{\mathrm{tree}}\left(1, \sigma, m\right), \tag{59}$$

  where α and β are arbitrary-sized lists of the external legs, β^T is used to represent the reverse ordering of the list β, and $\alpha \,\sqcup\!\sqcup\, \beta^T$ is the shuffle product of these lists (i.e. permutations that separately maintain the order of the individual elements belonging to each list). $|\beta|$ denotes the number of elements in the list β.
- They obey BCJ relations, which in the simplest incarnation take the form [1]:

$$\sum_{i=2}^{m-1} p_1 \cdot \left(p_2 + \ldots + p_i\right) A_m^{\mathrm{tree}}\left(2, \ldots, i, 1, i+1, \ldots, m\right) = 0. \tag{60}$$

- After considering all permutations of the above BCJ relation, there are only $(m-3)!$ independent partial tree amplitudes [1]. The position of three consecutive legs can be fixed in the cyclic ordering; for example, $A_m^{\mathrm{tree}}\big(1, 2, \sigma(3), \ldots, \sigma(m-1), m\big)$ can be chosen as the independent BCJ basis.

The first property is obvious from our definition of the partial amplitudes; however, the remaining ones require some explanation.

The fact that the partial tree amplitudes are invariant under cyclic permutations of their arguments is most easily seen after a basis change of the color factors. We rewrite the color factors in terms of traces of generators, T^a. From equation (40)

$$\tilde{f}^{abc} \equiv i\sqrt{2}f^{abc} = \mathrm{Tr}\left(\left[T^a, T^b\right]T^c\right) = \mathrm{Tr}\left(T^aT^bT^c\right) - \mathrm{Tr}\left(T^bT^aT^c\right), \tag{61}$$

which follows from the identity (40) after multiplying both sides with $T^{c'}$, tracing over the fundamental indices, and using equation (39). This basis change implicitly assumes that we have specialized to a gauge group were we can use 't Hooft's double-line notation [182], say $G = U(N_c)$.

The generators of $U(N_c)$ obey the completeness relation $(T^a)_i^{\ j}(T^a)_k^{\ l} = \delta_i^l\delta_k^j$, implying that products of several f^{abc} can be expressed by merging several traces

$$\begin{aligned}\left(\tilde{f}^{a_2}\tilde{f}^{a_3}\cdots\tilde{f}^{a_{m-1}}\right)_{a_1a_m} &= \tilde{f}^{a_1a_2b_1}\tilde{f}^{b_1a_3b_2}\cdots\tilde{f}^{b_{m-3}a_{m-1}a_m} \\ &= \mathrm{Tr}\left(T^{a_1}T^{a_2}T^{a_3}\cdots T^{a_m}\right) + (-1)^m\,\mathrm{Tr}\left(T^{a_m}\cdots T^{a_3}T^{a_2}T^{a_1}\right) + \ldots\end{aligned} \tag{62}$$

where on the last line the suppressed terms corresponds to 2^{m-1} distinct permutations of the trace over m generators. Out of all the permutations that appear, only the two displayed terms have the property that the generators T^{a_1} and T^{a_m} are adjacent (in the cyclic sense). This implies that, after replacing the DDM color factors with the trace-basis color factors in equation (55), we can uniquely identify the location of, say, the $\mathrm{Tr}(T^{a_1}T^{a_2}T^{a_3}\cdots T^{a_m})$ factor. It appears only once in $(\tilde{f}^{a_2}\tilde{f}^{a_3}\cdots\tilde{f}^{a_{m-1}})_{a_1a_m}$, which uniquely multiplies the partial tree amplitude $A_m^{\mathrm{tree}}(1,2,3,\ldots,m)$. Hence, $A_m^{\mathrm{tree}}(1,2,3,\ldots,m)$ must be the kinematic coefficient of $\mathrm{Tr}(T^{a_1}T^{a_2}T^{a_3}\cdots T^{a_m})$ in the trace-basis decomposition of the YM tree amplitude.

By crossing symmetry in the trace basis, the decomposition into partial amplitudes has the form

$$\mathcal{A}_m^{\mathrm{tree}} = g^{m-2}\sum_{\sigma\in S_{m-1}} A_m^{\mathrm{tree}}\left(1,\sigma(2),\sigma(3),\ldots,\sigma(m)\right)\mathrm{Tr}\left(T^{a_1}T^{a_{\sigma(2)}}T^{a_{\sigma(3)}}\cdots T^{a_{\sigma(m)}}\right), \tag{63}$$

which can be straightforwardly verified starting from equation (55). Crossing symmetry requires the summation over $(m-1)!$ terms, since we can fix the location of one leg, say leg 1, by the cyclic property of the trace. $(m-1)!$ is significantly larger than the $(m-2)!$ terms in equation (55). Where did the extra terms come from? In fact, they are the terms we suppressed in equation (62), which have combined in various ways to complete the formula (63). Finally, since the partial amplitudes in the DDM decomposition are the same partial amplitudes that appear in the trace decomposition, it follows that the partial amplitudes inherit the cyclic invariance of the trace. Further details of the trace basis and partial amplitudes may be found in [88].

A number of the other properties discussed above also follow from the exercise of mapping between the DDM and trace basis. The reversal (anti-)symmetry follows from observing that the term $(-1)^m\mathrm{Tr}(T^{a_m}\cdots T^{a_3}T^{a_2}T^{a_1})$ in equation (62) always goes together with $\mathrm{Tr}(T^{a_1}T^{a_2}T^{a_3}\cdots T^{a_m})$. The photon-decoupling identity follows from realizing that we can replace one generator in equation (63) by the $U(1)$ 'photon' generator $T_{U(1)} = 1$, which naturally belongs to the gauge group $U(N_c) = SU(N_c)\times U(1)$. However, gluons do not couple directly to photons since the latter have no charges. This can be seen directly by looking at the structure constants $\tilde{f}^{abU(1)} = \mathrm{Tr}([T^a, T^b]\,1) = 0$. Hence, the photon-decoupling identity follows from the vanishing of the amplitude with one photon.

The KK relations are explained by the fact that there are two different decompositions of the tree amplitude, the DDM (55) and the trace (63) decomposition, which use a different number of partial amplitudes, $(m-2)!$ and $(m-1)!$, respectively. The only way that this can be consistent is if there exist relations that map the $(m-1)!$ partial amplitudes into a $(m-2)!$ basis. This is precisely what the KK relations do. Recall that it was because the color factors are built only out of $\tilde{f}^{abc}$'s, which obey the Jacobi relations, that we could find the $(m-2)!$ basis. Thus, any theory where all fields transform in the adjoint representation and the whose amplitudes depend only color tensor is f^{abc} will obey the KK relations.

The BCJ amplitude relations are a consequence of CK duality, specifically of its interplay with gauge invariance. Consider the $(m-2)!$ partial amplitudes expressed in terms of numerators, they take the form

$$A_m^{\text{tree}}(1,\sigma(2),\ldots,\sigma(m-1),m) = -i\sum_{i\in\text{planar}} b_{i\sigma}\frac{n_i}{D_i}\,, \tag{64}$$

where n_i are the kinematic numerator factors of the diagrams canonical to some ordering layout, D_i are the propagators of the diagram, $b_{i\sigma}\in\{0,\pm1\}$ are coefficients that depend on the ordering σ (see also equation (54)), and the sum is only nonvanishing for planar diagrams with respect to the ordering σ.

We can impose kinematic Jacobi identities on the numerators, expressing all diagram numerators in terms of $(m-2)!$ independent master numerators. We can then imagine attempting to invert the matrix between color-ordered amplitudes and these master numerators. One will find a remarkable surprise—only $(m-3)!$ master numerators can be solved for in terms of some $(m-3)!$ color-ordered amplitudes—the rest of the master numerators contribute only as unfixed parameters representing a kind of generalization of gauge freedom. The remaining equations relate color-ordered amplitudes directly to simple functions of the ordered amplitudes with $(m-3)!$ legs fixed with no dependence on numerator choice. For example, following the original presentation [1], one can express the entirety of the KK $(m-2)!$ basis amplitudes in terms of $(m-3)!$ amplitudes as follows:

$$A_m^{\text{tree}}(1,2,\{\alpha\},3,\{\beta\}) = \sum_{\sigma\in\text{POP}(\{\alpha\},\{\beta\})} A_m^{\text{tree}}(1,2,3,\sigma)\prod_{k=4}^{|\alpha|+3}\frac{\mathcal{F}_k(3,\sigma,1)}{s_{24\ldots k}}\,, \tag{65}$$

where $|\alpha|$ is the length of the list $\{\alpha\}$, and the sum runs over partially ordered permutations (POP) of the merged $\{\alpha\}$ and $\{\beta\}$ sets. To be clear we are referring to leg labels, e.g. in $s_{24\ldots k}$, with labels 4 through k as the first $(k-3)$ entries of the ordered list $\{\alpha,\beta\}$. Equation (65) gives all permutations of $\{\alpha\}\bigcup\{\beta\}$ consistent with the order of the $\{\beta\}$ elements. Either α or β may be empty, trivially so for the α case. The function $\mathcal{F}_k$ associated with leg k is given by,

$$\mathcal{F}_k(\{\rho\}) = \left\{\begin{array}{ll} \sum_{l=t_k}^{m-1}\mathcal{S}_{k,\rho_l} & \text{if } t_{k-1}<t_k \\ -\sum_{l=1}^{t_k}\mathcal{S}_{k,\rho_l} & \text{if } t_{k-1}>t_k \end{array}\right\} + \left\{\begin{array}{ll} s_{24\ldots k} & \text{if } t_{k-1}<t_k<t_{k+1} \\ -s_{24\ldots k} & \text{if } t_{k-1}>t_k>t_{k+1} \\ 0 & \text{otherwise} \end{array}\right\}, \tag{66}$$

where

$$s_{ij\ldots k}\equiv(p_i+p_j+\cdots+p_k)^2\,, \tag{67}$$

and t_k is the position of leg k in the set $\{\rho\}$, except for t_3 and $t_{|\alpha|+4}$ which are always defined to be,

$$t_3\equiv t_5\,, \qquad t_{|\alpha|+4}\equiv 0\,. \tag{68}$$

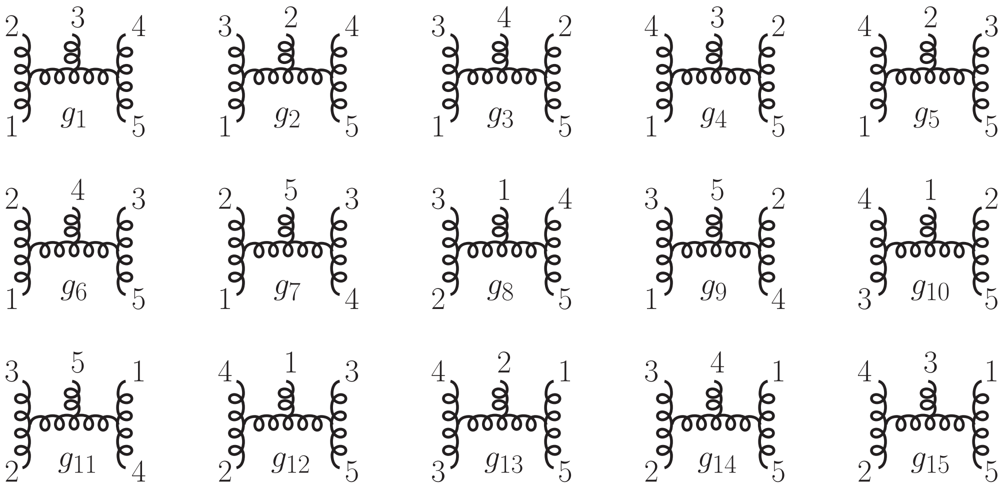


Figure 7. The color-dressed tree-level five-point amplitude an be organized using these fifteen graphs with only cubic vertices.

For $|\alpha| = 1$ this means that $t_3 = t_5 = 0$. The expression $\mathcal{S}_{i,j}$ is given by,

$$\mathcal{S}_{i,j} = \left\{ \begin{array}{ll} s_{ij} & \text{if } i < j \text{ or } j = 1 \text{ or } j = 3 \\ 0 & \text{otherwise} \end{array} \right\}. \tag{69}$$

The so-called *fundamental* BCJ relations (equation (60)) occur when the $|\alpha| = 1$. These amplitude relations were first identified in [1], and then proven, first as a low-energy limit of string-theory relations [104, 105], and then directly using the Britto-Cachazo-Feng-Witten (BCFW) recursion relations in field theory [113, 115].

Consider the five-point amplitude (e.g. governing two-to-three scattering), which offers a first nontrivial example. In this case, 15 distinct cubic diagrams contribute, as illustrated in figure 7. Only five of these contribute to a given color-ordered partial amplitude. Let us consider diagram nine from figure 7. To see which color-orderings (and with which signs) this diagram can contribute, we expand its canonical color-factor in the trace basis. The color factors follow from dressing with the structure functions $\tilde{f}^{abc}$. Going to a trace basis we see that the color factor associated with diagram nine is:

$$\begin{aligned} c_9 = & \operatorname{Tr}\left[T^{a_1}T^{a_2}T^{a_4}T^{a_5}T^{a_3}\right] - \operatorname{Tr}\left[T^{a_1}T^{a_3}T^{a_2}T^{a_4}T^{a_5}\right] + \operatorname{Tr}\left[T^{a_1}T^{a_3}T^{a_4}T^{a_2}T^{a_5}\right] \\ & + \operatorname{Tr}\left[T^{a_1}T^{a_3}T^{a_5}T^{a_2}T^{a_4}\right] - \operatorname{Tr}\left[T^{a_1}T^{a_3}T^{a_5}T^{a_4}T^{a_2}\right] - \operatorname{Tr}\left[T^{a_1}T^{a_4}T^{a_2}T^{a_5}T^{a_3}\right] \\ & - \operatorname{Tr}\left[T^{a_1}T^{a_5}T^{a_2}T^{a_4}T^{a_3}\right] + \operatorname{Tr}\left[T^{a_1}T^{a_5}T^{a_4}T^{a_2}T^{a_3}\right]. \end{aligned} \tag{70}$$

This implies that diagram nine will contribute to multiple color-ordered partial amplitudes, defined as the coefficient of each color trace in the full amplitude, with a variety of signs. The signs associated with each diagram in a partial amplitude are easily determined for a given color ordering by reordering the legs of each diagram to match the color ordering without allowing lines to cross, and keeping track of the signs from permuting the ordering of legs in each vertex.

Taking the layout as depicted in figure 7, each diagram contributes to a color-ordered partial amplitude according to whether we can flip the legs at each vertex (with a minus sign for each flip) so that the cyclic ordering of legs matches the ordering of the arguments of the partial amplitudes. For example, we have,

$$iA_5^{\text{tree}}(1,3,5,4,2) = \frac{n_1}{D_1} + \frac{n_2}{D_2} + \frac{n_6}{D_6} - \frac{n_9}{D_9} - \frac{n_{12}}{D_{12}}, \tag{71}$$

as well as

$$iA_5^{\text{tree}}(1,3,5,2,4) = \frac{n_3}{D_3} + \frac{n_4}{D_4} + \frac{n_5}{D_5} + \frac{n_9}{D_9} + \frac{n_{12}}{D_{12}}, \tag{72}$$

where the n_i are the kinematic numerators and the $1/D_i$ are the products of Feynman propagators that can be read of from graph g_i in figure 7.

Jacobi relations imply that the n_i of the diagrams in figure 7 are given as linear functions of numerators $\{n_1, n_2, n_3, n_4, n_5, n_6\}$, which we take as the master numerators. In total there are nine independent Jacobi relations,

$$\begin{aligned} &n_7 = n_6 - n_1, \quad n_8 = n_2 - n_1, \quad n_9 = n_3 - n_2, \quad n_{10} = n_4 - n_3, \quad n_{11} = n_5 - n_4, \\ &n_{12} = n_5 - n_6, \quad n_{13} = n_{10} - n_7, \quad n_{14} = n_{11} + n_8, \quad n_{15} = n_{12} - n_9. \end{aligned} \tag{73}$$

Solving this system in terms of the six master numerators gives

$$\begin{aligned} &n_7 = -n_1 + n_6, \quad n_8 = -n_1 + n_2, \quad n_9 = -n_2 + n_3, \quad n_{10} = -n_3 + n_4, \\ &n_{11} = -n_4 + n_5, \quad n_{12} = n_5 - n_6, \quad n_{13} = n_1 - n_3 + n_4 - n_6, \\ &n_{14} = -n_1 + n_2 - n_4 + n_5, \quad n_{15} = n_2 - n_3 + n_5 - n_6. \end{aligned} \tag{74}$$

Remarkably, by using equation (74), we can show that the partial amplitudes (71) and (72) contain all information necessary to describe all other ordered amplitudes at five point. For the sake of argument, solving equations (71) and (72) for n_1 and n_4 gives:

$$\begin{aligned} n_1 &= iD_1A_5^{\text{tree}}(1,3,5,4,2) - n_2\frac{D_1}{D_2} - n_6\frac{D_1}{D_6} + (n_3 - n_2)\frac{D_1}{D_9} + (n_5 - n_6)\frac{D_1}{D_{12}}, \\ n_4 &= iD_4A_5^{\text{tree}}(1,3,5,2,4) - n_3\frac{D_4}{D_3} - n_5\frac{D_4}{D_5} + (n_2 - n_3)\frac{D_4}{D_9} + (n_6 - n_5)\frac{D_4}{D_{12}}, \end{aligned} \tag{75}$$

where we have replaced n_9 and n_{12} with the master numerators, using equation (74). Using this, we express any other partial amplitude in terms of $A_5^{\text{tree}}(1,3,5,4,2)$ and $A_5^{\text{tree}}(1,3,5,2,4)$ by plugging in the solution (74) for these expressions for n_1 and n_4. Consider, for example, the partial amplitude:

$$iA_5^{\text{tree}}(1,3,2,5,4) = -\frac{n_2}{D_2} - \frac{n_3}{D_3} - \frac{n_4}{D_4} - \frac{n_8}{D_8} + \frac{n_{11}}{D_{11}}. \tag{76}$$

Jacobi relations constrain $n_8 = -n_1 + n_2$, and $n_{11} = -n_4 + n_5$. Replacing all non-master numerators with master numerators using equation (74), in conjunction with equation (75), we find:

$$
\begin{aligned}
iA_5^{\text{tree}}(1,3,2,5,4) = {} & iA_5^{\text{tree}}(1,3,5,4,2)\frac{D_1}{D_8} + iA_5^{\text{tree}}(1,3,5,2,4)\left(-1-\frac{D_4}{D_{11}}\right) \\
& -\left(\frac{1}{D_2}+\frac{1}{D_8}+\frac{D_1}{D_2D_8}+\frac{1}{D_9}+\frac{D_1}{D_8D_9}+\frac{D_4}{D_9D_{11}}\right)n_2 \\
& +\left(\frac{1}{D_9}+\frac{D_1}{D_8D_9}+\frac{D_4}{D_3D_{11}}+\frac{D_4}{D_9D_{11}}\right)n_3 \\
& +\left(\frac{1}{D_5}+\frac{1}{D_{11}}+\frac{D_4}{D_5D_{11}}+\frac{1}{D_{12}}+\frac{D_1}{D_8D_{12}}+\frac{D_4}{D_{11}D_{12}}\right)n_5 \\
& -\left(\frac{D_1}{D_6D_8}+\frac{1}{D_{12}}+\frac{D_1}{D_8D_{12}}+\frac{D_4}{D_{11}D_{12}}\right)n_6\,.
\end{aligned}
\tag{77}
$$

Using the explicit value of the propagators, a dramatic cancellation occurs under momentum conservation:

$$
\begin{aligned}
A_5^{\text{tree}}(1,3,2,5,4) = {} & -A_5^{\text{tree}}(1,3,5,2,4)\left(1+\frac{s_{25}}{s_{23}}\right)+A_5^{\text{tree}}(1,3,5,4,2)\frac{s_{345}}{s_{23}} \\
& -n_6\frac{(s_{23}+s_{24}+s_{25}+s_{345})}{s_{23}s_{24}s_{35}}+n_3\frac{(s_{23}+s_{24}+s_{25}+s_{345})}{s_{23}s_{24}s_{245}} \\
& +n_5\frac{(s_{24}\;(s_{25}+s_{35})+s_{23}(s_{24}+s_{235})+s_{235}(s_{25}+s_{345}))}{s_{23}s_{24}s_{35}s_{235}} \\
& -n_2\frac{(s_{23}(s_{24}+s_{45})+s_{45}(s_{25}+s_{345})+s_{24}(s_{245}+s_{345}))}{s_{23}s_{24}s_{45}s_{245}} \\
= {} & -A_5^{\text{tree}}(1,3,5,2,4)\left(1+\frac{s_{25}}{s_{23}}\right)+A_5^{\text{tree}}(1,3,5,4,2)\frac{s_{345}}{s_{23}}\,.
\end{aligned}
\tag{78}
$$

All the coefficients in front of the remaining explicit numerators vanish, giving $A_5^{\text{tree}}(1,3,2,5,4)$ solely in terms of a basis of partial amplitudes $A_5^{\text{tree}}(1,3,5,4,2)$ and $A_5^{\text{tree}}(1,3,5,2,4)$. Indeed, this occurs for every partial amplitude, yielding, the BCJ and KK amplitude relations [1].

An interesting corollary of the independence of the remaining five-point partial amplitudes on n_2,n_3,n_5,n_6, once n_1 and n_4 are chosen as in equation (75), is that we can choose to set the former numerators to zero since they have no effect on any of the partial amplitudes. This forces n_1 and n_4 to be nonlocal since they must absorb the propagators of the diagrams whose numerators are set to zero.

2.3.1. KLT formula and proof of tree-level adjoint CK duality. In this section we briefly review the KLT formulae for gravity tree-level amplitudes [86], first derived using string theory. We will show how they are intimately tied to the BCJ double copy in equation (45), and how they can be used to directly construct duality-satisfying numerators in purely-adjoint gauge theories.

Let us start by quoting the explicit KLT relations at three-, four-, five- and six-points,

$$
\begin{aligned}
\mathcal{M}_3^{\text{tree}}(1,2,3) &= iA_3^{\text{tree}}(1,2,3)\widetilde{A}_3^{\text{tree}}(1,2,3)\,, \\
\mathcal{M}_4^{\text{tree}}(1,2,3,4) &= -is_{12}A_4^{\text{tree}}(1,2,3,4)\widetilde{A}_4^{\text{tree}}(1,2,4,3)\,, \\
\mathcal{M}_5^{\text{tree}}(1,2,3,4,5) &= is_{12}s_{45}A_5^{\text{tree}}(1,2,3,4,5)\widetilde{A}_5^{\text{tree}}(1,3,5,4,2) \\
&\quad + i\,s_{14}s_{25}A_5^{\text{tree}}(1,4,3,2,5)\widetilde{A}_5^{\text{tree}}(1,3,5,2,4)\,,
\end{aligned}
$$

$$\mathcal{M}_6^{\text{tree}}(1,2,3,4,5,6) = -i s_{12} s_{45} A_6^{\text{tree}}(1,2,3,4,5,6)\Big(s_{35}\widetilde{A}_6^{\text{tree}}(2,1,5,3,4,6) + (s_{34}+s_{35})\widetilde{A}_6^{\text{tree}}(2,1,5,4,3,6)\Big) + \mathcal{P}(2,3,4)\,, \tag{79}$$

where the $\mathcal{M}_n^{\text{tree}}$ are tree-level gravity amplitudes and the A_n^{tree} are color-ordered gauge-theory partial amplitudes, and $\mathcal{P}(2,3,4)$ represents a sum over all permutations of leg labels $2,3$, and 4.

If A and $\tilde{A}$ are the tree-level amplitudes of D-dimensional pure YM theory, then the map between the two sets of on-shell gluon polarization vectors ε^i_μ, with $SO(D-2)$ little-group indices i, and those of the double-copy fields can be made explicit,

$$\begin{aligned}
(\varepsilon^h)^{ij}_{\mu\nu} &= \varepsilon^{((i}_\mu \varepsilon^{j))}_\nu \qquad (\text{graviton})\,,\\
(\varepsilon^B)^{ij}_{\mu\nu} &= \varepsilon^{[i}_\mu \varepsilon^{j]}_\nu \qquad (B\text{-field})\,,\\
(\varepsilon^\phi)_{\mu\nu} &= \frac{\varepsilon^i_\mu \varepsilon^j_\nu \delta_{ij}}{D-2} \qquad (\text{dilaton})\,.
\end{aligned} \tag{80}$$

On the first line the gluon polarizations are multiplied in symmetric-traceless combinations corresponding to the $\frac{1}{2}(D-2)(D-1)-1$ states of a graviton. On the second line they are antisymmetrized corresponding to the $\frac{1}{2}(D-2)(D-3)$ states of an antisymmetric tensor field. The completeness of the set of gluon polarization vectors implies that the right-hand side of the third line of equation (80) is proportional to $\eta_{\mu\nu}$ up to momentum-dependent terms, so $(\varepsilon^\phi)_{\mu\nu}$ describes a single state. Adding them all up we find the $(D-2)^2$ states in the tensor product of two massless vectors, as we should.

The double copy of D-dimensional pure YM theory gives gravity amplitudes $\mathcal{M}^{\text{tree}}$ that follow from the Lagrangian [183, 184]

$$S = \int d^D x\sqrt{-g}\left[-\frac{1}{2}R + \frac{1}{2(D-2)}\partial^\mu\phi\,\partial_\mu\phi + \frac{1}{6}e^{-4\phi/(D-2)}H^{\lambda\mu\nu}H_{\lambda\mu\nu}\right], \tag{81}$$

where $H_{\lambda\mu\nu}$ is the field strength of the two-index antisymmetric tensor $B_{\mu\nu}$ and the non-canonical normalization of the dilaton quadratic term is chosen to avoid non-rational dependence on the spacetime dimension D. The $\mathbb{Z}_2$ symmetry $B_{\mu\nu}\to -B_{\mu\nu}$ generates a consistent truncation of this Lagrangian to Einstein gravity coupled to ϕ. The further $\mathbb{Z}_2$ symmetry of this truncation, $\phi\to-\phi$, allows a further consistent truncation to Einstein gravity. The double copy analog of this truncation is realized by choosing gluon polarizations in symmetric-traceless combinations, as for the graviton polarizations in equation (80).

To show the connection to the BCJ double copy, consider, for example, the five-point tree amplitude. The double-copy amplitude in terms of Jacobi-satisfying numerators is,

$$\begin{aligned}
\mathcal{M}_5^{\text{tree}}(1,2,3,4,5) &= -i\sum_{i=1}^{15}\frac{n_i\tilde{n}_i}{D_i}\\
&= -i\,\tilde{n}_1\left(\frac{n_1}{D_1}+\frac{n_1}{D_7}+\frac{n_1}{D_8}+\frac{n_1+n_4}{D_{13}}+\frac{n_1+n_4}{D_{14}}\right)\\
&\quad -i\,\tilde{n}_4\left(\frac{n_4}{D_4}+\frac{n_4}{D_{10}}+\frac{n_4}{D_{11}}+\frac{n_1+n_4}{D_{13}}+\frac{n_1+n_4}{D_{14}}\right),
\end{aligned} \tag{82}$$

where we used the solution (74) and the propagators $1/D_i$ can be read off from the diagrams in figure 7. As usual, where we suppress an overall factor of $(\kappa/2)^3$. Remarkably, after using the solution (74) for both the n_i and $\tilde{n}_i$, the result depends only on the numerators n_1, n_4 and $\tilde{n}_1, \tilde{n}_4$. Using equation (75), we have,

$$\begin{aligned}\mathcal{M}_5^{\text{tree}}(1,2,3,4,5) &= -\left(\tilde{n}_1 A_5^{\text{tree}}(1,2,3,4,5) + \tilde{n}_4 A_5^{\text{tree}}(1,4,3,2,5)\right) \\ &= i\, s_{12}s_{45}A_5^{\text{tree}}(1,2,3,4,5)\widetilde{A}_5^{\text{tree}}(1,3,5,4,2) \\ &\quad + i\, s_{14}s_{25}A_5^{\text{tree}}(1,4,3,2,5)\widetilde{A}_5^{\text{tree}}(1,3,5,2,4)\,. \end{aligned} \tag{83}$$

This implies that we can express the double copy in terms of partial tree amplitudes of the two gauge theories.

This structures applies also at higher points, and is captured by the m-point formula [86]:

$$\mathcal{M}_m^{\text{tree}} = -i \sum_{\sigma,\rho\in S_{m-3}(2,\ldots,m-2)} A_m^{\text{tree}}(1,\sigma,m-1,m)\, S[\sigma|\rho]\widetilde{A}_m^{\text{tree}}(1,\rho,m,m-1)\,, \tag{84}$$

where we suppress an overall factor of $(\kappa/2)^{m-2}$. The formula makes use of a matrix $S[\sigma|\rho]$ known as the field-theory KLT kernel. This is an $(m-3)!\times(m-3)!$ matrix of kinematic polynomials that acts on the color-ordered amplitudes for $(m-3)!$ permutations of the external legs [24, 93, 185, 186]:

$$S[\sigma|\rho] = \prod_{i=2}^{m-2}\left[2p_1\cdot p_{\sigma_i} + \sum_{j=2}^{i} 2p_{\sigma_i}\cdot p_{\sigma_j}\theta(\sigma_j,\sigma_i)_\rho\right]\,, \tag{85}$$

where $\theta(\sigma_j,\sigma_i)_\rho = 1$ if σ_j is before σ_i in the permutation ρ, and zero otherwise. A compact definition, which reproduces equation (85) upon use of momentum conservation and on-shell conditions, can be given recursively [11] as [169],

$$S[A,j|B,j,C] = 2(p_1+p_B)\cdot p_j\, S[A|B,C]\,, \qquad S[2|2] = s_{12}\,, \tag{86}$$

where multiparticle labels $B = (b_1, b_2, \ldots, b_p)$ involve multiple external legs, and we use the notation $p_B = p_{b_1} + p_{b_2} + \ldots + p_{b_p}$. Using the recursive formula, we can obtain four-, five- and six-point KLT relations as particular cases.

In addition, the field-theory KLT kernel allows us to find explicit expressions for duality-satisfying tree-level numerators in the purely-adjoint case. The construction that we give here was independently worked out in references [24, 149]. The idea is to define the numerators for a subset of diagrams called half-ladder (or multi-peripheral) diagrams, whose structure is illustrated in figure 8. Corresponding to permutations of these half-ladder diagrams we specify $(m-2)!$ master numerators via,

$$\begin{aligned} n(1,\sigma(2,\ldots,m-2),m-1,m) &= -i\sum_{\rho\in S_{m-3}} S[\sigma|\rho]\widetilde{A}_m^{\text{tree}}(1,\rho,m,m-1)\,, \\ n(1,\tau(2,\ldots,m-1),m)\Big|_{\tau(m-1)\neq m-1} &= 0\,, \end{aligned} \tag{87}$$

[11] This recursive presentation of the KLT kernel has a string theory origin [24].

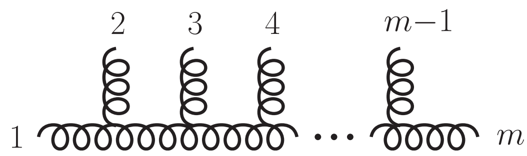


Figure 8. An m-point half-ladder tree diagram.

and take the remaining $(2m-5)!!-(m-2)!$ numerators to be determined by the Jacobi relations. By definition these numerators satisfy all the kinematic Jacobi relations and hence they obey the CK duality. However, to conclude that they define valid numerators we must also prove that they give correct amplitudes, both in gauge theory and in gravity.

Consider the DDM decomposition introduced in equation (55) for gauge-theory amplitudes with only adjoint particles. We use CK duality to replace the color factors in that formula with the above defined numerators,

$$\begin{aligned}\mathcal{M}_m^{\text{tree}} = \mathcal{A}_m^{\text{tree}}\Big|_{c_i\to n_i} &= \sum_{\tau\in S_{m-2}} A_m^{\text{tree}}\left(1,\tau\left(2,\ldots,m-1\right),m\right) n\left(1,\tau\left(2,\ldots,m-1\right),m\right)\\ &= \sum_{\sigma,\rho\in S_{m-3}} A_m^{\text{tree}}\left(1,\sigma,m-1,m\right) S[\sigma|\rho]\widetilde{A}_m^{\text{tree}}\left(1,\rho,m,m-1\right).\end{aligned} \tag{88}$$

On the first line we have a DDM decomposition for gravity amplitudes, where the half-ladder numerators play the same role as the half-ladder color factors in equation (55). On the second line we have plugged in the explicit numerators, and used the fact that only $(m-3)!$ of them are non-vanishing. As is obvious, the KLT formula (86) is reproduced. This implies that we get correct gravity amplitudes, given that both A_m^{tree} and $\widetilde{A}_m^{\text{tree}}$ are gauge-theory amplitudes. If we take A_m^{tree} to be amplitudes in bi-adjoint ϕ^3-theory and $\widetilde{A}_m^{\text{tree}}$ are YM amplitudes, then the above KLT formula gives back YM amplitudes. Hence the numerators in equation (87) give correct amplitudes.

This completes the constructive proof showing that CK duality can be satisfied for gauge theories with adjoint particles, given that all the BCJ amplitude relations (65) hold, which implies that the KLT formula hold. This argument relies on the availability of a DDM representation of the amplitude and on the existence of the field-theory KLT kernel. Pure YM theory, or $\mathcal{N}=1,2,4$ super-YM (SYM) theory are examples where tree-level CK duality is proven by this argument. Note, however, that the above numerators are nonlocal functions and the crossing symmetry of the amplitude does not follow automatically from relabeling the numerators. Hence it often desirable to find other representations of tree-level numerators. In specific cases, we can find representations of an amplitude with desired properties by imposing these properties on an ansatz whose coefficients are determined by requiring that it match the amplitude, a strategy also tremendously useful at loop level—see sections 3.2 and 6.

Similar considerations hold for amplitudes with multiple distinguishable adjoint scalars, although it may be necessary to introduce four-scalar interactions for the duality to hold [187]. As we discuss in section 2.6, imposing the duality on fermionic amplitudes implies supersymmetry.

2.4. Example 2: matter in fundamental representation

We now generalize the discussion in the previous subsection by introducing matter in the fundamental representation, as it appears in quantum chromodynamics (QCD) [188–190]. To be

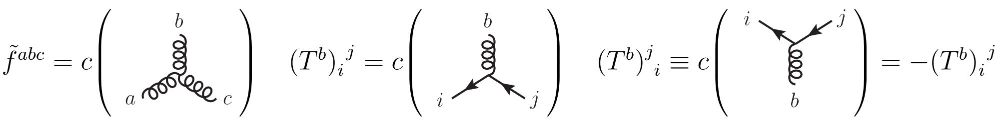


Figure 9. Color vertices with planar ordering consistent with the color-ordered Feynman rules.

Table 1. Number of cubic diagrams, $\nu(m,k)$, in the full m-point amplitude with k distinguishable quark-antiquark pairs and $(m-2k)$ gluons.

$k \setminus m$	3	4	5	6	7	8
0	1	3	15	105	945	10 395
1	1	3	15	105	945	10 395
2	—	1	5	35	315	3465
3	—	—	—	7	63	693
4	—	—	—	—	—	99

specific, let us consider YM theory with gauge group G and with N_f fundamental fermions[12]. For simplicity we call this theory QCD, given that it precisely matches QCD once we specify the gauge group to be $SU(3)$ and the number of quark flavors to be six; its Lagrangian is

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4}\left(F^a_{\mu\nu}\right)^2 + \overline{q}^{\alpha}\left(i\not{D} - M_{\alpha}^{\ \beta}\right) q_{\beta}\,, \quad \text{where} \quad D_\mu = \partial_\mu - i g A^a_\mu t^a\,, \tag{89}$$

where $\alpha, \beta = 1, \ldots, N_f$ are flavor indices, and spinor indices and fundamental gauge group indices are suppressed. The mass matrix $M_{\alpha}^{\ \beta}$ is taken to be diagonal. The only color tensors are in this case f^{abc} and $(t^a)_i^{\ j}$ which both have three free indices. Thus, all color factors will again correspond to cubic diagrams. A difference with the pure-adjoint case is that we now need to decorate the lines of the diagrams with the appropriate representation: adjoint, fundamental or anti-fundamental. This is illustrated in figure 9. A general color decomposition of tree-level amplit be found in [189] (see also [191–193]).

Without loss of generality we write the QCD m-point tree amplitude in terms of diagrams with cubic vertices,

$$\mathcal{A}^{\mathrm{tree}}_{m,k} = -i\, g^{m-2} \sum_{i\,\in \text{cubic diag.}}^{\nu(m,k)} \frac{c_i\, n_i}{D_i}\,, \tag{90}$$

where c_i are color factors, n_i are kinematic numerators, and D_i are denominators encoding the propagator structure of the cubic diagrams. The denominators (and numerators) may in principle contain masses, corresponding to massive quark propagators. For k quark-antiquark pairs and $(m-2k) > 0$ gluons, we may count the number of cubic diagrams. Assuming that the quarks are all of distinct flavor, one can then show that the number of nonzero diagrams is $\nu(m,k) = \frac{(2m-5)!!}{(2k-1)!!}$ [189]. As exemplified in table 1, the numbers grow modestly with the number of quarks.

Amplitudes with multiple quarks of the same flavor and mass can be obtained from distinct-flavor amplitudes by setting masses to be equal and summing over permutations of quarks

[12] A similar example of YM theory with scalars in matter representations will be discussed in section 5.2.

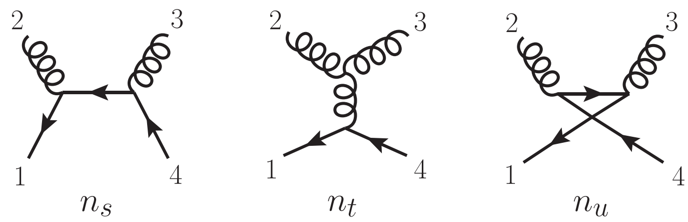


Figure 10. The diagrams contributing to the two-quark two-gluon amplitude.

with appropriate fermionic signs. Therefore, we do not lose generality by taking all k quark-antiquark pairs to have distinct flavor and mass. To be explicit, in table 1 we provide total counts $\nu(m,k)$ of cubic diagrams for different amplitudes up to eight particles and four quark pairs. It agrees with the usual counting of standard QCD Feynman diagrams restricted to those diagrams that only have trivalent vertices.

The color factors c_i in equation (90) are constructed from the cubic diagrams using only two building blocks: the structure constants $\tilde{f}^{abc}$ for three-gluon vertices and generators $(T^a)_i^{\ j}$ for quark-gluon vertices, as shown in figure 9. When separating color from kinematics, the diagrammatic crossing symmetry only holds up to signs dependent on the permutation of legs. These signs are apparent in the total antisymmetry of $\tilde{f}^{abc}$. For a uniform treatment of the fundamental representation, it convenient to introduce a similar antisymmetry for the fundamental generators,

$$(T^a)^j{}_i \equiv -(T^a)_i^{\ j} \quad \Leftrightarrow \quad \tilde{f}^{cab} = -\tilde{f}^{bac} \,. \tag{91}$$

This allows us to introduce a similar antisymmetry in color-ordered kinematic vertices, so that they are effectively the same as for the adjoint representation. As noted in equation (40) the color factors obey Jacobi and commutation identities. They both imply color-algebraic relations of the form given in equation (41), and differ only by the subdiagrams as drawn in figure 6, but otherwise have common diagram structure. The interdependence among the color factors c_i means that the corresponding kinematic coefficients n_i/D_i are in general not unique, as reflected by the underlying gauge dependence of the numerators.

A first interesting example of an amplitude is the four-point amplitude for two gluons and a quark-antiquark pair displayed in figure 10[13],

$$\mathcal{A}^{\text{tree}}_{4,1}\left(1\bar{q},2g,3g,4q\right) = -i\,g^2\left(\frac{n_s c_s}{s-m_q^2} + \frac{n_t c_t}{t} + \frac{n_u c_u}{u-m_q^2}\right), \tag{92}$$

where the numerators are

$$\begin{aligned} n_s &= \frac{1}{2}\bar{u}_1 \not{\varepsilon}_2 \left(\not{p}_{12}\delta^{\alpha_1}_{\alpha_4} - M^{\alpha_1}_{\alpha_4}\right)\not{\varepsilon}_3 v_4\,, \quad n_u = \frac{1}{2}\bar{u}_1 \not{\varepsilon}_3 \left(\not{p}_{13}\delta^{\alpha_1}_{\alpha_4} - M^{\alpha_1}_{\alpha_4}\right)\not{\varepsilon}_2 v_4\,, \\ n_t &= n_u - n_s\,, \end{aligned} \tag{93}$$

and the color factors

$$c_s = \left(T^{a_3}T^{a_2}\right)_{i_4}^{\ i_1}\,, \quad c_u = \left(T^{a_2}T^{a_3}\right)_{i_4}^{\ i_1}\,, \quad c_t = c_u - c_s = \tilde{f}^{\,a_2a_3b}\left(T^b\right)_{i_4}^{\ i_1}\,. \tag{94}$$

Here the Greek indices α_1, α_2 are global (flavor) indices carried by the fermions.

[13] When useful, we use the slightly-nonstandard notation $\mathcal{A}_n\left(1\Phi_1,\ldots,n\Phi_n\right)$ to display explicitly the external states in an amplitude.

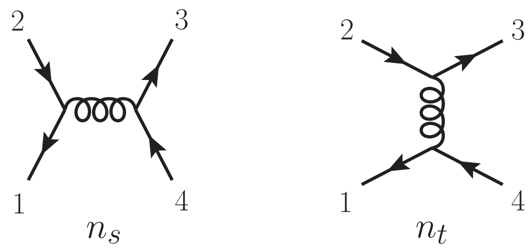


Figure 11. The diagrams contributing to the four-point pure-quark amplitude.

Following the same steps as in pure YM theory one can show that the numerator relation together with the kinematics constraints at four points imposes massive BCJ relations for the partial amplitudes,

$$\left(s-m_q^2\right) A_{4,1}^{\text{tree}}\left(1\bar{q},2g,3g,4q\right) = \left(u-m_q^2\right) A_{4,1}^{\text{tree}}\left(1\bar{q},3g,2g,4q\right) . \tag{95}$$

More generally, one can understand this relation as a consequence of gauge redundancy. We have two independent numerators, which are not invariant under gauge transformations. We can thus at most build one gauge-invariant quantity out of these, and hence all partial amplitudes must be related.

At general multiplicity m, the BCJ amplitude relations in their simplest incarnation take the form,

$$\sum_{i=2}^{m-1} p_1\cdot\left(p_2+\ldots+p_i\right) A_{m,k}^{\text{tree}}\left(2,\ldots,i,1g,i+1,\ldots,m\right) = 0 , \tag{96}$$

where leg 1 must be a massless gluon in the adjoint. Unlike equation (60), here the particles $2,\ldots,n$ may have any spin, mass, and gauge-group representation. The partial amplitude is constructed as a sum over planar Feynman graphs in the same fashion as for the purely adjoint case; however, the color decomposition for these mixed adjoint-generic-representation amplitudes is quite different. See [189, 191–193] for details.

As a further nontrivial example at four points, consider the fundamental representation four-quark amplitude displayed in figure 11. This amplitude is given as a sum over two displayed diagrams,

$$\mathcal{A}_{4,2}^{\text{tree}}\left(1\bar{q},2q,3\bar{q},4q\right) = -ig^2\left(\frac{n_s c_s}{s}+\frac{n_t c_t}{t}\right), \tag{97}$$

where the color factors are

$$c_s = \left(T^a\right)_{i_2}^{\ i_1}\left(T^a\right)_{i_4}^{\ i_3} , \quad c_t = \left(T^a\right)_{i_4}^{\ i_1}\left(T^a\right)_{i_2}^{\ i_3} , \tag{98}$$

and the kinematic factors are

$$n_s = -\frac{1}{2}\left(\bar{u}_1\gamma_\mu v_2\right)\left(\bar{u}_3\gamma^\mu v_4\right)\delta_{\alpha_2}^{\alpha_1}\delta_{\alpha_4}^{\alpha_3}, \quad n_t = -\frac{1}{2}\left(\bar{u}_1\gamma_\mu v_4\right)\left(\bar{u}_3\gamma^\mu v_2\right)\delta_{\alpha_4}^{\alpha_1}\delta_{\alpha_2}^{\alpha_3} . \tag{99}$$

For this amplitude, neither the color nor the kinematic factors satisfy any relations among themselves, hence CK duality is trivially satisfied. Indeed, each kinematic numerator is gauge invariant by itself and thus the amplitude representation is necessarily unique. We will however see in section 5 that in some cases it is possible or even necessary to impose additional numerator relations for matter amplitudes without external gluons.

We now look more in detail at the theory obtained from the double-copy formula with two sets of QCD numerators. It will consist of gravity coupled to a single massless complex scalar, as well as a set of massive photons and scalars. Massless and massive fields in this theory originate from the double copy of adjoint and fundamental gauge-theory fields, respectively. As an example, we give the amplitude between four massive (complex) photons γ,

$$\mathcal{M}_4^{\text{tree}}(1\bar{\gamma},2\gamma,3\bar{\gamma},4\gamma) = -i\left(\frac{n_s\left(n_s|_{N_f\to 1}\right)}{s}+\frac{n_t\left(n_t|_{N_f\to 1}\right)}{t}\right), \tag{100}$$

where we trivialize the number of flavors on one side in order to avoid a redundant description with a factorized flavor group in the gravitational theory.

Writing out the expression we have

$$\begin{aligned}\mathcal{M}_4^{\text{tree}}(1\bar{\gamma},2\gamma,3\bar{\gamma},4\gamma) = -i\Bigg\{&\frac{\left[(\bar{u}_1\gamma_\mu v_2)(\bar{u}_3\gamma^\mu v_4)\right]^2}{4s}\delta_{\alpha_2}^{\ \alpha_1}\delta_{\alpha_4}^{\ \alpha_3}\\ &+\frac{\left[(\bar{u}_1\gamma_\mu v_4)(\bar{u}_3\gamma^\mu v_2)\right]^2}{4t}\delta_{\alpha_4}^{\ \alpha_1}\delta_{\alpha_2}^{\ \alpha_3}\Bigg\}.\end{aligned} \tag{101}$$

The square can be upgraded to a tensor product since the external spinors can be chosen differently for the two numerator copies.

In order to better understand the double-copy amplitude, we may write it in terms of chiral spinors and explicitly write out the little group indices. For example, using the massive spinor-helicity variables reviewed in appendix B, we simplify the above expression to obtain

$$\begin{aligned}\mathcal{M}_4^{\text{tree}}\left(1\bar{\gamma}^{aa'},2\gamma^{bb'},3\bar{\gamma}^{cc'},4\gamma^{dd'}\right) = -i\Bigg\{&\Big(\langle 1^a 3^c\rangle\left[2^b 4^d\right]+[1^a 3^c]\langle 2^b 4^d\rangle+\langle 1^a 4^d\rangle\left[2^b 3^c\right]\\ &+\left[1^a 4^d\right]\langle 2^b 3^c\rangle\Big)^2\frac{\delta_{\alpha_2}^{\alpha_1}\delta_{\alpha_4}^{\alpha_3}}{s}+(2\leftrightarrow 4)\Bigg\}.\end{aligned} \tag{102}$$

2.5. Double copy implies diffeomorphism symmetry

Why does the double copy of gauge-theory amplitudes yield amplitudes of some gravity theory? A minimal criterion is that the expression obtained from the double-copy method be invariant under linearized diffeomorphisms. Here we show that invariance of the double-copy amplitudes under linearized diffeomorphisms is a direct consequence of color-kinematics duality and gauge invariance of the two gauge-theory factors entering the construction.

We start from a general linearized gauge transformation acting on a single external gluon with momentum p. Its polarization vector transforms as: $\varepsilon_\mu(p)\to\varepsilon_\mu(p)+p_\mu$. Gauge invariance of the amplitude implies that every diagram numerator should shift as

$$n_i\to n_i+\delta_i\,,\qquad \delta_i = n_i\Big|_{\varepsilon\to p}. \tag{103}$$

Then, the entire amplitude is unaffected provided that the shifts δ_i obey

$$\sum_i\frac{c_i\delta_i}{D_i}=0, \tag{104}$$

which must hold since by assumption the gauge-theory amplitude is gauge invariant.

Aside from the explicit expressions for the numerator factors, the above equation must rely exclusively on the generic algebraic properties of the color factors c_i, namely antisymmetry and Jacobi identities. This means that, if we have CK-duality-satisfying numerators $\tilde{n}_i$ in some gauge theory and consider their double copy with another set of gauge numerators n_i, then any linearized gauge transform of the n_i will leave the double-copy amplitude invariant:

$$\sum_i \frac{n_i \delta_i}{D_i} = 0\,. \tag{105}$$

We now analyze in more detail the significance of this transformation. A general coordinate transformation can be used to impose both transversality and tracelessness on the on-shell asymptotic states that enter the definition of a scattering amplitude. This, in turn, results in imposing the conditions $\varepsilon_{\mu\nu}(p)p^\nu = 0 = \varepsilon_{\mu\nu}(p)\eta^{\mu\nu}$ on the graviton's polarization tensor. After this choice of gauge, amplitudes will still be invariant under the subset of linearized diffeomorphisms that do not modify the above conditions. These will act as

$$\varepsilon_{\mu\nu}(p) \to \varepsilon_{\mu\nu}(p) + p_{(\mu}q_{\nu)}\,, \tag{106}$$

where q is a reference vector that obeys $p\cdot q = 0$, but is otherwise generic. The parenthesis denote symmetrization of spacetime indices.

The first step of formulating a double-copy construction is to establish a map between gravity asymptotic states and pairs of gauge-theory states. In general, the double-copy graviton will be obtained by taking the symmetric-traceless part of the product of the two gauge-theory gluons, i.e. its polarization tensor will be obtained from the gluon's polarizations as $\varepsilon_{\mu\nu} = \varepsilon_{((\mu}\tilde{\varepsilon}_{\nu))}$, where the double brackets indicate the symmetric-traceless part[14].

We now study tree-level amplitudes obtained from the double-copy method. We take a set of duality-satisfying numerators n_i only for one of the gauge-theory factors. The other set of numerators is taken in the form

$$\tilde{n}_i' = \tilde{n}_i + \tilde{\Delta}_i\,, \qquad\qquad \sum_i \frac{\tilde{\Delta}_i c_i}{D_i} = 0\,. \tag{107}$$

While the numerators $\tilde{n}_i'$ can violate CK duality, they can be obtained from a set of duality-satisfying numerators $\tilde{n}_i$ with a transformation of the form (42) with parameters $\tilde{\Delta}_i$. Hence, we are assuming that there exists an amplitude presentation for which the duality is satisfied also for the second gauge theory. However, in the double-copy method, we use a set of numerators with different properties for one of the theories, a fact that will be advantageous in practical calculations.

Starting from the double-copy gravity amplitude in equation (45) at tree level, a tree amplitude can then be expressed as

$$i\mathcal{M}_m = \sum_i \frac{n_i \tilde{n}_i}{D_i} + \sum_i \frac{n_i \tilde{\Delta}_i}{D_i} = \sum_i \frac{n_i \tilde{n}_i}{D_i}\,, \tag{108}$$

where we have not included the overall $(\kappa/2)^{n-2}$. Because numerator factors n_i obey the same algebraic relations as the color factors c_i, equation (107) implies the last equality above. Using equation (108), the variation of the double-copy amplitude under a linearized diffeomorphism of the form (106) becomes

$$i\mathcal{M}_m \to i\mathcal{M}_m + \sum_i \frac{\delta_i\ \tilde{n}_i|_{\tilde{\varepsilon}\to q}}{D_i} + \sum_i \frac{n_i|_{\varepsilon\to q}\tilde{\delta}_i}{D_i}\,. \tag{109}$$

[14] We also note that the antisymmetric and trace parts of the product of the two gauge-theory gluon polarizations are identified with an antisymmetric tensor field and the dilaton. These two field are generically present in amplitudes from the double copy unless additional steps are taken to ensure their removal, as we will see in section 5.3.4.

The two terms are of the form (105) and hence vanish because of CK duality. We then conclude that invariance of the amplitude under linearized diffeomorphisms at tree level follows from gauge invariance of the gauge theories entering the double-copy construction provided that CK duality is obeyed. Diffeomorphism invariance of the amplitudes at loop level can also be established through generalized unitarity [194]. We will see in the next subsection and in sections 4 and 5 that the double copy can also be used to engineer amplitudes which are invariant under other symmetries, including supersymmetry and gauge symmetry. In fact, one can think of the double copy as a clever procedure to write down amplitudes that obey a prescribed set of on-shell Ward identities starting from gauge-theory data. By construction, these amplitudes also obey standard factorization properties as well as crossing symmetry. The basic intuition is that gauge invariance together with mild assumptions on the singularity structure are sufficient to fix the form of amplitudes [173, 174].

2.6. *Adjoint fermions + duality ⇒ supersymmetry*

In section 2.4 we discussed CK duality in the context of YM theory with matter fermions. We now look at the case of adjoint fermions in arbitrary dimension. In this case, we will see that the duality is equivalent to the existence of supersymmetry, as argued in [30] (see also [117] for a related discussion). For concreteness, we specialize to D-dimensional YM theory minimally coupled to a single adjoint Majorana fermion, described by the Lagrangian

$$\mathcal{L} = \mathrm{Tr}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{i}{2}\bar{\psi}\not{D}\psi\right]. \tag{110}$$

In all dimensions, the four-gluon and two-gluon-two-fermion amplitudes respect the duality between color and kinematics without any further constraint. However, four-fermion amplitudes leads to an interesting constraint. This amplitude is given by

$$\begin{aligned}\mathcal{A}_4^{\text{tree}}(1\psi,2\psi,3\psi,4\psi) = i\Bigg(&\frac{(\bar{u}_1\gamma_\mu v_2)(\bar{u}_3\gamma^\mu v_4)\,c_s}{2s} + \frac{(\bar{u}_2\gamma_\mu v_3)(\bar{u}_1\gamma^\mu v_4)\,c_t}{2t} \\ &+ \frac{(\bar{u}_3\gamma_\mu v_1)(\bar{u}_2\gamma^\mu v_4)\,c_u}{2u}\Bigg),\end{aligned} \tag{111}$$

where the $\bar{u}_i$ and v_i are spinor external states which obey $\bar{u}_i\gamma^\mu v_j = \bar{u}_j\gamma^\mu v_i$ due to the Majorana condition, $\bar{u}_i = v_i^T\mathcal{C}$. In dimensions in which a Weyl representation can be chosen one of the terms above vanishes.

The requirement that $\mathcal{A}_4^{\text{tree}}(1\psi,2\psi,3\psi,4\psi)$ obeys CK duality forces the gamma matrices to obey the relation

$$(\bar{u}_1\gamma_\mu v_2)(\bar{u}_3\gamma^\mu v_4) + (\bar{u}_2\gamma_\mu v_3)(\bar{u}_1\gamma^\mu v_4) + (\bar{u}_3\gamma_\mu v_1)(\bar{u}_2\gamma^\mu v_4) = 0. \tag{112}$$

Equation (112) is the equivalent to the Fierz identity that appears in the supersymmetry transformation of the Lagrangian (110). This analysis can be repeated for pseudo-Majorana spinors with analogous results. Overall, an identity of this form can be satisfied only for $D = 3,4,6,10$, i.e. the dimensions for which the theory (110) is supersymmetric.

The relation between CK duality with adjoint fermions and supersymmetry should not seem surprising in hindsight. In principle, adjoint fermions can be combined with gluons with the double-copy procedure, resulting in a gravity theory which includes spin-3/2 fields. However, it is known that local supersymmetry is required to have a consistent theory of interacting spin-3/2 fields. This is another example of the duality between color and kinematics underpinning the consistency of the gravity theory from the double copy. We shall see more examples along this line in the following sections.

By repeating the discussion in section 2.5, it is straightforward to see that the double copy of a gauge theory with another gauge theory that has global supersymmetry leads to a theory that exhibits local supersymmetry. Indeed, a linearized local supersymmetry transformation of a gravitino polarization vector-spinor $u^{\alpha}_{\mu}(p)$ is

$$u^{\alpha}_{\mu}(p) \to u^{\alpha}_{\mu}(p) + p_{\mu}\xi_{\alpha}\,, \tag{113}$$

where ξ_α is the transformation parameter which, in order to preserve the γ-tracelessness of the gravitino wave function must obey the massless Dirac equation, $\not{p}\xi = 0$. Then, the transformation of a double-copy amplitude (after a discussion similar to the one that led to equation (109)) is

$$\mathcal{M}^{\text{tree}}_m \to \mathcal{M}^{\text{tree}}_m + \sum_i \frac{\delta_i \tilde{n}_i\,|_{\tilde{u}^{\alpha}\to\xi^{\alpha}}}{D_i}\,, \tag{114}$$

where $\tilde{u}^{\alpha}$ is the spinor that, through the double copy, generates the gravitino under consideration. Since the parameter ξ of the supersymmetry transformation has the same properties as the original spinor $\tilde{u}^{\alpha}$ it replaces, the factor $\tilde{n}_i\,|_{\tilde{u}^{\alpha}\to\xi^{\alpha}}$ has the same properties as $\tilde{n}_i$, in particular it obeys Jacobi relations. Thus, the variation of the double-copy amplitude under (linearized) supersymmetry transformations vanishes, implying that the double-copy theory exhibits local supersymmetry.

More generally, we may expect that, under the right circumstances, the double copy of a gauge theory with a theory that exhibits a global symmetry leads to a theory where the global symmetry is promoted to a local symmetry. We shall return to this point in section 4.

The emergence of supersymmetry from CK duality offers a novel perspective on the maximal number of gravitini that can consistently enter a supergravity theory. As we have seen, a gauge theory coupled to fermions can exhibit CK duality in at most ten dimensions dimensions. Thus, this is the highest dimension which a supergravity theory can be given a double-copy interpretation in the sense described here. Taking two such theories gives therefore the largest number of supersymmetries, which is two in ten dimensions or, upon dimensional reduction, eight in four dimensions. This observation recovers the usual bound following from the requirement that the exist multiplets of supersymmetry algebra containing fields of spin $s \leqslant 2$.

2.7. General lessons from applying CK duality

In the previous subsections, we presented various concrete examples of theories which obey CK duality. Statements about CK duality often depend on the details of the theories under consideration and on what observables are being studied. Since the duality is often used as a shortcut for computing gravitational amplitudes, one can restrict to gauge theories suitable for giving broad classes of consistent gravitational theories once the numerators are assembled via the double-copy method. Expanding on the examples discussed earlier, in the rest of this review we will focus mostly on theories with the following general features:

- There exists (at most) one massless gauge field, the gluon, that transform in the adjoint of a gauge group G, and all fields of the gauge theory are charged under this group. We will see in section 5 that this requirement translates to the equivalence principle in gravity.
- The gauge group G is a completely general Lie group in the sense that no assumptions on its rank need be imposed on it. Note that throughout this section the only properties of the gauge group we have utilized are the Jacobi relations of its structure constants and the commutation relations of its representation matrices, which do not require to spell out our choice of Lie group.

- Amplitudes involving adjoint fields (gluons or adjoint matter) should admit perturbative expansions where the kinematic numerators obey the same Lie algebra relations (e.g. Jacobi identities) as the corresponding adjoint-valued color factors. We have seen in section 2.5 that this property is essential for obtaining a gravitational theory after the double copy. This condition also implies the universality of gravitational self-interactions.
- Amplitudes involving fields in generic representations of the gauge group should admit perturbative expansions where the kinematic numerators obey the same Lie algebra relations as the generators of those representations. The simplest example of non-purely-adjoint theory has been discussed in section 2.4.

These general properties guarantee that every diagram in the perturbative expansion of an amplitude has a unique nontrivial color factor, which obeys the minimal constrains imposed by the Lie algebra of the gauge group, and furthermore that the coupling to the unique gluon is universally controlled by the gauge-group representations. The kinematic factors can then be constrained to obey the duality by enforcing the one-to-one map between color and kinematic identities. Along these lines, in section 5.1 we will articulate a more precise set of working rules which will define the properties of the gauge theories employed for obtaining a web of double-copy-constructible theories.

3. Geometric organization

The dual Jacobi identities give nontrivial relations between diagram numerators. Here we describe the systematics of these relations and how they can be used to express amplitudes' integrands in terms of the contributions of a small set of master diagrams. This is generally very helpful at higher perturbative orders because it allows us to express an integrand in terms of a (small) subset of all of its terms. To this end, we will first describe a useful geometric organization via a graph of graphs[15] that offers insight into the information flow of the duality identities. We will then illustrate the general case through some examples.

3.1. Amplitudes in terms of boundary data

The duality between color and kinematics provides a set of relations between diagrams. Section 2 frames the discussion of the duality in terms of vector and matrix operations between linear spaces of numerators of diagrams and linear spaces of scattering amplitudes. Here we give an alternative perspective, using the language of graphs [196, 197]. This offers a useful way to visualize how a small set of graphs is sufficient to describe the entire amplitude. We shall see that the minimal set of graphs whose numerators need to be specified can be thought of as boundary data on the graph of graphs describing the amplitude.

As a simple example, consider the four-gluon tree amplitude. After absorbing any contact terms into graphs with only cubic vertices, this amplitude can be described by the three graphs in figure 3, corresponding to the s-channel, t-channel, and u-channel. Both the kinematic and color numerators of these three graphs satisfy the (dual) Jacobi identities in equations (16) and (17), $c_s + c_t + c_u = 0$ and $n_s + n_t + n_u = 0$. These equations are equivalent to the statement that any two 'single-copy' numerator dressings determine the third one. We can draw this relationship as a graph of graphs, where each vertex represents a specific graph participating

[15] See [195] for a related application of such an approach towards identifying *scattering forms* of amplitudes.

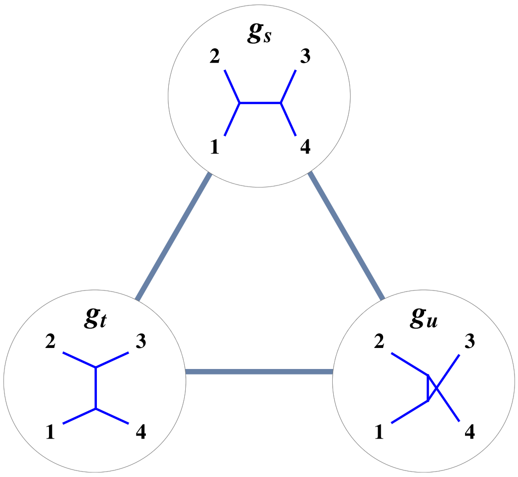


Figure 12. Graph of graphs relevant to four-point tree-level scattering forms a triangle. Each vertex represents one of the three graphs in figure 3, and every edge represents a Jacobi or Whitehead move (cf figure 13). Every triangle in the graph of graphs represents a Jacobi relation that can be used to constrain a dressing of one vertex graph in terms of dressings of the other two vertex graphs.

in the Jacobi relation and the edges connect to the other two vertices (i.e. graphs) that determine the first vertex.

These relations can be summarized in the triangle shown in figure 12. Each vertex or node in figure 12 is one of the three cubic graphs from figure 3 that would contribute to the four-point amplitude. Every edge in this graph of graphs corresponds to a Jacobi move on the internal propagator of one of the vertex graphs that transmutes it into another. In the mathematics literature these moves are known as Whitehead moves [198]. The basic moves for acting on graphs represented by the edges in figure 12 are denoted $\hat{t}$ and $\hat{u}$ and are shown in figure 13. The first move, which we call $\hat{t}$, takes the s-channel graph in figure 3 and converts it to the t-channel. Similarly, we call $\hat{u}$ the move that converts the s-channel graph to the u-channel one. The move that takes the t-channel graph to the u-channel graph can be understood as the composition $\hat{u}\circ\hat{t}$. Alternatively, we can view it as one of the same basic operations as on the s-channel graph but acting on a graph with permuted labels. Strictly speaking, one should associate a direction with each move, but we will ignore this distinction because, up to relabelling, the reverse operation is identical to the forward one. It is not difficult to see that each edge in the triangle graph of graphs contains the graphs contributing to a particular four-point color-ordered partial amplitude, and the entire triangle itself represents an occasion for Jacobi to be satisfied by a dressing of the graphs (whether color or kinematic).

This basic structure generalizes straightforwardly to higher-point amplitudes. Consider it at five points: in total there are 15 distinct cubic graphs (constructible by starting with one five point cubic graph and applying $\hat{t}$ and $\hat{u}$ on each of it is internal edges, and repeating on new graphs until closure) contributing to the full color-dressed integrand as displayed in figure 14.

Exercise 3.1. Draw individual five-point graphs for each vertex in figure 14. Label the edge operations to get from graph to graph.

Let us focus on the subset of five graphs comprising the ordered partial amplitude $A_5^{\text{tree}}(1,3,5,4,2)$, as indicated in equation (71). As shown in figure 15, all five graphs can

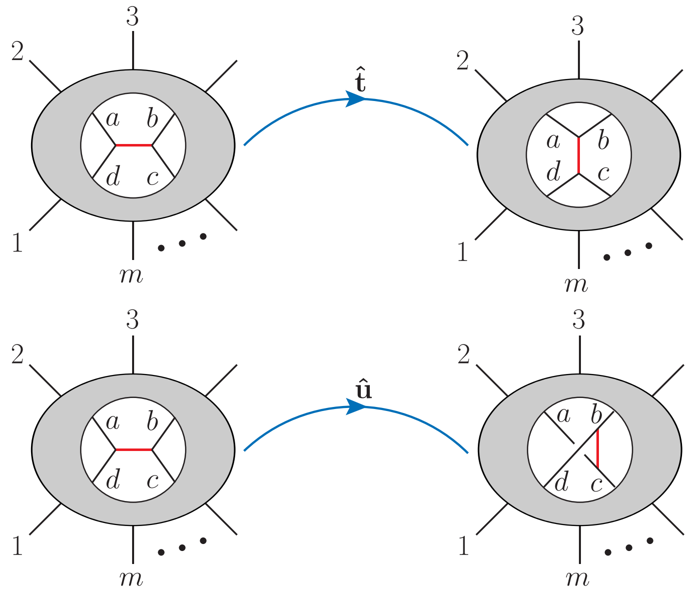


Figure 13. Operations that relate edges of one graph to edges of another.

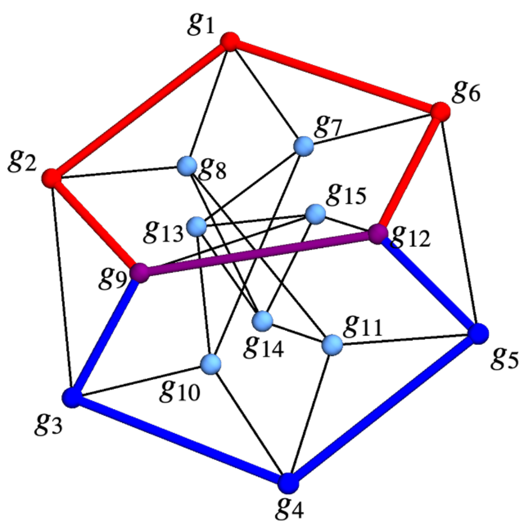


Figure 14. Graph of graphs for the full color-dressed five-point partial amplitude, with vertex labels corresponding to the graphs given in figure 7. Two color-ordered partial amplitude graphs are highlighted, corresponding to $A_5^{\text{tree}}(1,3,5,4,2)$ and $A_5^{\text{tree}}(1,3,5,2,4)$, cf equations (71) and (72) respectively, as well as figure 15.

be found by starting with any graph that has the relevant external ordering and applying $\hat{t}$ to its two internal edges to find and connect two other graphs with the same color order. It is interesting to note that each edge in the graph represents a shared factorization channel for the internal propagator not mutated between the two connected graphs. We keep applying $\hat{t}$ until closure. This procedure of repeated $\hat{t}$ application, building a graph of all the graphs of a

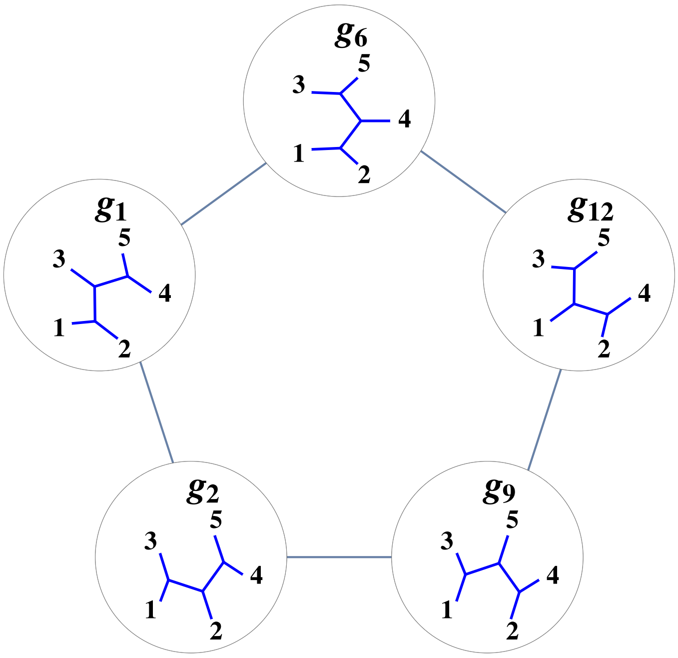


Figure 15. Graph of graphs relevant to the color-ordered five-point partial amplitude $A_5^{\text{tree}}(1,3,5,4,2)$. The vertices correspond to the indicated five-point graphs given in figure 7, with the understanding that each labeled graph contributes to this color order with signs determined by its color factor as in equation (71). The edges in the color-ordered graph of graph represent application of $\hat{t}$ on vertex-graph edges until closure.

given color-order, carves out the skeletal graph[16] of a polytope known as the *associahedron*. (see [200, 201]). Associahedra are often also called Stasheff polytopes. As $\hat{t}$ preserves color order, each Stasheff skeletal graph is composed of all the graphs that contribute to a fixed color order [197]. Indeed every Stasheff subgraph of the full graph of graphs represents the contributions of a particular ordered partial amplitude to the full amplitude. This is exemplified by the two highlighted pentagonal subgraphs of figure 14 which represent the ordered amplitudes $A_5^{\text{tree}}(1,3,5,2,4)$ and $A_5^{\text{tree}}(1,3,5,4,2)$.

Exercise 3.2. Find another pentagonal subgraph in figure 14 besides the two highlighted ones. Which color-ordered amplitude does it represent?

Of course, the full five-point amplitude requires all 15 cubic graphs (cf equation (50)), as displayed in the complete five-point graph of graphs figure 14. Every triangle subgraph represents a Jacobi identity that single-copy numerator dressings could satisfy.

One question when presented with the graph of graphs, is whether it is easy to see how many ordered amplitude (Stasheff) subgraphs are required to specify a full color-dressed amplitude. A natural conclusion is $(m-2)!$ of them, because this is the minimal number of ordered amplitude subgraphs whose union of vertex-graphs includes all $(2m-5)!!$ vertex-graphs that contribute to the full amplitude. In terms of our five-point graph of graphs, drawn in figure 14, it would be necessary to identify six different pentagonal subgraphs for every vertex-graph to be

[16] Typically called a one-skeleton (see [199]).

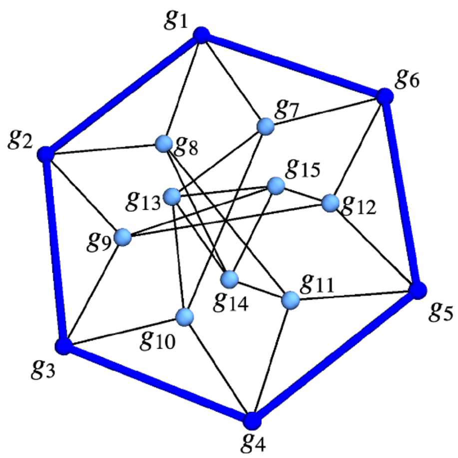


Figure 16. Due to the nine Jacobi relations (six outer triangles, and three inner triangles), only the six outer boundary graphs are needed to specify bulk data via Jacobi relations.

included at least once. This corresponds to the KK [181] basis, which gives the full amplitude in terms of partial amplitudes [180] as per equation (55).

Exercise 3.3. Find six color-ordered partial amplitudes (pentagonal subgraphs) in the five-point graph of graphs shown in figure 14 that together include every node at least once. What color-ordered amplitudes do they correspond to? Is it a KK basis?

The counting works differently for an amplitude where both the kinematic numerators and the color factors obey Jacobi identities. Consider the external boundary graphs of figure 16. As noted above, Jacobi identities are represented by triangles in the graph of graphs. By working inwards using Jacobi identities in figure 16, we see that each pair of neighboring graphs in the six external boundary graphs completely specify all other graphs. So, for theories that can satisfy the dual Jacobi identities, we need only specify data on this external boundary. More generally, for an m-point amplitude, it turns out that only $(m-2)!$ boundary graph numerators are sufficient to specify all other graph numerators via Jacobi and $(m-3)!$ ordered amplitudes are independent. For example, in the case of the five-point amplitudes, the two highlighted partial amplitudes in figure 14 are sufficient to generate all others.

The spanning boundary graphs at tree level, sometimes referred to as the master graphs, have a geometric association as well. Consider the half-ladder graph shown in figure 8. A half-ladder graph (also called a multiperipheral graph) is a cubic tree-graph where all vertices but two connect two internal edges. They are called half-ladders because they can be drawn to resemble one half of a ladder split down all rungs. For any multiplicity, a choice of master graphs can be obtained by taking any half-ladder graph and acting with the $\hat{u}$ move on all internal edges until no new graphs are generated. All graphs so generated will remain half-ladders with different labels and their corresponding graph of graphs forms the one-skeleton of a polytope known as a *permutahedron* [201].

Exercise 3.4. Why is at least one half-ladder graph required at any multiplicity in the set of Jacobi master-graphs?

As discussed in section 2, the mismatch between the number of independent gauge-theory amplitudes and the number of independent numerators leads to a gauge freedom that allows some numerators to take on arbitrary values, neither altering the amplitudes nor the BCJ amplitude relations. At five points, four of the numerators can be arbitrarily chosen, i.e. even set to zero, making the remaining numerators nonlocal. In addition, the fact that there are only $(m-3)!$ independent gauge-theory amplitudes is directly related to the fact that in the KLT formula (84) only $(m-3)!$ independent partial amplitudes appear for each of the two gauge theories.

As typical with graphical organization of amplitudes, the total number of independent graphs will increase factorially, going as $(m-2)!$ for multiplicity m. While this is a reduction over the total $(2m-5)!!$ cubic graphs that contribute to the complete amplitude, a more useful question is the minimal information required to build the complete amplitude (at tree level, and, more generally, the amplitude's integrand at loop level). Remarkably, as we now show, by imposing diagram symmetry, we can specify only a single half-ladder diagram, which then determines all other diagrams at any given multiplicity at tree level. This allows us to avoid specifying a factorially-growing number of diagrams.

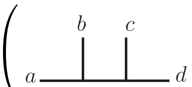
3.2. *Applying relabeling invariance at tree-level*

As a warm-up before turning to gauge theory, we consider the NLSM [202], as defined by the Lagrangian in the Cayley parameterization [169, 203, 204],

$$\mathcal{L}_{\mathrm{NLSM}} = \frac{1}{2}\mathrm{Tr}\left\{\partial_\mu\varphi\,\frac{1}{1-\lambda\varphi^2}\,\partial^\mu\varphi\,\frac{1}{1-\lambda\varphi^2}\right\}, \tag{115}$$

where φ is a Lie-algebra valued Goldstone-boson scalar field in the adjoint representation. Here the color symmetry is global. Although this theory has only even-point interactions, we can assign its data to graphs with only cubic vertices by multiplying and dividing by appropriate inverse propagators. In terms of diagrams with only cubic vertices, dimensional analysis dictates that each vertex effectively carries two powers of momentum. The first nonvanishing amplitude is at four points. We will see that, by imposing Jacobi identities and relabeling symmetry on the kinematic numerators of the diagrams, we can obtain the four-point scattering amplitude of this theory. At higher points, we will also either need to impose additional conditions to uniquely fix the amplitudes to this theory. It is sufficient to impose the manifestation of only quartic poles in amplitudes [205], which in combination with color-kinematics encodes the necessary [206–208] vanishing soft-scalar limits, or *Adler zero* conditions [209]. Higher-derivative deformations of the NLSM and their compatibility with the KK and the BCJ amplitudes relations have been discussed in [210].

To start the construction of the four-point amplitude, consider the four-point half-ladder graph. Since we require the dimensions of the numerator to match that following from the NLSM Lagrangian and therefore carry four powers of momentum, we take a numerator ansatz:

$n(a,b,c,d) \equiv n\Bigg($  $\Bigg) = s_{ab}(\alpha s_{ab} + \beta s_{bc}),$ (116)

where $s_{ab} = (p_a + p_b)^2$ and α and β are to be constrained by symmetry and the kinematic Jacobi identities. The other kinematic invariant is $s_{ac} = -s_{ab} - s_{bc}$, and it is not independent.

Imposing the Jacobi constraints,

$$n(a,b,c,d) = n(c,a,d,b) + n(d,a,b,c)\,, \tag{117}$$

relates α and β according to

$$\begin{aligned} 0 &= \alpha s_{ab}^2 - \alpha s_{bc}^2 - \beta s_{bc} s_{ac} - \alpha s_{ac}^2 \\ &= -2\alpha s_{ab} s_{bc} + \beta s_{ab} s_{bc} - 2\alpha s_{bc}^2 + \beta s_{bc}^2 \,, \end{aligned} \tag{118}$$

where we used $s_{ac} = -s_{ab} - s_{bc}$. Given that the Mandelstam invariants s_{ab} and s_{bc} are independent, we find $\beta = 2\alpha$; thus, the numerator is uniquely fixed up to its overall scale, which may be identified as the coupling of the model:

$$n(a,b,c,d) \propto s_{ab}\left(s_{ab} + 2s_{bc}\right) . \tag{119}$$

One can verify that this expression satisfies all necessary antisymmetry constraints, e.g. it changes sign with $a \leftrightarrow b$ or $c \leftrightarrow d$. The full amplitude, up to overall normalization is then

$$\begin{aligned} A_{\mathrm{NLSM}}^{\mathrm{tree}} \propto \frac{c(1,2,3,4)\, n(1,2,3,4)}{s_{12}} + \frac{c(3,1,4,2)\, n(3,1,4,2)}{s_{13}} \\ + \frac{c(4,1,2,3)\, n(4,1,2,3)}{s_{14}} \,, \end{aligned} \tag{120}$$

where the color factors are obtained by dressing each vertex with a structure constant f^{abc}. The key lesson is that, by taking the numerators to be functions of the graph labels, only a single numerator needs to be specified.

Exercise 3.5. Repeat the above analysis assuming only degree-one monomials in the Mandelstam invariants. What theory could this construction correspond to?

Next, consider the case of YM theory at four points. In this case the numerator ansatz is constructed out of external momenta and polarization vectors $\{\varepsilon_1, \ldots, \varepsilon_4\}$, subject to the requirements that every ε_i appears once in each term and that every term has exactly two momenta. These constraints guarantee consistency with the structure of Feynman rules. The possible third-degree monomials are constructed from the following independent Lorentz invariants:

$$\begin{aligned} &\{s_{12}, s_{13}, (p_1 \cdot \varepsilon_2), (p_1 \cdot \varepsilon_3), (p_2 \cdot \varepsilon_1), (p_2 \cdot \varepsilon_3), (p_2 \cdot \varepsilon_4), (p_3 \cdot \varepsilon_1), \\ &(p_3 \cdot \varepsilon_2), (p_3 \cdot \varepsilon_4), (\varepsilon_1 \cdot \varepsilon_2), (\varepsilon_1 \cdot \varepsilon_3), (\varepsilon_1 \cdot \varepsilon_4), (\varepsilon_2 \cdot \varepsilon_3), (\varepsilon_2 \cdot \varepsilon_4), (\varepsilon_3 \cdot \varepsilon_4)\} \,. \end{aligned} \tag{121}$$

There are 30 possible combinations, leading to an ansatz with an equal number of parameters. Besides constraining it with the kinematic Jacobi identity (117), we also impose the antisymmetry constraints at the two vertices[17],

$$n(a,b,c,d) = -n(a,b,d,c) = -n(b,a,c,d) \,, \tag{122}$$

matching the antisymmetry of the color factors.

Applying these constraints on the ansatz built from the monomials in equation (121) fixes all but five of the ansatz' coefficients. Further imposing gauge invariance on one external leg then fixes the form of the numerator. In fact, it is sufficient to impose gauge invariance on one leg when the amplitude is factorized on the pole of a given channel, i.e. for $s_{ab} \to 0$,

$$n(a,b,c,d)\,|_{s_{ab}\to 0,\ \text{and}\ \varepsilon_a \to p_a} \to 0 \,. \tag{123}$$

[17] See section 2.1, equation (51) and discussion below it.

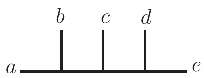
This gives

$$n(a,b,c,d) \propto \{[(\varepsilon_a\cdot\varepsilon_b)\,p_a^\mu + 2\,(\varepsilon_a\cdot p_b)\,\varepsilon_b^\mu - (a\leftrightarrow b)]\,[(\varepsilon_c\cdot\varepsilon_d)\,p_{c\mu} + 2\,(\varepsilon_c\cdot p_d)\,\varepsilon_{d\mu} \\ - (c\leftrightarrow d)] + s_{ab}\,[(\varepsilon_a\cdot\varepsilon_c)\,(\varepsilon_b\cdot\varepsilon_d) - (\varepsilon_a\cdot\varepsilon_d)\,(\varepsilon_b\cdot\varepsilon_c)]\}\,, \tag{124}$$

in agreement with equation (12). We note that, as explained in [173, 174], one can also determine the amplitude using other constraints, in particular from gauge invariance and mild assumptions on the singularity structure.

Exercise 3.6. Verify explicitly that the four-point YM numerator given above satisfies the Jacobi constraint.

Emboldened by the success to obtain four-point amplitude by imposing dual-Jacobi relations, we continue to the next multiplicity for the NLSM. As for four points, the duality involves not only imposing kinematic Jacobi relations, but also the same antisymmetry carried by color factors. For the half-ladder numerators,

$$n(a,b,c,d,e) \equiv n\Bigg($$


$$\Bigg)\,, \tag{125}$$

these antisymmetry constraints read:

$$n(a,b,c,d,e) = -n(a,b,c,e,d) = -n(b,a,c,d,e) = -n(d,e,c,a,b)\,. \tag{126}$$

The Jacobi identities corresponding to the two independent propagators of the five-point half-ladder graphs are:

$$n(a,b,c,d,e) = n(a,c,b,d,e) + n(c,b,a,d,e)\,, \\ n(a,b,c,d,e) = n(a,b,d,c,e) + n(b,a,e,c,d)\,. \tag{127}$$

One can immediately see the need to impose two Jacobi relations from the two triangles that touch every vertex in the graph of graphs for the five-point tree as drawn in figure 16.

At five points, we have a 35-parameter ansatz comprised of all degree-three monomials with factors from

$$\{s_{ab}, s_{ac}, s_{ad}, s_{bc}, s_{bd}\}\,. \tag{128}$$

The constraints in equations (126) and (127) fix 34 parameters, leaving us with a unique expression, up to an overall coefficient,

$$n(a,b,c,d,e) \propto (s_{ac}+s_{bc})\,(s_{ad}\,(s_{bd}+s_{be}) - \{a\leftrightarrow b\})\,. \tag{129}$$

Remarkably, the five-point amplitudes obtained from these numerators actually vanish, in line with the fact that odd-point amplitudes vanish in the NLSM. Moreover, this amplitude vanishes without having to impose the requirement that the underlying theory has no three-point vertex (i.e. that there is no two-particle factorization channel). This in turn suggests that there do not exist CK-satisfying scalar two-derivative theories with only fields in the adjoint representation that are not the NLSM.

Exercise 3.7. Verify that the above numerator satisfies the two independent Jacobi relations at five points.

Exercise 3.8. Verify explicitly that a color-ordered amplitude, say $A_5^{\text{tree}}(1,2,3,4,5)$, expressed in terms of its cubic graphs using the diagram numerator in equation (129) vanishes.

One can continue in this way, systematically building up higher-point amplitudes. It is also useful to impose other physical constraints such as the vanishing of all factorization limits where at least one factor is an odd-point amplitude:

$$\lim_{s_{1\ldots 2k}\to 0} s_{1\ldots 2k} A_n^{\text{tree}}(1,\ldots,n) = 0\,, \tag{130}$$

or, equivalently, the vanishing of the residue of the simple pole in $s_{1\ldots 2k}$:

$$\sum_{\text{states}} A_{2k+1}^{\text{tree}}(1,\ldots 2k,p)\, A_{n-2k+1}^{\text{tree}}(-p,2k+1,n)\big|_{p^2=0} = 0\,. \tag{131}$$

These conditions guarantee recursively consistency with the vanishing of the odd-point tree amplitudes with multiplicity smaller than n. Let us illustrate this for the six-point amplitude. For our scalar theory, we have nine independent external momentum invariants; from them we can construct 495 degree-four monomials thus obtaining a 495-parameter ansatz for the half-ladder graph. CK duality alone constrains all but 23. Imposing the vanishing of the color-ordered factorization

$$\sum_{\text{states}} A_3^{\text{tree}}(1,2,p)\, A_5^{\text{tree}}(-p,3,4,5,6) = 0\,, \tag{132}$$

leaves six unconstrained parameters. Although individual diagrams depend on them, these parameters always appear in the same linear combination in front every color-ordered partial amplitude, which indeed reproduce the six-point partial NLSM amplitudes, up to the overall normalization. Other factorization limits, e.g. $0 = \text{Res}_{s_{23}=0}(A_6^{\text{tree}}(1,2,3,4,5,6)) = \sum_{\text{states}} A_3^{\text{tree}}(2,3,p)\, A_5^{\text{tree}}(-p,4,5,6,1)$, do not constrain them any further, implying that it should be possible to remove five of the remaining six parameters by a local generalized gauge transformation.

When combined, relabeling symmetry and the dual Jacobi relations are extremely constraining and can be used to determine scattering amplitudes in the NLSM and gauge theory. As the number of legs and loops increases, this process of constraining an ansatz becomes increasingly more tedious. However, there are now a variety of constructive approaches for building tree-level and low-loop numerators that satisfy the kinematic Jacobi identities [14, 24–26, 28, 29, 149, 156, 211–216].

In general, higher-loop integrands is a more involved problem, perhaps more in gauge theories than for the NLSM. Direct approaches based on constraining ansätze have proven an effective means of generating gauge-theory loop integrands [6, 217]. We will see explicit examples in section 6. It turns out, however, that it can be difficult to find gauge-theory numerators that manifest the duality between color and kinematics thus complicating the construction of corresponding gravity integrands. Nevertheless, a generalized double-copy procedure outlined in section 7 can be used to convert gauge-theory integrands in generic representations to integrands in gravity theories; for example, this procedure was used to obtain the five-loop four-point integrand of $\mathcal{N}=8$ supergravity and determine its UV behavior [38, 218].

4. Gravity symmetries and their consequences

Symmetries are essential for understanding the properties of gauge and gravity theories. In the context of the framework provided by CK duality and the double copy, which relates QFTs order-by-order in perturbation theory, it is hence interesting to explore how symmetries originate and transfer. Not all symmetries of double-copy theories are currently well-understood from this perspective; likewise, the consequences of certain symmetries of single-copy parent theories have yet to be properly understood[18]. In this section we review the current status of the relation between the symmetries of the single- and double-copy theories. We begin by outlining which (part) of the symmetries of a Lagrangian can be identified and analyzed through scattering-amplitude techniques emphasizing that, while the linearized part of symmetries can be manifest, nonlinear symmetries affect only special momentum configurations of scattering amplitudes. We then proceed to discuss the linearly-realized global symmetries and to extend the diffeomorphism and local-supersymmetry discussion in section 2 to also include nonabelian gauge symmetry. All these symmetries are inherited from the symmetries of their single-copy parents. We then discuss certain enhanced symmetries, i.e. symmetries which, while unrelated to any of the single-copy symmetries, act linearly on the double-copy asymptotic states. The ability to efficiently compute amplitudes and analyze them for special momentum configurations is essential to explore the emergence of nonlinear symmetries in the double copy.

4.1. Symmetries: Lagrangian vs. scattering amplitudes

Lagrangians exhibiting nonlinear symmetries—such as supersymmetry in a formulation without auxiliary fields, or nonabelian gauge symmetries—are usually constructed through an iterative Noether procedure. One starts with the free-field theory with the desired spectrum, which is invariant under the linearized form of the desired symmetries and simultaneously deforms the action and the transformation rules such that the resulting action is invariant off-shell under the deformed transformations. The resulting symmetry algebra closes up to the equations of motion. Thus, this approach leads to actions and transformation rules of the form

$$\begin{aligned} &S = S_2 + S_3 + S_4 + \dots\,, \\ &\delta = \delta^{(0)} + \delta^{(1)} + \delta^{(2)} + \dots\,, \end{aligned} \tag{133}$$

where the n-field term S_n in the action determines the $(n-2)$-field term $\delta^{(n-2)}$ in the transformation rules. For example,

$$\begin{aligned} \delta^{(0)}S_3 + \delta^{(1)}S_2 &= 0\,, \\ \delta^{(0)}S_4 + \delta^{(1)}S_3 + \delta^{(2)}S_2 &= 0\,. \end{aligned} \tag{134}$$

The first relation implies that the cubic term is invariant under the undeformed transformations up to terms proportional to the free equations of motion.

Quantum mechanically, symmetries are realized through Ward identities, which relate time-ordered correlation functions of the fundamental fields of the theory. Nonlinear transformation rules imply that the relevant Ward identities contain correlation functions of different multiplicities. For example, for a transformation $\delta\phi \propto \phi^k$, an n-point Green's function is related to $(n+k-1)$-point Green's functions. Moreover, locality of the transformation rules imply that, for $k \geqslant 2$, these k fields are at the same spacetime point(s).

Upon Lehmann–Symanzik–Zimmermann (LSZ) reduction, Ward identities simplify considerably. For asymptotic states with momenta $p_1, \dots, p_n$, the amputation leading to the

[18] Moreover, symmetries of sectors of a single-copy parent theories that relate to the gauge group—such as symmetries of the planar sector–seem difficult to capture because of 'contamination' from other sectors.

n-point amplitude selects the most singular term, proportional[19] to $\prod_{i=1}^{n}(p_i^2-m_i^2)^{-1}$. For an $(n+k-1)$-point $(k=2,3,\ldots)$ Green's function resulting from a nonlinear term in a (symmetry) transformation, momentum conservation requires that it has a different pole structure. Thus, all such terms are amputated away and all effects of nonlinear terms in the symmetry transformations that underlie the structure of off-shell Ward identities are projected out by the LSZ reduction. The resulting on-shell Ward identities imply that, for generic momenta, S-matrix elements are invariant only under the linearized symmetry transformations. This argument fails when the additional fields appearing in nonlinear symmetry transformations all carry vanishing momenta; indeed, in this case, the off-shell $(n+k)$-point Green's function develops a pole $\prod_{i=1}^{n}(p_i^2-m_i^2)^{-1}$ and gives a nonvanishing contribution after the LSZ n-point amputation. We shall return in section 4.5 to this special momentum configuration and interpret it as the soft limit of a higher-point scattering amplitude.

It is not difficult to identify these features in nonabelian YM theory: the amplitudes vanish if the polarization vector of a gluon $\varepsilon_\mu(p)$ is replaced by the momentum[20]

$$\delta^{(0)}A_\mu=\partial_\mu\Lambda \quad\longrightarrow\quad \delta\varepsilon_\mu(p)=p_\mu\Lambda(p)\,. \tag{135}$$

They are also invariant under the global part of the gauge group, which is the only remnant of the nonlinearity of the gauge transformation.

Not all symmetry transformations have a linearized approximation. An outstanding class of examples are the U-duality symmetries of extended supergravity theories, such as the $E_{7(7)}$ duality group of $\mathcal{N}=8$ supergravity. It turns out that only their maximal compact subgroup, which is isomorphic to the on-shell R-symmetry group, has such an approximation. It is therefore an interesting question whether on-shell methods can probe symmetry transformations which are inherently nonlinear.

A possible approach, put forward in [219] and further explored in [220, 221], effectively amounts to constructing the quantum one-particle irreducible (1PI) effective action and studying its symmetries. Indeed, the quantum 1PI effective action is determined[21] by the S-matrix of the theory up to terms proportional to the free equations of motion. Consequently, up to the corresponding contact terms and assuming absence of anomalies, the off-shell Ward identities of all symmetries—in particular of the nonlinear ones—should hold. In this formalism, anomalies appear as violations of the Ward identities of the corresponding symmetries which cannot be removed by the addition of finite local counterterms to the (effective) action. These counterterms may be simultaneously interpreted both as part of the definition of the theory and as an ambiguity in the construction of the effective action from the S-matrix.

Another approach geared towards the exploration of nonlinearly-realized symmetries was first described in [222] for the $E_{7(7)}$ symmetry of $\mathcal{N}=8$ supergravity in four dimensions. It amounts to (1) the vanishing of scattering amplitudes in the limit in which momenta of one scalar field vanish and (2) the identification/extraction of the structure constants of the nonlinearly-realized part of the symmetry group from the limit in which two scalar fields have vanishing momenta. In section 4.5 we will outline this approach and summarize some of its many generalizations to nonlinearly-realized (Volkov-Akulov) supersymmetry [220, 223, 224], Bondi–Metzner–Sachs (BMS) symmetry [225–227], anomalous symmetries [228],

[19] Here we assume that external states have generic masses m_i. Assuming from the outset that external states are massless does not alter the conclusion.

[20] It is worth mentioning that, from the perspective of the gauge-fixed theory, the transformation $\varepsilon^\mu\to\varepsilon^\mu+\Lambda(p)p^\mu$ can also be interpreted as ε^μ not being a proper Lorentz vector [173]. Indeed, the polarization vector is constrained to obey $p\cdot\varepsilon=0$ so, on shell, any transformation of ε can include a shift by p^μ.

[21] One may use other methods, such as those outlined in [219], to construct the effective action.

effective theories [229], string theory [230–232] and theories with spontaneously-broken conformal invariance [233]. For discussions of soft theorems at the quantum level see [234–238]. While neither of these two approaches is specifically tied to the double-copy construction, they may provide strategies to understanding aspects of symmetries of the double-copy theories and their relation to their single-copy parents.

4.2. Global symmetries; on-shell R symmetry

In the absence of anomalies, the scattering amplitudes of a theory exhibit its off-shell symmetries to all orders in perturbation theory[22]. Below, we shall review how the double copy expresses this property.

As we reviewed at length in previous chapters, at tree level the KLT relations build gravity scattering amplitudes from gauge-theory amplitudes. More generally, for all double-copy theories (including the non-gravitational ones), there exist analogous relations that build their scattering amplitudes in terms of pairs of theories. It is therefore clear that, at tree level, the global symmetry group G of a double-copy theory is at least as large as the product of the global symmetry groups $G_{1,2}$ of the parent theories:

$$G \supset G_1 \otimes G_2 \,. \tag{136}$$

In a Feynman-diagram approach to the construction of scattering amplitudes, one can arrange that each diagram exhibits all off-shell global symmetries of the classical Lagrangian. The construction of tree-level CK-satisfying numerators in terms of tree-level amplitudes [24, 149] implies that, at tree level, the same is true for each single-copy parent theory if one also demands that the amplitude obey CK duality. Thus, equation (136) also holds in this approach.

In the presence of a symmetry-preserving regulator, generalized unitarity then guarantees that the regularized cuts of higher-loop amplitudes of the double-copy theory also inherit all the global symmetries of the single-copy parent theories.

Exercise 4.1. Explore if there is a general statement that can be made about anomalous global symmetries, i.e. whether all anomalous global symmetries of the single-copy parent theories remain anomalous in the double-copy theory. To this end, consider the example of a four-dimensional gauge theory with chiral fermions and construct examples of the double-copy amplitudes that involve scattering amplitudes of this theory that are sensitive to the chiral anomaly.

Not all global symmetries of a double-copy theory are inherited; in fact, inheritance of some symmetries demands that others be enhanced. Consider, for example, the case of the double copy of two theories with $\mathcal{N}_1$ and $\mathcal{N}_2$-extended supersymmetry, respectively. Their supersymmetry algebras have $SU(N_1)$ and $SU(N_2)$ R symmetry (perhaps with additional decoupled $U(1)$ factors) and, according to the previous discussion, the double-copy theory will be invariant under at least $SU(N_1)\times SU(N_2)$ transformations. However, the $(\mathcal{N}_1+\mathcal{N}_2)$-extended supersymmetry algebra that is expected based on the number of supercharges has a larger R symmetry, $SU(N_1+N_2)$. Thus, to extend $R_1\otimes R_2$ to the complete R symmetry group, it is necessary to identify further $2N_1N_2+1$ generators. The first $2N_1N_2$ generators were constructed in [239, 240] in terms of the supersymmetry generators of the two single-copy parent theories. In four dimensions, they are

$$G_{I\tilde{J}} = Q_{+I}\tilde{Q}_{-\tilde{J}} \quad \text{and} \quad G^{I\tilde{J}} = Q_-{}^I\tilde{Q}_+{}^{\tilde{J}}\,, \tag{137}$$

[22] This assumes the existence of a regulator that preserves these symmetries.

Table 2. Action of the $G^{I\tilde{J}}$ generator defined in equation (137) on the states of $\mathcal{N}=8$ supergravity. The action of the generators $G_{I\tilde{J}}$ is obtained by reversing the direction of the arrows.

	$\tilde{g}^+$	$\tilde{f}^+_{\tilde{I}}$	$\tilde{\phi}_{\tilde{I}\tilde{J}}$	$\tilde{f}^-_{\tilde{I}\tilde{J}\tilde{K}}$	$\tilde{g}^-_{\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
g^+	h^+	$\psi^+_{\tilde{I}}$	$A^+_{\tilde{I}\tilde{J}}$	$\chi^+_{\tilde{I}\tilde{J}\tilde{K}}$	$\phi_{\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
f^+_I	ψ^+_I	$A^+_{I\tilde{I}}$	$\chi^+_{I\tilde{I}\tilde{J}}$	$\phi^+_{I\tilde{I}\tilde{J}\tilde{K}}$	$\chi^-_{I\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
ϕ_{IJ}	A^+_{IJ}	$\chi^+_{IJ\tilde{I}}$	$\phi^+_{IJ\tilde{I}\tilde{J}}$	$\chi^-_{IJ\tilde{I}\tilde{J}\tilde{K}}$	$A^-_{IJ\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
f^-_{IJK}	χ^+_{IJK}	$\phi^+_{IJKL\tilde{I}}$	$\chi^+_{IJK\tilde{I}\tilde{J}}$	$A^-_{IJK\tilde{I}\tilde{J}\tilde{K}}$	$\psi^-_{IJK\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
g^-_{IJKL}	ϕ_{IJKL}	$\chi^-_{IJKL\tilde{I}}$	$A^-_{IJKL\tilde{I}\tilde{J}}$	$\psi^-_{IJKL\tilde{I}\tilde{J}\tilde{K}}$	$h^-_{IJKL\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$

where Q and $\tilde{Q}$ are the supersymmetry generators of the two single-copy parent theories, respectively, and the $\pm$ indices represent their helicity. These generators have vanishing total helicity. From their structure it is clear that they change the helicities of the two single-copy components in opposite ways, such that the helicity of the double-copy state is unchanged. For the case of $\mathcal{N}=8$ supergravity, their action on states is given in table 2.

The remaining (Cartan) generator which is necessary to recover the complete (and expected) R-symmetry group may in principle be obtained from the closure of the off-diagonal $G_{I\tilde{J}}$ and $G^{I\tilde{J}}$. In section 4.4, we shall review another way of identifying it, as well as its physical interpretation.

An interesting feature which has been observed in explicit examples, some of which are described in section 5, is that certain gravity theories have two distinct double-copy realizations. In these cases, each version of the construction exhibits different manifest symmetries and, while following the pattern above, the details of the symmetry enhancement are different. An example discussed in [241] from a double-copy perspective and in [242] from a string-theory point of view, is $\mathcal{N}=4$ supergravity with two vector multiplets, which can be realized both as ($\mathcal{N}=4$ SYM)$\times$(YM+2 scalars) and ($\mathcal{N}=2$ SYM)$\times$($\mathcal{N}=2$ SYM). While in the former construction the complete $SU(4)$ R symmetry of supergravity is manifest, the latter only has a manifest $SU(2)\times SU(2)$ symmetry. A more dramatic example is provided by three-dimensional $\mathcal{N}\geqslant 8$ supergravities, which can be realized [243, 244] either in terms of two three-dimensional SYM theories or in terms of two Chern–Simons-matter theories [245–250] (see also [119] for the double-copy realization of maximally supersymmetric three-dimensional supergravity [251]).

To study the origin of the symmetries of a gravitational theory from the double-copy factors, it is sometimes convenient to introduce a manifestly covariant formulation by defining the action of the double copy on the off-shell linearized (super)fields, following an approach introduced in [56, 239, 252–256]. To give an explicit example, we consider the double copy of two vector fields, which in this language is written as

$$H_{\mu\nu} = h_{\mu\nu} + B_{\mu\nu} + \phi\eta_{\mu\nu} = A^a_\mu \star \Phi^{-1}_{aa'} \star \tilde{A}^{a'}_\nu\,, \tag{138}$$

where A^a_μ and $\tilde{A}^{a'}_\mu$ are the fields in the left and right gauge theory, respectively. The resulting double-copy field $H_{\mu\nu}$ (sometimes referred to as the 'fat graviton' [58], see also section 8) needs to be decomposed in irreducible representations of the Lorentz group, giving the graviton field, the dilaton and an antisymmetric tensor $B_{\mu\nu}$. This version of the construction is formulated in position space and hence relies on the convolution among linearized superfields, which is defined as

$$[f\star g](x) = \int d^4y\, f(y)\, g(x-y)\ . \tag{139}$$

Crucially, $\Phi_{aa'}$ is a bi-adjoint scalar field which is employed to contract the gauge indices of the left and right fields. The action of a symmetry transformation on left and right fields is then written as

$$\begin{aligned}
\delta A^a_\mu &= \partial_\mu \Lambda^a + f^a{}_{bc} A^b_\mu \theta^c + \hat{\delta} A^a_\mu\,,\\
\delta \tilde{A}^a_\mu &= \partial_\mu \tilde{\Lambda}^a + f^a{}_{bc} \tilde{A}^b_\mu \tilde{\theta}^c + \hat{\delta} \tilde{A}^a_\mu\,,
\end{aligned} \tag{140}$$

where Λ^a, $\tilde{\Lambda}^a$, θ^a and $\tilde{\theta}^a$ are the parameters of local abelian and global nonabelian gauge transformations, while $\hat{\delta}$ indicates a global transformation under the (super)Poincaré group. The bi-adjoint scalars are designed to offset the left and right gauge transformations, and transform as [253]

$$\delta\Phi^{-1}_{aa'} = -f^b{}_{ac}\Phi^{-1}_{ba'}\theta^c - f^{b'}{}_{a'c'}\Phi^{-1}_{ab'}\tilde{\theta}^{c'} + \hat{\delta}\Phi^{-1}_{aa'}\,. \tag{141}$$

While this approach treats the action of the gauge-theory symmetries in an elegant way, the full dictionary is known only at the linearized level. As it was shown in [256], the linearized gravitational equations of motion can be obtained from the linearized gauge-theory ones. It remains an open question how to include interactions in this formalism (which are naturally incorporated from the perspective of scattering amplitudes).

Exercise 4.2. Use equation (141) to show that the fat graviton defined in equation (138) is inert under nonabelian gauge transformations. Moreover, show that its local transformation rules are a linear combination of linearized diffeomorphisms and gauge transformations of a two-index antisymmetric tensor field.

4.3. Local symmetries

In section 2, we discussed in detail the emergence of diffeomorphism invariance and local supersymmetry in gravity scattering amplitudes obtained from the double-copy construction. The former is a direct consequence of the gauge invariance of the two single-copy gauge theories and manifest CK-satisfying form for at least one of the two gauge theories [156, 173, 174]. The latter is a consequence of the gauge invariance of one of the single-copy gauge theories, supersymmetry of the second, and manifest CK-satisfying form for at least one of them. The on-shell supersymmetry Ward identities of the double-copy theory follow from those of the single-copy parents. In this section we review how similar mechanisms lead to other local symmetries in double-copy theories.

As discussed in the beginning of this section, scattering amplitudes in theories with local symmetries that act nonlinearly on fields are invariant under the global part of the symmetry group (if it acts linearly) as well as under its linearized local transformations. The converse, however, does not necessarily hold: scattering amplitudes that are invariant under some global symmetry group G and under abelian local transformations do not necessarily describe a QFT with a local G symmetry. For example, they may correspond to a field theory with $\dim(G)$

abelian vector fields. It is of course not difficult to distinguish between these two possibilities by inspecting the scaling dimension of certain scattering amplitudes, which is different according to whether the theory involves abelian and nonabelian vector fields.

From the discussion in section 2, it is clear that, in any double-copy theory, each vector field whose asymptotic states are realized as a product of a scalar- and a vector-field asymptotic states exhibits a Maxwell gauge symmetry, which is a consequence of the corresponding gauge symmetry of the vector field in the single-copy parent theory[23]. In order to associate these vector fields to a local nonabelian symmetry, the corresponding amplitudes must exhibit several properties: (1) be invariant under the adjoint action of some a global nonabelian symmetry group on the asymptotic states of vector fields and (2) have the correct dimension to be consistent with minimal coupling. The second property demands that a three-vector amplitude have unit dimension,

$$\left[A_3^{(0)}\right] = 1\,, \tag{142}$$

as in a standard nonabelian gauge theory. To obtain such amplitudes through double copy, at least one of the single-copy parent theories must have amplitudes of dimension zero. Lorentz invariance and locality imply then that the corresponding single-copy fields labeling such amplitudes must be scalars. To satisfy property (1), the corresponding amplitude must be momentum-independent and coming from a Lagrangian of the type

$$\mathcal{L} = \cdots + f^{abc}F^{ABC}\phi^{aA}\phi^{bB}\phi^{cC} + \ldots\,, \tag{143}$$

where the ellipsis stand for other interactions. This reasoning led to the double-copy realization of YME theories, as described in [120]. It was also used in [257] to obtain amplitudes in the same theory through a KLT-like construction. This procedure can be extended to give a double-copy realization of spontaneous breaking of YM gauge symmetry of supergravity theory [122]; spontaneous breaking of this symmetry is related to explicit breaking of a global symmetry of one of the single-copy parent theories. We shall review its applications more thoroughly in section 5.3.7.

The same analysis implies that the double-copy fields that are realized as products of single-copy fields with nonzero spin cannot couple directly to the nonabelian vector potential and can couple only to its field strength. Indeed, in a conventional gauge theory, any three-point amplitude with at least one field with nonzero spin has unit dimension. Thus, the corresponding double-copy three-point amplitude has dimension 2 and cannot be given by a minimal-coupling term.

The analysis above can be extended to interactions of gravitini with abelian or nonabelian gauge potentials. Such interactions are the tell-tale of gauged supergravities—that is, supergravities in which part of the R symmetry is gauged. The gravitino minimal coupling around Minkowski space is

$$\mathcal{L}_3 \sim \bar{\psi}_\mu \gamma^{\mu\nu\rho} D_\nu \psi_\rho\,, \tag{144}$$

and, thus, as for the case of lower-spin fields, the two-gravitini-vector amplitudes have again unit dimension. Since all three-point amplitudes in conventional gauge theories that could have this spin content in their product have at least unit dimension, it follows that their double

[23] We note that the single-copy origin of Maxwell gauge symmetry is not obvious if the vectors are realized as in terms of two fermions.

copy can only describe the coupling of gravitini and field strengths and thus, a more refined argument is needed to accommodate minimal couplings of gravitini.

The observation that sidesteps the difficulty exposed above [123] is that, assuming that the theory has a Minkowski ground state, the three-point amplitude following from the minimal coupling of a gravitino with a vector field spontaneously breaks supersymmetry. Consequently, some of the gravitini must be massive and therefore their double-copy realization must involve a single-copy theory with massive vector fields and another one with massive fermions. The former must therefore be a spontaneously-broken gauge parent theory while the latter turns out to exhibit explicit supersymmetry breaking (this construction will be illustrated in detail in section 5.3.8). The general pattern is that, through the double copy, explicit breaking of a global symmetry can be promoted to spontaneous breaking of the local version of the same symmetry.

4.4. Dualities

We have seen in section 4.2 that double-copy theories inherit all the global symmetries of their single-copy parents and that some of the single-copy symmetries combine in nontrivial ways (e.g. two supersymmetry generators combine to become a bosonic R-symmetry generator) to enhance the inherited symmetries. Lagrangian-based supergravity considerations suggest the existence of much larger symmetries—the U-duality symmetries—which are typically noncompact. In pure supergravities in various dimensions, these symmetries were originally discussed in [258, 259]; their maximal compact subgroup is isomorphic to the on-shell R-symmetry group of the theory. Depending on the dimension, they are either symmetries of the Equations of motion (in four dimensions) or symmetries of the Lagrangian (e.g. in five dimensions). In four dimensions, the field strengths and their duals form an irreducible representation of the U-duality group, which therefore contains electric/magnetic duality as one of its generators. While the general understanding U-duality symmetries from the double-copy perspective is currently an open problem, their dimension-dependent properties suggest that their realization should involve transformations that are not off-shell symmetries in the single-copy parent theories.

Carrasco *et al* [260] showed that a universal generator of the U-duality groups of four-dimensional supergravities can be realized as the difference of the little-group generators (helicity) of the two single-copy parent theories; the charges of the double-copy fields under this generator are

$$Q = q\left(h - \tilde{h}\right). \tag{145}$$

Note that this transformation acts on the positive and negative helicity vector fields in the double-copy theory with opposite phases. Because of this property, the above transformation can be identified as an electric/magnetic duality transformation acting on vector fields, combined with additional transformations of other fields [261, 262]. It is not *a priori* clear why this transformations should be a symmetry of the double-copy theory at tree level. One can nonetheless check that for $\mathcal{N} \geqslant 5$ it is part of the on-shell R symmetry of the theory and thus part of the maximal compact subgroup of the U-duality. For example, decomposing the positive-helicity (denoted with the index '+' below) and scalar states of $\mathcal{N} = 8$ supergravity in representations of the $(SU(4), \widetilde{SU(4)})^{U(1)}$ subgroup of the $SU(8)$ R symmetry (of which only $SU(4) \times SU(4)$ is manifest in the double copy) one finds

$$
\begin{aligned}
\mathbf{1}^+ &= (\mathbf{1},\mathbf{1})^0\ , \\
\mathbf{8}^+ &= (\mathbf{4},\mathbf{1})^q \oplus (\mathbf{1},\mathbf{4})^{-q}\ , \\
\mathbf{28}^+ &= (\mathbf{6},\mathbf{1})^{2q} \oplus (\mathbf{1},\mathbf{6})^{-2q} \oplus (\mathbf{4},\mathbf{4})^0\ , \\
\mathbf{56}^+ &= \left(\bar{\mathbf{4}},\mathbf{1}\right)^{3q} \oplus \left(\mathbf{1},\bar{\mathbf{4}}\right)^{-3q} \oplus (\mathbf{6},\mathbf{4})^q \oplus (\mathbf{4},\mathbf{6})^{-q}\ , \\
\mathbf{70}\ &= (\mathbf{1},\mathbf{1})^{4q} \oplus (\mathbf{1},\mathbf{1})^{-4q} \oplus \left(\bar{\mathbf{4}},\mathbf{4}\right)^{2q} \oplus \left(\mathbf{4},\bar{\mathbf{4}}\right)^{-2q} \oplus (\mathbf{6},\mathbf{6})^0\ .
\end{aligned}
\tag{146}
$$

As pointed out in [260], the $U(1)$ charges resulting from this decomposition are exactly given by equation (145). The decomposition of the negative helicity states is obtained by conjugating the first four lines of equation (146); the $U(1)$ charge changes sign under conjugation and may be also identified as being proportional to the net number of indices of the states in table 2. From this table, we also see that supersymmetry generators change the $U(1)$ charge by $1/2$ unit, while the off-diagonal R-symmetry generators enhancing $SU(4)\times\tilde{SU}(4)\to SU(8)$ change the $U(1)$ charge by one unit. While we illustrated here its relevance for $\mathcal{N}=8$ supergravity, the $U(1)$ symmetry described by equation (145) is required for obtaining the complete on-shell R-symmetry for all $\mathcal{N}\geqslant 5$ supergravities, as follows from the fact that the latter theories can be obtained as consistent truncations of the former.

For $1\leqslant\mathcal{N}\leqslant 4$, this symmetry is present at tree level but it is anomalous [260, 263]. This anomaly sources certain loop-level amplitudes which vanish at tree level. In the realization of these theories as double copies with one non-supersymmetric gauge-theory factor, these anomalous amplitudes [260] can be traced to a self-duality anomaly of YM theory [264]. When realized as double copies of supersymmetric gauge theories, the identification of anomalous amplitudes is more subtle: they arise from μ-terms[24] which, in each gauge theory, give only $\mathcal{O}(\epsilon)$ terms but give finite terms only after the double copy [20]. It turns out [265] that, at least for $\mathcal{N}=4$ supergravity, the anomalous amplitudes can be canceled at one loop by the addition of a finite local counterterm to the classical action; this counterterm restores the $U(1)$ symmetry at the expense of breaking other symmetries[25] that do not appear to impose any obvious selection rules on amplitudes. The same counterterms also cancels the two-loop anomalous amplitudes [266]. The full consequences of these cancellations remain to be explored[26].

It is instructive to consider the $U(1)$ transformation with charges (145) vis à vis the observation discussed in section 4.2 that the same supergravity theory may have (two or perhaps more) different double-copy realizations. While a general analysis is yet to be carried out, it is not difficult to see on a case-by-case basis that this symmetry may play different roles. To this end, consider $\mathcal{N}=4$ supergravity with two vector multiplets, realized as ($\mathcal{N}=4$ SYM)×(YM+2 scalars) and ($\mathcal{N}=2$ SYM)×($\mathcal{N}=2$ SYM), both of which can be obtained as different (orbifold) truncations[27] of the double-copy constructions of $\mathcal{N}=8$ supergravity. In the former, the $SU(4)$ R symmetry in manifest and the $U(1)$ symmetry is part of the $SU(1,1)$ duality group of $\mathcal{N}=4$ supergravity. In the latter, only $SU(2)\times SU(2)\subset SU(4)$ is manifest and the $U(1)$ symmetry is required to enhance it to the complete $SU(4)$ R symmetry. In this formulation the origin of $U(1)\subset SU(1,1)$ is not clear. Similarly, the further enhancement to the $SO(6,2)\times SU(1,1)$

[24] μ-terms are numerator terms proportional to the extra-dimensional parts of loop momenta. Such terms vanish identically if the integrand is evaluated in four dimensions.

[25] They are the two generators that, together with $U(1)$, form the $SU(1,1)$ classical U-duality group of the theory.

[26] For $\mathcal{N}=0$ supergravity, defined as the double copy of two pure YM theories, this symmetry is also present; it represents the $U(1)$ rephasing of the dilaton-axion. It survives at the quantum level because there are no fields that can contribute to its anomaly.

[27] See section 5.2.2 for details on field-theory orbifolds in this and related contexts.

complete U-duality group (see section 5) is currently an open problem, on the same footing as $SU(8) \to E_{7(7)}$ in $\mathcal{N}=8$ supergravity[28].

The definition of this universal $U(1)$ symmetry in (145), does not single out supergravities as the only double-copy theories that exhibit this symmetry. There exist many non-gravitational four-dimensional field theories exhibiting electric/magnetic duality; for all those that have a double-copy realization, the transformations of the asymptotic states under duality have the same form (145). An example is the Born–Infeld theory; in this case duality implies that only split-helicity amplitudes (i.e. amplitudes with an equal number of positive and negative vector fields) are nonvanishing. While this can be proven to all multiplicities through various techniques [261, 262], it would be interesting to understand this property from the perspective of the double-copy construction. Quite generally, it remains an interesting open question to understand the consequences of duality from the perspective of the single-copy parent theories.

The one-loop all-multiplicity all-plus and single-minus amplitudes of the Born-Infeld theory were constructed using D-dimensional unitarity and supersymmetric decomposition in [269] and integrated using dimension-shifting relations in [270]. The amplitudes in both classes turn out to be nonvanishing, implying that, similarly to $\mathcal{N}=4$ supergravity, duality appears to be anomalous in the non-supersymmetric Born-Infeld theory in this regularization scheme. It remains an open question [269] whether the anomaly is physical or whether it can be removed by a finite local counterterm at the expense of other symmetries. Arguments presented in [262] suggest that it should be possible to restore duality with a counterterm that breaks Lorentz invariance.

4.5. Soft theorems as tests of enhanced global symmetries

As described at length in section 4.1, supergravity considerations suggest the existence of a much larger symmetry group then the one manifestly realized on on-shell scattering amplitudes. Part of this (U-duality) group acts nonlinearly and thus does not impose standard selection rules on scattering amplitudes; consequently, the corresponding generators cannot be realized manifestly (i.e. linearly) on scattering amplitudes simultaneously with supersymmetry and Lorentz invariance[29]. Because of these features, it is not currently known how to identify the single-copy origin of the noncompact U-duality transformations. The discussion in section 4.1 and the ability to compute scattering amplitudes efficiently (both at tree and loop level) gives us an alternative route to probe the existence of these symmetries, borrowing from the supergravity knowledge that scalar fields of the theory parametrize coset space of the form G/H, where G is the U-duality group and H its maximal compact subgroup. In this section, departing from the philosophy in the rest of this review, we shall assume that supergravity amplitudes are available (through the double-copy or by some other means) and describe how to identify the hidden existence of the noncompact U-duality symmetries.

[28] For supergravity theories for which the scalar fields parametrize the locally-homogeneous space G/H with H being the maximal compact subgroup of G, the noncompact part of G can be identified once H and its representations carried by scalars are determined [267]; see also [239, 240, 268] for further details from the double-copy perspective at the noninteracting level. It is nevertheless not clear how to construct the noncompact G-generators in terms of operators in the single-copy theories.

[29] We note that a Lagrangian formulation of $\mathcal{N}=8$ that has manifest $E_{7(7)}$ symmetry was constructed in [271] and further explored in [272]. This formulation, however, breaks manifest Lorentz invariance. Moreover, diffeomorphism transformations on vector fields are realized in a nonstandard way. It would be interesting to explore the scattering amplitudes of $\mathcal{N}=8$ supergravity in this formulation and compare them to the standard form.

This problem was first discussed in detail in [222], where it was shown that the existence of nonlinearly realized symmetries of this type can be identified through the vanishing of single-soft-scalar limit of scattering amplitudes, while the precise group structure can be inferred from the limit in which the momenta of two scalar fields become simultaneously soft. We review this construction, which was also extended to other nonlinearly-realized or spontaneously-broken symmetries, as well as to fields with nontrivial Lorentz-transformation properties, in [220, 224, 228, 231, 236]. A thorough analysis of the soft limits in effective field theories was carried out in [229, 273].

Consider, following [222], a symmetry group G with generators falling into two sets, T and X, broken to the subgroup H generated by T. Schematically, the commutation relations are

$$[T,T] \sim T, \qquad [T,X] \sim X, \qquad [X,X] \sim T. \tag{147}$$

From a Lagrangian point of view (if one is available), there exists a (Nambu-Goldstone) scalar for each of the broken generators X. In general, this Lagrangian has many degenerate vacua; moving from one to another amounts to giving nonzero vacuum expectation values (VEVs) to the Nambu–Goldstone scalars. From the perspective of scattering amplitudes, a vacuum-expectation value of a field corresponds to a condensate of the zero-momentum mode. Thus, exploring the change in vacuum state is equivalent to exploring the properties of scattering amplitudes in the zero-momentum limit for some of the scalars.

A similar conclusion may be reached by revisiting the argument in section 4.1 showing that, for generic momentum configurations, LSZ reduction renders scattering amplitudes insensitive to nonlinear field transformations. Assuming a generic transformation rule $\delta\phi \sim \phi^{k\geqslant 2}$, the same argument implies that, if all but one of the fields on the right-hand side of the transformation carry vanishing momentum, then the transformed Green's function has the same poles as the original one and therefore survives the LSZ reduction. Thus, the nonlinear parts of a symmetry transformation should have a reflection on higher-multiplicity scattering amplitudes in which the additional asymptotic states have vanishing momenta.

Starting with some vacuum state $|0\rangle$, a neighboring one is obtained through a G transformation with parameters given by the VEVs of the old scalars in the new vacuum:

$$|0\rangle_\theta = e^{iX^\alpha\theta_\alpha}|0\rangle\,. \tag{148}$$

Since the G-symmetry requires that amplitudes around the two vacua be the same, the conclusion is therefore that the scattering amplitudes with at least one zero-momentum scalar field vanish identically. For a single soft scalar, this reproduces the celebrated Adler zero [209]. One may turn the single-soft-scalar limit argument around and infer [121] that, in a theory that has vanishing single-soft-scalar limits, the scalar fields belong to a locally-homogeneous space (i.e. a space that has a transitive local group action).

A more involved argument [222] extracts the structure constants of the broken symmetry group from the double-soft-scalar limit of scattering amplitudes:

$$\mathcal{M}_{n+2}(1,2,3,\ldots,n+2) \xrightarrow{p_1,p_2\to 0} \frac{1}{2}\sum_{i=3}^{n+2}\frac{p_i\cdot(p_2-p_1)}{p_i\cdot(p_2+p_1)}T\mathcal{M}_n(3,\ldots,n+2)\,, \tag{149}$$

where T is the G-generator given by the commutator of the X generators corresponding to the two soft scalars. The momenta of the two scalars should be taken soft at the same rate.

Exercise 4.3. As we discussed in section 4.1 we have seen that scattering amplitudes with generic momenta are insensitive to nonlinear terms in symmetry transformations because the

LSZ reduction projects out their contribution. An interesting unexplored problem is the contribution of terms with special momentum configurations. Consider a nonlinear symmetry transformation whose nonlinear parts contains bilinears and cubic terms. Assuming that only one of the fields in the nonlinear terms carries nonzero momentum, explore the features of the single- and double-soft-scalar limits of amplitudes by applying LSZ construction to Green's functions acted upon by such special transformations.

The construction reviewed above does not refer to any specific order in perturbation theory and thus relies on absence of U-duality anomalies. Its conclusions have been used to constrain and characterize possible counterterms of $\mathcal{N}=8$ supergravity, which should be such that their contributions to scattering amplitudes have soft limits following the same pattern. Through this reasoning it was shown that a suggested three-loop R^4 counterterm is inconsistent with the $E_{7(7)}$ symmetry of $\mathcal{N}=8$ supergravity [273]. Along the same lines, [274] argued that the first deformation of $\mathcal{N}=8$ supergravity that is consistent with the soft-scalar behavior required by $E_{7(7)}$ symmetry can appear at seven loops and corresponds to a supersymmetric completion of a D^8R^4 operator.

Generic diffeomorphism transformations are nonlinear. As we discussed in the beginning of this section and in section 4.4, infinitesimal/linearized diffeomorphisms are symmetries of scattering amplitudes: shifting the graviton polarization tensor $\varepsilon(p)^{\mu\nu}\mapsto\varepsilon(p)^{\mu\nu}+p^{(\mu}\Lambda^{\nu)}$ with Λ^μ being the parameter of the transformation, leaves amplitudes invariant. By definition, large diffeomorphisms do not have a linearized approximation; the BMS transformations (named after Bondi, van der Burg, Metzner and Sachs [275–277]) arise, in a certain gauge, as residual diffeomorphism symmetries of asymptotically-flat spacetimes which do not fall off at infinity. It was argued in [278–280] that the Ward identities of these symmetries imply the tree-level single-soft-graviton behavior of scattering amplitudes. Quantum corrections have been discussed in [235, 281], with the conclusion that they affect the linear order in the small momentum if all other momenta are generic. The identification of the BMS algebra in the double-soft-graviton limit of scattering amplitudes was discussed in [227].

Other symmetries can also be probed through double-soft limits. For example, by explicitly inspecting the tree-level amplitudes of a certain Akulov–Volkov theory [223, 224] showed that the double-soft-goldstino limit yields the supersymmetry algebra. Moreover, for $4\leqslant\mathcal{N}\leqslant 8$ supergravities in four dimensions and for $\mathcal{N}=16$ supergravity in three dimensions, tree-level scattering amplitudes have a universal behavior in the double-soft-fermion limit which is analogous to the scalar one. The photon and graviton soft theorems were discussed from an effective-field-theory standpoint in [229], where a complete classification of local operators responsible for modifications of soft theorems at subleading order for photons and subsubleading order for gravitons was derived.

The original discussion [222] of the U-duality symmetries in supergravity and its subsequent generalizations assumed absence of anomalies of the spontaneously-broken symmetry. Possible anomalies have been included in this framework in [228], from the perspective of the effective action; the conclusion of the analysis is that, while the single-soft limits receive corrections signaling the anomalous breaking of the symmetry, double-soft limits are unaffected. This is probably a reflection of the anomaly (defined as the nonvanishing of the divergence of the symmetry current) being invariant under the classical symmetry.

5. A web of double-copy-constructible theories

As we have seen in the previous sections, the duality between color and kinematics and the double-copy construction express amplitudes of gravitational theories in terms of simpler

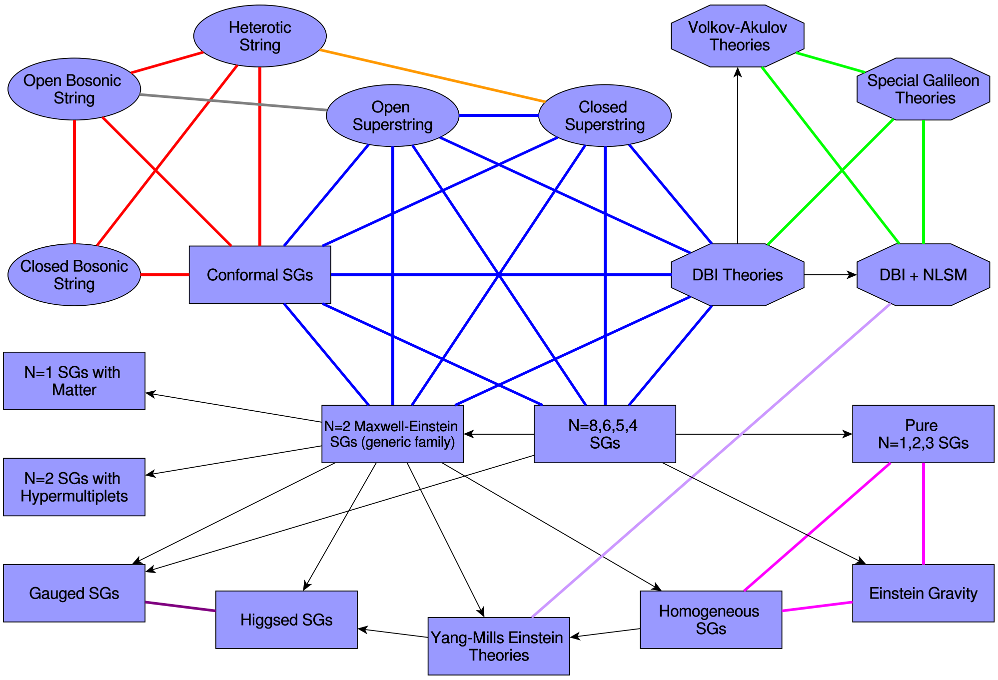


Figure 17. Schematic rendition of the web of theories. Nodes represent the main double-copy-constructible theories discussed in this section, which include gravitational theories (rectangular nodes), string theories (oval nodes) and non-gravitational theories (octagonal nodes). Undirected links are drawn between theories that have a common gauge-theory factor in their construction (different gauge-theory factors correspond to different colors). Directed links connect theories obtained by modifying/deforming both gauge-theory factors (e.g. adding matter, assigning VEVs). Details are given throughout section 5.3.

building blocks from gauge theory. It has become clear that this property is not an accident of few very special theories, but extends to large classes of gravitational and non-gravitational theories. Seemingly unrelated theories have been shown to share—and thus be connected by—the same set of building blocks, yielding a 'web of theories' which can be analyzed with double-copy methods (see figure 17). In this section, we aim to probe this web more in detail. Particularly prominent results will be the classification of homogeneous $\mathcal{N}=2$ Maxwell-Einstein supergravities [282], which can be reproduced and streamlined by double-copy methods, the double-copy construction for YME [120, 125, 257, 283] and gauged supergravities [123, 284], and the construction for Dirac–Born–Infeld (DBI) theories [125, 285]. We will also see that some of the building blocks which appear, for example, in the double-copy construction for conformal supergravities play a role in a family of 'stringy' double-copy constructions. Similar webs of theories have appeared, for example, in the contexts of the scattering equations formalism [125], amplitude transmutation [285], and soft limits [286].

The simplest examples of double-copy-constructible theories we have discussed so far include $\mathcal{N}\geqslant 4$ supergravity and Einstein gravity coupled to a dilaton and two-form field. Once a double-copy structure has been established for a given gravitational theory, it is relatively straightforward to obtain the tree-level amplitudes of its consistent truncations. In this way, we can study amplitudes in a handful of additional theories. At the same time, it is well-known

Table 3. Freedom in specifying the two-derivative action in extended (ungauged) supergravities with $2 \leqslant \mathcal{N} \leqslant 8$ in four dimensions.

Supergravities	Free parameters	Scalar geometry
$\mathcal{N} > 4$ supergravities	None	Symmetric spaces
$\mathcal{N} = 4$ supergravity	Number of vector multiplets	Symmetric space
$\mathcal{N} = 3$ supergravity	Number of vector multiplets	Symmetric space
$\mathcal{N} = 2$, vector multiplets, 5D uplift	C_{IJK}-tensor	Very-special Kähler geometry
$\mathcal{N} = 2$, vector multiplets, 4D only	Free degree-two holomorphic function (prepotential)	Special Kähler geometry
$\mathcal{N} = 2$, hypermultiplets, from c-map	C_{IJK}-tensor or prepotential	Special/very special quaternionic Kähler geometry
$\mathcal{N} = 2$, hypermultiplets, general	See text	Quaternionic Kähler geometry

that supergravity theories with $\mathcal{N} < 4$ have a very rich structure which goes beyond the few theories that can be understood as truncations of more supersymmetric gravities. Ungauged supergravities with $\mathcal{N} \geqslant 5$ and two-derivative actions are unique. Starting from $\mathcal{N} = 4$, it becomes possible to have various matter contents. While $\mathcal{N} = 3,4$ supergravities are completely specified by the number of vector multiplets, additional information on interactions needs to be provided for theories with $\mathcal{N} = 2$ supersymmetry. Supergravity theories generically involve scalar fields, which can be regarded as the coordinates of a manifold. While extended $\mathcal{N} > 2$ supersymmetry allows only a discrete set of symmetric scalar manifolds, supersymmetry poses less stringent constraints when $\mathcal{N} \leqslant 2$. Specifically, in four dimensions, supergravities with vector multiplets possess special-Kähler scalar manifolds, while the geometry is quaternionic-Kähler in the case of supergravities with hypermultiplets [287].

In table 3, we list the information which needs to be provided to specify unambiguously ungauged supergravity theories with $2 \leqslant \mathcal{N} \leqslant 8$, together with the corresponding geometries. It should be noted that theories with $\mathcal{N} = 2$ have different geometrical properties depending on whether or not they have a five-dimensional uplift. Theories with vectors multiplets which can be lifted to five dimensions are uniquely specified by a symmetric constant tensor C_{IJK} whose indices run over the total number of vector fields. Since this tensor can be obtained from inspecting specific three-point interactions, supergravities of this sort have the pleasant property of being entirely constructible from their three-point amplitudes, a property that we will utilize extensively later in this section. Intrinsically-four-dimensional theories are significantly less constrained. They are fully specified by a homogeneous degree-two holomorphic function—the prepotential—which is otherwise arbitrary. Theories with hypermultiplets possess even more freedom: $\mathcal{N} = 2$ supersymmetry only require the hypermultiplet scalar manifold be quaternionic-Kähler (that is, to admit an hermitian metric and three complex structures which satisfy the quaternionic algebra). At the same time, a subset of these theories can be regarded as the image of supergravities with vector multiplets under an operation known as c-map; specifying these theories requires the same information as their vector counterparts. When studying supergravities with reduced supersymmetry it is important to keep in mind how this freedom is reflected in the gauge-theory data entering the double-copy construction.

Supergravities studied in the double-copy context have thus far been mostly theories of the Maxwell-Einstein class, i.e. theories in which all vector fields are abelian and there are no charged matter fields. From a Lagrangian perspective, supergravities with nonabelian gauge interactions have also been studied, see [287, 288] for reviews. They can be further divided into YME theories and proper gauged supergravities. In the former class, a nonabelian subgroup of the isometry group of the scalar manifold is promoted to a gauge symmetry. In the latter case, part of the R symmetry is promoted to gauge symmetry. This procedure, customarily referred to as gauging, does not introduce additional vector fields. It minimally couples some of the existing vector fields while also giving them nonabelian self-interactions and extending the resulting theory so that it is invariant under the required number of supercharges. In an amplitude context, YME theories have been studied from a variety of perspectives, including scattering equations [125, 283, 289], collinear limits of gauge theory amplitudes [290], BCFW recursion [214, 216], string theory [133, 134], ambitwistor strings [140] and, of course, the double-copy construction [120]. Through this work, it has become clear that amplitudes in such theories may be written as linear combinations of (color-ordered) amplitudes or ordinary YM theory [133, 214, 216, 289]. We will see later in this section that the above property has a very straightforward double-copy interpretation. Gauged supergravities display a considerably more involved structure. Once a subset of the R symmetry is gauged (i.e. some of the R-symmetry generators appear in the covariant derivatives), supersymmetry requires a scalar potential to appear in the theory. According to whether the potential vanishes or not at a critical point, the theory admits Minkowski, Anti-de Sitter or de Sitter vacua. Minkowski vacua break supersymmetry spontaneously (partly or completely), resulting in massive gravitini. The study of gauged supergravities in the double-copy framework is still in the early stages, but encouraging results are available which will be reviewed later in this section.

A growing body of work seems to suggest that the existence of a double-copy structure is not merely an accidental feature of highly-supersymmetric theories, but a generic property of very large classes of gravities. To determine whether the double-copy property is a hidden structure of gravitational interactions it is necessary to identify the gauge-theory counterparts of all data required to specify a generic gravity theory, whether it be ungauged, YME or gauged. While this program has not yet been completed, important progress has been made in formulating double-copy constructions for theories which include, among others, pure supergravities, homogeneous $\mathcal{N}=2$ Maxwell-Einstein supergravities, homogeneous $\mathcal{N}=2$ theories with hypermultiplets, large classes of YME or gauged theories, and conformal supergravities. A list of ungauged and gauged theories for which a double-copy construction is currently known can be found in tables 4 and 5, respectively. Gauge theories with fields in various matter (non-adjoint) representations of the gauge group are a rather common building block for this class of extended constructions. Useful tools for treating matter representations in a way that makes manifest color and numerator relations will be introduced in section 5.2. We will then discuss systematics of the process of identifying the gravity theory given, through double copy, by a pair of gauge theories and study several examples in section 5.3.

Double-copy constructibility is a property that goes beyond gravitational theories. Various theories without a graviton, most prominently some variants of the DBI theory have also been shown to possess this property (see table 6). We shall briefly review their construction in section 5.3.11.

5.1. The rules of the game

To capture as many gravities as possible, we need to consider gauge theories which are more general than the ones discussed at length in previous sections. At the same time, having in mind a double-copy construction which leads to a sensible gravity theory with desirable basic properties, it makes sense to impose some requirements on the gauge theories under consideration. Some additional requirements will also be imposed for simplicity reasons; in both cases, one can contemplate generalizations in which some of the stated rules of the game bent or broken.

Table 4. Non-exhaustive list of ungauged gravities and supergravities for which a double-copy construction is presently known. Theories are given in four dimensions unless otherwise stated.

Gravity	Gauge theories	References	Variants and notes
$\mathcal{N} > 4$ supergravity	• $\mathcal{N} = 4$ SYM theory • SYM theory ($\mathcal{N} = 1, 2, 4$)	[1, 2, 31, 291, 292]	
$\mathcal{N} = 4$ supergravity with vector multiplets	• $\mathcal{N} = 4$ SYM theory • YM-scalar theory from dim. reduction	[1, 2, 31, 293]	• $\mathcal{N} = 2 \times \mathcal{N} = 2$ construction is also possible
Pure $\mathcal{N} < 4$ supergravity	• (S)YM theory with matter • (S)YM theory with ghosts	[188]	• Ghost fields in fundamental rep
Einstein gravity	• YM theory with matter • YM theory with ghosts	[188]	• Ghost/matter fields in fundamental rep
$\mathcal{N} = 2$ Maxwell–Einstein supergravities (generic family)	• $\mathcal{N} = 2$ SYM theory • YM-scalar theory from dim. reduction	[120]	• Truncations to $\mathcal{N} = 1, 0$ • Only adjoint fields
$\mathcal{N} = 2$ Maxwell–Einstein supergravities (homogeneous theories)	• $\mathcal{N} = 2$ SYM theory with half hypermultiplet • YM-scalar theory from dim. reduction with matter fermions	[121, 294]	• Fields in pseudo-real reps • Include Magical Supergravities
$\mathcal{N} = 2$ supergravities with hypermultiplets	• $\mathcal{N} = 2$ SYM theory with half hypermultiplet • YM-scalar theory from dim. red. with extra matter scalars	[121, 240]	• Fields in matter representations • Construction known in particular cases
$\mathcal{N} = 2$ supergravities with vector/ hypermultiplets	• $\mathcal{N} = 1$ SYM theory with chiral multiplets • $\mathcal{N} = 1$ SYM theory with chiral multiplets	[239, 241, 295]	• Construction known in particular cases
$\mathcal{N} = 1$ supergravities with vector multiplets	• $\mathcal{N} = 1$ SYM theory with chiral multiplets • YM-scalar theory with fermions	[188, 239, 241, 295]	• Fields in matter reps • Construction known in particular cases
$\mathcal{N} = 1$ supergravities with chiral multiplets	• $\mathcal{N} = 1$ SYM theory with chiral multiplets • YM-scalar with extra matter scalars	[188, 239, 241, 295]	• Fields in matter reps • Construction known in particular cases
Einstein gravity with matter	• YM theory with matter • YM theory with matter	[1, 188]	• Construction known in particular cases

(Continued.)

Table 4. (Continued.)

$R+\phi R^2+R^3$ gravity	• YM theory $+F^3+F^4+\ldots$ • YM theory $+F^3+F^4+\ldots$	[296]	• Extension to $\mathcal{N}\leqslant 4$ replacing one of the factors by undeformed SYM theory
Conformal (super)gravity	• DF^2 theory • (S)YM theory	[152, 153]	• $\mathcal{N}\leqslant 4$ • Involves specific gauge theory with dimension-six operators
3D maximal supergravity	• BLG theory • BLG theory	[119, 243, 297]	• 3D only

Table 5. Gauged/YME gravities and supergravities for which a double-copy construction is presently known.

Gravity	Gauge theories	References	Notes
YME supergravities	• SYM theory • YM $+\phi^3$ theory	[120, 125, 133, 134, 140, 214, 216, 257, 283, 285, 289]	• Trilinear scalar couplings • $\mathcal{N}=0,1,2,4$ possible
Higgsed supergravities	• SYM theory (Coulomb branch) • YM $+\phi^3$ theory with extra massive scalars	[122]	• $\mathcal{N}=0,1,2,4$ possible • Massive fields in supergravity
$U(1)_R$ gauged supergravities	• SYM theory (Coulomb branch) • YM theory with SUSY broken by fermion masses	[123]	• $0\leqslant\mathcal{N}\leqslant 8$ possible • SUSY is spontaneously broken • Only theories with Minkowski vacua
Gauged supergravities (nonabelian)	• SYM theory (Coulomb branch) • YM $+\phi^3$ theory with massive fermions	[284]	• SUSY is spontaneously broken • Only theories with Minkowski vacua

First of all, for simplicity, we choose to focus on theories for which amplitudes can be organized exclusively in terms of cubic graphs. This is a natural generalization of the gauge theories from the previous sections, which possess this property, and is a natural choice for describing gravities that are entirely specified by their three-point interactions. Hence, we restrict the space of gauge theories under consideration according to the following rule:

> **Working rule 1:** consider gauge theories with only cubic invariant tensors or, alternatively, theories for which amplitudes can be organized in terms of cubic graphs.

Allowed invariant tensors will include, for example, structure constants, representation matrices and cubic Clebsch–Gordan coefficients. It should be emphasized that the gauge theories under consideration can and will possess quartic vertices. Our requirement constrains higher-point interaction vertices to be made of color building blocks which are cubic. If this property is satisfied, amplitudes can be expressed in terms of cubic graphs by including a suitable number of inverse propagators in the numerator factors. While this rule is quite desirable for the sake of simplicity, it can in principle be broken. A notable violation are the Bagger–Lambert–Gustavsson (BLG) and Aharony–Bergman–Jefferis–Maldacena (ABJM) theories, which are most naturally organized in terms of quartic graphs [119, 243, 297].

Within the class of cubic theories, however, we need to consider cases which are as general as possible. This motivates the second rule:

> **Working rule 2:** the gauge theories will include matter fields transforming in general (not necessarily irreducible) representations of the gauge group (which is not necessarily semisimple). Only one adjoint representation will be allowed.

Considering general gauge groups and representations will allow us to capture very large families of (super)gravities which would not otherwise be accessible through double-copy methods. The main observation is that there is nothing in the double-copy construction that requires that representations be divided into irreducible blocks. At the same time, we want to obtain theories with a single graviton. This forces us to combine all gauge-theory gluons in a single adjoint representation, even when the gauge group is the product of several factors each possessing its own adjoint representation. In case of more than one semi-simple factor in the gauge group, we need to take all gauge coupling constants to be the same. Since all fields in the gauge theory have canonical couplings with gluons, our second rule can also be regarded as the double-copy incarnation of the equivalence principle.

Additionally, massive fields are typically assigned to non-adjoint representations such that all the fields in a given representation have the same mass. This will be accompanied by mass-matching conditions of the spectrum of the two sides of the double copy.

Combining the first two rules, we obtain a generic amplitude structure that involves cubic graphs in which internal and external legs carry definite representations of the gauge group. Cubic vertices between three representations are allowed only when it is possible to extract a gauge singlet in their tensor product (or, alternatively, there exist a nonvanishing invariant tensor with the three corresponding indices). Whenever a vertex involves two lines carrying the same representation, its symmetry or antisymmetry will be dictated by the representations under consideration (real representations will imply antisymmetry, pseudo-real representation will imply symmetry). Additionally, color factors will obey three-term identities following from the Jacobi relations, the generators' commutation relations and additional algebraic relations which may also involve the Clebsch–Gordan coefficients. Consequently, the duality between color and kinematics must to be imposed in the following way:

> **Working rule 3:** numerator factors in a duality-satisfying presentation of an amplitude need to have the same algebraic properties as the color factors. This includes symmetry properties as well as obeying two- and three-term identities.

As discussed in section 2, this rule ensures that the gravity theory obtained through double copy is invariant under linearized diffeomorphisms. If there exists (massive) vector fields that transform in non-adjoint representations, additional gauge-group Lie algebra relations are needed to guarantee that gauge invariance aside from Jacobi and commutation relations. The same relations should be imposed on the kinematic numerators for all fields that transform in the same representations as the vectors. For some classes of constructions, it will be convenient to consider a slight variant of Working Rule 3 which instructs to impose the algebraic properties of the color factors of one theory on the numerators of the other theory entering the double-copy construction (and vice versa).

Finally, we need a procedure for consistently pairing representations in the two gauge theories when we substitute color factors with numerator factors following the double-copy prescription. *A priori*, several choices are possible. However, the following criterion is convenient, elegant and easy to implement:

> **Working rule 4:** each state in the double-copy (gravitational) theory corresponds to a gauge-invariant bilinear of gauge-theory states. For this to be possible, we will identify the gauge groups of the two theories entering the construction.

A concrete consequence of this rule is that gauge-theory states in the adjoint representation will double copy among themselves, but not with states in matter non-adjoint representations. Similarly, states in two matter representations will be combined only when the tensoring of the representations includes a singlet. Considering general graphs, two numerators will be combined only when working rule 4 is satisfied by each internal and external line. We will see that this requirement is essential for preventing the gravity from the double copy from having too many gravitini.

The space of all possible gauge theories is quite vast (though perhaps not quite as vast as that of gravitational theories). The purpose of the working rules we laid out is to restrict this space to a subset which is sufficiently large to capture a considerable number of theories and yet sufficiently small to allow a thorough analysis. It is not difficult to enlarge it by relaxing some of the rules. We emphasize that many gauge theories which might be naively rejected as unphysical, such as theories with ghost fields, may be admissible—even in some sense necessary—from a double-copy perspective. This is because, through double copy, gauge-theory data is deconstructed and reassembled in a highly-nontrivial way and undesirable features of gauge theories can be rendered harmless by this process.

The rules stated in this section should be slightly modified when constructing theories that are not gravitational. In this case, the gauge group should be replaced by a global symmetry group in the theories entering the construction.

5.2. Tools for extensions

Having established the general rules of the game, we will now analyze particular examples. We start by considering a YM-scalar theory with only adjoint fields and trilinear cubic couplings [120]. Its Lagrangian can be written as

$$\begin{aligned}\mathcal{L}_{\mathrm{YM}+\phi^3} = &-\frac{1}{4}F^{\hat{a}}_{\mu\nu}F^{\mu\nu\hat{a}} + \frac{1}{2}\left(D_\mu\phi^A\right)^{\hat{a}}\left(D^\mu\phi^A\right)^{\hat{a}} - \frac{g^2}{4}f^{\hat{a}\hat{b}\hat{e}}f^{\hat{e}\hat{c}\hat{d}}\phi^{A\hat{a}}\phi^{B\hat{b}}\phi^{A\hat{c}}\phi^{B\hat{d}}\\ &+\frac{1}{3!}\lambda g\, F^{ABC}f^{\hat{a}\hat{b}\hat{c}}\phi^{A\hat{a}}\phi^{B\hat{b}}\phi^{C\hat{c}}\,.\end{aligned} \tag{150}$$

The indices $\hat{a},\hat{b},\hat{c}$ are gauge-group adjoint indices. $A,B,C = 1,\ldots,n$ are global indices carried by the scalars[30]. The theory has a $SO(n)$ global symmetry which is broken by the trilinear couplings to the subgroup preserved by the F^{ABC} tensor. Field strengths and covariant derivatives are

$$F^{\hat{a}}_{\mu\nu} = \partial_\mu A^{\hat{a}}_\nu - \partial_\nu A^{\hat{a}}_\mu + gf^{\hat{a}\hat{b}\hat{c}}A^{\hat{b}}_\mu A^{\hat{c}}_\nu\,,$$

[30] In this section, we frequently use hatted indices for gauge-group indices of the gauge theories that enter the double copy, to help distinguish them from global indices and gauge indices that appears in gravitational theories.

$$\left(D_\mu \phi^A\right)^{\hat{a}} = \partial_\mu \phi^{A\hat{a}} + g f^{\hat{a}\hat{b}\hat{c}} A_\mu^{\hat{b}} \phi^{A\hat{c}} \,. \tag{151}$$

To understand the constraints imposed by CK duality on the parameters of this theory we first analyze the four-scalar amplitudes. There is a clean separation between the contribution from the trilinear scalar coupling and the one from gluon exchange (the latter including the contact term). After a short calculation, the s-channel numerator can be written as

$$n_s = \delta^{AB}\delta^{CD}\,(t-u) - \left(\delta^{AC}\delta^{BD} - \delta^{AD}\delta^{BC}\right) s - \lambda^2 F^{ABE}F^{ECD} \,, \tag{152}$$

while the other numerators can be obtained by relabeling the external lines. The three corresponding color factors obey standard Jacobi relations; imposing the duality between color and kinematics then results in the condition

$$\lambda^2 \left(F^{ABE}F^{ECD} + F^{BCE}F^{EAD} + F^{CAE}F^{EBD}\right) = 0 \,. \tag{153}$$

The λ^0 part of the numerator factors satisfies the duality automatically. This follows from the YM-scalar theory with $\lambda = 0$ being the dimensional reduction of a pure YM theory in higher dimension, which is known to satisfy the duality at arbitrary multiplicity. At order λ^2, the kinematic Jacobi relations imply that F^{ABC}-tensors must themselves obey Jacobi relations. This implies that they can be regarded as the structure constants of some global group which is unrelated to the gauge group. What remains to be done is to consider amplitudes involving vectors and amplitudes at higher points. It turns out that no further constraint on the theory appears. It has been explicitly checked that the Lagrangian obeys CK duality up to at least six points [120].

A second example which we review in detail is YM theory with complex scalars in a matter representation [122]; an analogous example involving matter fermions was discussed in section 2. For such a field content, trilinear couplings are forbidden by gauge symmetry; the first possible scalar self-interaction is quartic, so the Lagrangian is

$$\mathcal{L}_{\text{scalar}} = D_\mu \overline{\varphi} D^\mu \varphi - a\frac{g^2}{2}\left(\overline{\varphi} t^{\hat{a}} \varphi\right)\left(\overline{\varphi} t^{\hat{a}} \varphi\right) \,, \tag{154}$$

where φ is a complex scalar, $t^{\hat{a}}$ are representation matrices and a is a constant. Products are understood in the sense of matrix multiplication, as representation indices are not displayed explicitly to avoid cluttering the expression. The two-scalar two-gluon amplitude in this theory is[31]

$$\begin{aligned}
&\mathcal{A}_4\left(1\overline{\varphi}^{\hat{\imath}}, 2\varphi_{\hat{\jmath}}, 3A^{\hat{a}}, 4A^{\hat{b}}\right) \\
&\quad = i g^2 \left\{ \left(\frac{4\left(\varepsilon_4\cdot k_1\right)\left(\varepsilon_3\cdot k_2\right) + t\left(\varepsilon_3\cdot\varepsilon_4\right)}{t} \left(t^{\hat{a}}t^{\hat{b}}\right)_{\hat{\jmath}}^{\ \hat{\imath}} + (3\leftrightarrow 4) \right) \right. \\
&\qquad \left. + i\,\frac{4\left(\varepsilon_3\cdot k_1\right)\left(\varepsilon_4\cdot k_2\right) - 4\left(\varepsilon_4\cdot k_1\right)\left(\varepsilon_3\cdot k_2\right) + (u-t)\left(\varepsilon_3\cdot\varepsilon_4\right)}{s} f^{\hat{a}\hat{b}\hat{c}} \left(t^{\hat{c}}\right)_{\hat{\jmath}}^{\ \hat{\imath}} \right\} ,
\end{aligned} \tag{155}$$

where we have displayed explicitly the gauge representation indices $\hat{\imath}, \hat{\jmath}$. As a consequence of the commutation relation for the group generators, the color factors obey a three-term identity,

$$\left[t^{\hat{a}}, t^{\hat{b}}\right] = i f^{\hat{a}\hat{b}\hat{c}} t^{\hat{c}} \qquad \to \qquad c_t - c_u = c_s \,. \tag{156}$$

[31] We use the slightly-nonstandard notation, e.g. $\mathcal{A}_n\left(1\Phi_1, \ldots, n\Phi_n\right)$, which displays explicitly the external states.

It is easy to verify that the same identity is automatically satisfied by the numerators in the above amplitude,

$$n_t - n_u = n_s\,. \tag{157}$$

This is possibly the simplest nontrivial example of the duality between color and kinematics for theories with non-adjoint fields. While CK duality for the two-scalar two-gluon amplitude is a rather straightforward generalization of the case of amplitudes with fields in the adjoint representation, the four-scalar amplitude exposes new subtleties. This amplitude is

$$\mathcal{A}_4\left(1\overline{\varphi}^{\hat{\imath}}, 2\overline{\varphi}^{\hat{\jmath}}, 3\varphi_{\hat{k}}, 4\varphi_{\hat{l}}\right) = i\,g^2\left\{\frac{s-u-a\,t}{t}\left(t^{\hat{a}}\right)_{\hat{l}}^{\ \hat{\imath}}\left(t^{\hat{a}}\right)_{\hat{k}}^{\ \hat{\jmath}} + (3\leftrightarrow 4)\right\}. \tag{158}$$

Due to the scalar being complex, the amplitude involves only two terms; in principle, we may consider imposing the extra identity

$$\left(t^{\hat{a}}\right)_{\hat{l}}^{\ \hat{\imath}}\left(t^{\hat{a}}\right)_{\hat{k}}^{\ \hat{\jmath}} - \left(t^{\hat{a}}\right)_{\hat{l}}^{\ \hat{\jmath}}\left(t^{\hat{a}}\right)_{\hat{k}}^{\ \hat{\imath}} = 0\,. \tag{159}$$

However, this identity is not satisfied except for special gauge groups and representations, so it would seem that our Working Rule 3 does not compel us to impose (159) in the general case. At the same time, the numerator factors can easily obey the corresponding two-term kinematic identity if we fix $a = 1$. Whether of not this choice should be made depends on the situation in which the kinematic numerators are used. For example, we might choose to use this theory in a double-copy construction that involves massive W fields in a matter representation (we will see that this is required, for example, for constructing Higgsed supergravities). In these cases, the spontaneously-broken gauge symmetry results in Ward identities that can be satisfied only if the massive vectors belong to specific representations for which color factors obey additional relations which are of the form (159). Hence, it will be appropriate to impose two-term identities on the numerators of the second gauge theory entering the double copy. In contrast, whenever (159) is not necessary for deriving some Ward identity in the gravity theory, there is no particular reason for imposing the corresponding numerator identity.

So far, we have presented two examples of theories which obey CK duality. In some cases, it is sufficient to write down simple gauge theories, verify that they obey the duality up to at least a certain multiplicity, and feed the corresponding numerators in the double-copy apparatus. However, this approach quickly becomes inconvenient as the number of matter representations increases. Hence, we would like to have systematic tools for obtaining more general theories which obey the duality from simpler ones. These tools will be reviewed in the next three subsections.

5.2.1. Breaking representations into pieces. A first step for generating theories with fields transforming in matter representations in a way that preserves the duality is to start from the adjoint representation of a larger gauge group and decompose it into representations of a subgroup. This amounts to splitting the adjoint index of the larger groups $\hat{A}$ as

$$\hat{A} \;\rightarrow\; (\hat{a}, \hat{\alpha}_1, \ldots, \hat{\alpha}_p)\,, \tag{160}$$

where $\hat{a}$ is the adjoint index of the smaller subgroup and $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ are indices of other representations. While it is always possible to choose them to correspond to irreducible representations, we will not do so here. The structure constants of the original gauge group are broken down as follows:

$$\left\{f^{\hat{A}\hat{B}\hat{C}}\right\} \;\rightarrow\; \left\{f^{\hat{a}\hat{b}\hat{c}}, f^{a\alpha_i\,\hat{\beta}_i}, f^{\hat{\alpha}_i\,\hat{\beta}_j\hat{\gamma}_k}\right\}. \tag{161}$$

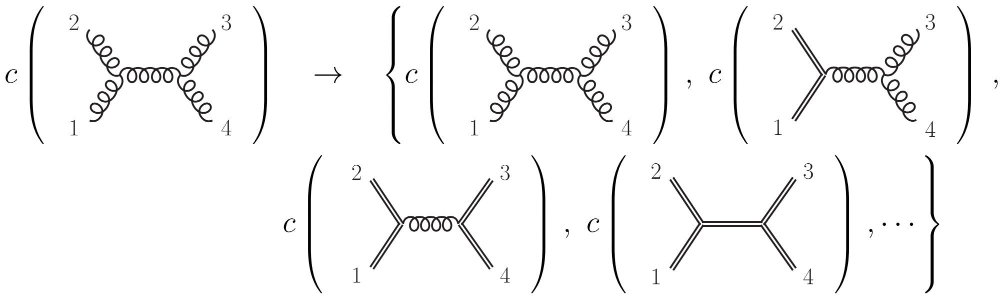
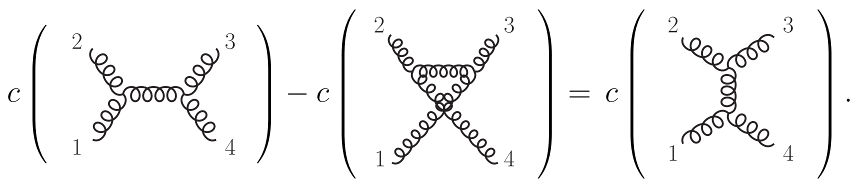


Figure 18. Breaking of an adjoint representation into representations of a smaller subgroup. Curly lines denote the adjoint representation of the smaller group; double lines denote matter representations.

Here $f^{\hat{a}\hat{b}\hat{c}}$ are the structure constants of the unbroken subgroup, $f^{\hat{a}\hat{\alpha}_i\hat{\beta}_i}$ give the representation matrices of the ith matter representation and $f^{\hat{\alpha}_i\hat{\beta}_j\hat{\gamma}_k}$ give Clebsch–Gordan coefficients for representations i,j, and k. We note that $f^{\hat{a}\hat{b}\hat{\alpha}}=0$ from closure of the algebra of the unbroken gauge group. The notation above suggests that we have assumed the matter representations above to be real; the complex case can be treated analogously by introducing a pairing between some representations i,j,k and their conjugate, denoted as $\bar{\imath},\bar{\jmath},\bar{k}$. The breaking of the adjoint representation acts in the following way on color and numerator factors:

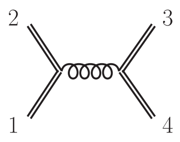
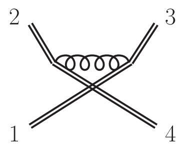
- Color factors are split into different pieces according to the representations carried by internal and external lines (see figure 18); color identities are preserved by this operation, but one needs to take into account that some color factor may vanish upon direct evaluation.
- Numerator factors are unchanged. Graphs with the same topology but different representation labels will inherit the same numerator factors as the original graphs of the unbroken theory. Whenever numerators obey a three-term identity in the unbroken theory, the identity will be inherited by the broken theory.

Two-term identities of the form (159) deserve a more detailed discussion. Before decomposition into representations of a subgroup, color factors obey the standard Jacobi relations, which at four points can be written as



(162)

If we consider the case in which the initial adjoint representation is broken into three pieces (adjoint, a single complex matter representation, its conjugate), the corresponding identity for external matter is

$c\Bigg($  $\Bigg) = c\Bigg($  $\Bigg)\,,$ (163)

because the structure constant does not contain a component with two indices in the same complex representation. Hence, the original three-term color identity has collapsed into a two-term identity. However, the decomposition of color factors with respect to a subgroup does not affect the kinematic numerators so, if the original theory obeys CK duality, the numerators still obey three-term kinematic identities after the decomposition. In other words, a nonvanishing numerator is associated to a vanishing color factor. We will see that this will require extra care, e.g. in the construction of Higgsed supergravities in section 5.3.7. Note that a nonzero kinematic numerator can be associated with a vanishing color factor also in theories with only adjoint fields, as we shall see in section 6.

5.2.2. Field-theory orbifolds. If we consider a gauge theory which possesses a certain number of global/flavor symmetries (which may include the R symmetry), it is always possible to truncate it to its sector which is invariant under the combined action of some elements of the global and gauge groups. Specifically, a generic adjoint field Φ of the original theory will transform as

$$\Phi \to RFg\Phi g^{\dagger}\,, \tag{164}$$

where g is the gauge-group element and R,F are the corresponding elements of the R-symmetry and global-flavor group (R-symmetry and global indices are not explicitly displayed). It is convenient to consider elements (g,R,F) which belong to a discrete subgroup Γ of the symmetry group of the theory we are considering, that is we have elements g,R,F such that $g^k = I, F^k = I, R^k = I$ for some k[32]. Theories obtained with this construction are referred to as field-theory orbifolds in the literature [311]. Given (g,R,F) above, we can immediately write a projector

$$\mathcal{P}_\Gamma \Phi = \frac{1}{|\Gamma|}\sum_{(g,R,F)\in\Gamma} RFg\Phi g^{\dagger}\,, \tag{165}$$

where $|\Gamma|$ denotes the rank of Γ. It is easy to verify that this projector sets to zero all components of Φ which are not invariant under Γ.

To give a simple example, we start from $\mathcal{N}=4$ SYM theory with $SU(2N)$ gauge group and consider an $\Gamma = \mathbb{Z}_2$ orbifold with generators

$$r = \mathrm{diag}\,(1,1,-1,-1)\,, \qquad g = \begin{pmatrix} I_N & 0 \\ 0 & -I_N \end{pmatrix}. \tag{166}$$

The matrix r gives the action of the unique nontrivial generator of $\mathbb{Z}_2$ on the fundamental R-symmetry indices; the action of $\mathbb{Z}_2$ on other representation of the R symmetry can be obtained

[32] Other subgroups can also be considered.

by taking tensor products of r. In this case, it is convenient to represent the action of the projector on the components of a on-shell $\mathcal{N}=4$ superfield $\mathcal{V}^{\hat{A}}_{\mathcal{N}=4}$ which is written as (460). The part of this superfield which survives the orbifold projection (165) is

$$\begin{aligned}\mathcal{V}^{\hat{A}}_{\mathcal{N}=4} \to A^{\hat{a}}_{+} + \eta_i \lambda^{\hat{a}i}_{+} + \eta_r \lambda^{\hat{\alpha}r}_{+} + \eta_1\eta_2\phi^{\hat{a}12} + \eta_i\eta_r\phi^{\hat{\alpha}ir} \\ + \eta_3\eta_4\phi^{\hat{a}34} + \eta_1\eta_2\eta_r\lambda^{\hat{\alpha}r}_{-} + \eta_i\eta_3\eta_4\lambda^{\hat{a}i}_{-} + \eta_1\eta_2\eta_3\eta_4 A^{\hat{a}}_{-}\,,\end{aligned} \tag{167}$$

where $i,j=1,2$ and $r,s=3,4$. The gauge-group indices $\hat{a},\hat{b}$ and $\hat{\alpha},\hat{\beta}$ run over the (reducible) $SU(N)\times SU(N)\times U(1)$ adjoint representation and the bi-fundamental representation, respectively. This result can be organized in $\mathcal{N}=2$ on-shell superfields as

$$\mathcal{V}^{\hat{A}}_{\mathcal{N}=4} \to \mathcal{V}^{\hat{a}}_{\mathcal{N}=2} + \eta_r\Phi^{\hat{\alpha}r}_{\mathcal{N}=2} + \eta_3\eta_4\overline{\mathcal{V}}^{\hat{a}}_{\mathcal{N}=2}\,, \qquad r=3,4, \tag{168}$$

where $\Phi_{\mathcal{N}=2}$ is the on-shell hypermultiplet superfield. Hence, we see that the theory resulting from the orbifold projection (165) with $\Gamma=\mathbb{Z}_2$ acting as (166) is an $\mathcal{N}=2$ SYM theory with gauge group $SU(N)\times SU(N)\times U(1)$ and one matter hypermultiplet in the bi-fundamental representation.

Exercise 5.1. Work out spectrum and on-shell superfield organization for the $\mathbb{Z}_2$ orbifold projection of $\mathcal{N}=4$ SYM theory with generators

$$r=\mathrm{diag}\,(-1,-1,-1,-1)\,, \qquad g=\begin{pmatrix} I_N & 0 \\ 0 & -I_N\end{pmatrix}. \tag{169}$$

What is the residual supersymmetry?

Exercise 5.2. Formulate an orbifold projection of $\mathcal{N}=4$ SYM preserving $\mathcal{N}=1$ supersymmetry. Work out the multiplet structure of the on-shell superfields.

Field-theory-orbifold amplitudes are constructed through a set of Feynman rules which are obtained directly by taking the Feynman rules of the parent theory and dressing both internal and external lines with projectors of the form (165). For a tree-level amplitude, one can use invariance of the propagators and vertices under global and gauge symmetries to move all projectors from internal to external lines. The result is that all tree-level amplitudes of a theory constructed as an orbifold can be obtained from the amplitudes of the parent theory by inserting projectors on the external legs or, alternatively, by ensuring that the asymptotic states are invariant under the orbifold group. Particularly relevant to us, this property has the consequence that all numerator relations of the parent theory are preserved by the orbifold construction [30].

Exercise 5.3. Consider $\mathcal{N}=4$ SYM theory with $SU(2N+1)$ gauge group. Show that the projection with orbifold group generators

$$r=diag\,(1,1,-1,-1)\,, \qquad g=\begin{pmatrix} -I_{2N} & 0 \\ 0 & 1\end{pmatrix}, \tag{170}$$

yields a $\mathcal{N}=2$ theory with a hypermultiplet in the fundamental representation.

The reader may wonder whether there is a straightforward way to extend this result to loop level. At one loop, using the symmetries of propagators and vertices, projectors can be removed from all but one internal line (which can be chosen freely). Additionally, particular classes of loop-level amplitudes (for example, planar amplitudes in the large-N limit) are inherited from the parent theory (for a subset of so-called regular orbifolds) [311, 312]. Loop-level amplitudes can of course be constructed from tree-level ones with unitarity methods. For general amplitudes and choices of orbifold groups, the properties at loop-level will not be directly related to the ones of the parent theory and the orbifold construction will be used to obtain tree-level building blocks to be employed with unitarity methods. This construction has been instrumental, for example, in the study of one-loop amplitudes for supergravities that can be embedded in the $\mathcal{N}=8$ maximal theory [30, 241].

5.2.3. Masses as compact momenta. Once representations of a larger gauge group are broken into smaller pieces, it is in principle possible to introduce nonzero masses for some of the fields. At the same time, if we intend to consider more general theories of gravity coming from the double copy, we need some procedure for generating massive states (for which the most natural choice is the Higgs mechanism). If we consider gauge theories that can be written in higher dimension, a straightforward way to create mass terms from an amplitude perspective consists of assigning to some of the fields momenta in one of the extra (compact) dimensions.

Our starting point is to consider adjoint fields in a higher-dimensional theory which are written as

$$\begin{aligned} A^{\mu\hat{A}}\left(\vec{x},x_{D+1}\right)\Big|_{D+1} &= \left(e^{i x_{D+1} m}\right)^{\hat{A}\hat{B}} A^{\mu\hat{B}}\left(\vec{x}\right), \\ \phi^{a\hat{A}}\left(\vec{x},x_{D+1}\right)\Big|_{D+1} &= \left(e^{i x_{D+1} m}\right)^{\hat{A}\hat{B}} \phi^{a\hat{B}}\left(\vec{x}\right), \quad a=1,\ldots,n, \end{aligned} \tag{171}$$

where $m^{\hat{A}\hat{B}}$ is a mass matrix with $\hat{A},\hat{B}$ indices in the adjoint, $\vec{x}$ is the D-dimensional coordinate and x_{D+1} is the internal direction. If the mass matrix vanishes, this is equivalent to ordinary dimensional reduction. The condition above can also be implemented in position space through the differential equation

$$\partial_{D+1}\begin{pmatrix} A^{\mu\hat{A}} \\ \phi^{a\hat{A}} \end{pmatrix}\Bigg|_{D+1} = i\, m^{\hat{A}\hat{B}}\begin{pmatrix} A^{\mu\hat{B}} \\ \phi^{a\hat{B}} \end{pmatrix}. \tag{172}$$

Introducing this mass term has the effect of breaking the adjoint representation of the gauge group into various representations with respect to which $m^{\hat{A}\hat{B}}$ is block-diagonal. We choose $m^{\hat{A}\hat{B}}$ to be given by

$$m^{\hat{A}\hat{B}} = igVf^{\hat{0}\hat{A}\hat{B}}. \tag{173}$$

Fields that commute with the gauge-group generator $t^{\hat{0}}$ will not have a mass since that implies that $f^{0\hat{A}\hat{B}}$ vanish. We can now explicitly show that the kinetic term of the scalars in $(D+1)$ dimensions is identical to a kinetic term in D dimensions plus a ϕ^4-term in which a scalar acquires a VEV:

$$\begin{aligned} \frac{1}{2}\left(\mathcal{D}_\mu\phi^{a\hat{A}}\right)^2\Big|_{D+1} &\to \frac{1}{2}\left(\mathcal{D}_\mu\phi^{a\hat{A}}\right)^2 - \frac{1}{2}\left(i\, m^{\hat{A}\hat{B}}\phi^{a\hat{B}} + gf^{\hat{A}\hat{B}\hat{C}}\phi^{0\hat{B}}\phi^{a\hat{C}}\right)^2 \\ &= \frac{1}{2}\left(\mathcal{D}_\mu\phi^{a\hat{A}}\right)^2 + \frac{g^2}{2}\mathrm{tr}\left(\left[Vt^0+\phi^0,\phi^a\right]^2\right), \end{aligned} \tag{174}$$

where we have renamed the gauge field in the internal direction, $A^{\hat{A}}_{D+1} \to \phi^{0\hat{A}}$ (the global index a does not include $a = 0$). We then inspect the $(D+1)$-dimensional vector-field kinetic term,

$$
\begin{aligned}
-\frac{1}{4}\left(\mathcal{F}^{\hat{A}}_{\mu\nu}\right)^2\Big|_{D+1} &\to -\frac{1}{4}\left(\mathcal{F}^{\hat{A}}_{\mu\nu}\right)^2 + \frac{1}{2}\left(\partial_\mu\phi^{0\hat{A}} - i m^{\hat{A}\hat{B}} A^{\hat{B}}_\mu + g f^{\hat{A}\hat{B}\hat{C}} A^{\hat{B}}_\mu \phi^{0\hat{C}}\right)^2 \\
&= -\frac{1}{4}\left(\mathcal{F}^{\hat{A}}_{\mu\nu}\right)^2 + \frac{1}{2}\left(\left(\mathcal{D}_\mu\phi^0\right)^{\hat{A}} - i m^{\hat{A}\hat{B}} A^{\hat{B}}_\mu\right)^2 \\
&= -\frac{1}{4}\left(\mathcal{F}^{\hat{A}}_{\mu\nu}\right)^2 + \frac{1}{2}\left(\left(\mathcal{D}_\mu\phi^0 + \mathcal{D}_\mu\langle\phi^0\rangle\right)^{\hat{A}}\right)^2 .
\end{aligned} \tag{175}
$$

This term is identical to the D-dimensional vector-field kinetic term plus the kinetic term for ϕ^0 in the presence of a VEV

$$
\langle\phi^0\rangle = V t^{\hat{0}} . \tag{176}
$$

Adding the quartic potential terms for the scalars, one sees that the $(D+1)$-dimensional massless (S)YM Lagrangian in the presence of a compact momentum of the form (171) is indeed equivalent to a spontaneously-broken D-dimensional SYM Lagrangian with VEV given by (176). Strictly speaking, we have shown that this procedure works only in the presence of a quartic potential generated by dimensional reduction of a higher-dimensional pure (S)YM theory. The case of more general scalar potentials need to be considered separately. The argument in this subsection gives a prescription for finding amplitudes of theories with fields becoming massive though the Higgs mechanism, in terms of higher-dimensional massless amplitudes. If the higher-dimensional theory obeys CK duality, the massive amplitudes will inherit the same algebraic properties [122]. BCJ amplitude relations with massive particles were also derived in [116] using the CHY formalism.

5.2.4. Identifying the right supergravity. At this point, we have developed some basic techniques to start from the amplitudes of an arbitrary theory which is known or can be shown to obey the duality between color and kinematics and generate the amplitudes of more involved theories, which may include fields in non-adjoint representations and mass terms from the Higgs mechanism, in a way that preserves numerator relations. In principle, we can use the numerators from various theories obtained with this procedure for producing amplitudes through the double-copy technique. As discussed in sections 2 and 4, these amplitudes will obey the Ward identities related to invariance under linearized diffeomorphisms and hence should be the amplitudes from some gravitational theory. Identifying the precise theory, however, is not always straightforward. In principle, one could consider a generic Lagrangian involving the Einstein–Hilbert term (or, in case of conformal gravity, some Weyl^2 gravitational action) and arbitrary matter interactions. Up to terms which vanish due to the equations of motion, this Lagrangian can be fixed order by order by comparing its amplitudes with the ones from the double-copy method. In practice, the implementation of this program is limited by one's desire to evaluate higher-point tree amplitudes. For many theories however, minimal information about symmetries and lower-point interactions can be sufficient for identifying the theory completely and, in principle, for writing down its Lagrangian (with some help from the relevant supergravity literature). More specifically:

- Symmetry considerations are sufficient for identifying supergravities with extended $\mathcal{N} \geqslant 4$ supersymmetry and particular theories with reduced supersymmetry that can be viewed as truncations of more supersymmetric theories [31, 239, 241]. Such considerations can also be sufficient to formulate constructions for theories with homogeneous scalar manifolds, with some residual freedom that needs to be fixed with minimal information on their interactions [121, 240].

- Very broad classes of Maxwell-Einstein supergravities with $\mathcal{N}=2$ supersymmetry which can be lifted up to at least five spacetime dimensions can be uniquely specified by their three-point interactions (specifically, three-vector amplitudes in five-dimensions). In a sense, these theories constitute a natural testing ground for double-copy constructions with reduced supersymmetry [120, 121].
- More generally, there exist amplitudes which capture physical features of the desired supergravity theory. For example, YME theories or gauged supergravities with nonabelian gauge group will possess non vanishing three-point amplitudes between three gluons [120]. Gauged supergravities will have nonvanishing amplitudes between two gravitini and one vector [123]. Knowledge of these amplitudes can either allow identification of the theory or point to the general class to which the theory belongs.
- Some theories are characterized in terms of their soft limits. These include e.g. theories with homogeneous target spaces [121, 222], the NLSM and some of its extensions [125, 285, 306], and the special Galileon theory [309].

5.3. Examples

We now proceed to discussing some examples. A list of the main double-copy constructible theories at the time of this writing can be found in tables 4–6.

5.3.1. Theories with $\mathcal{N}\geqslant 4$ supersymmetry.

Pure supergravities with $\mathcal{N}=4$ and 8 have been originally formulated from a Lagrangian perspective in [313, 314] and [258, 315], respectively, while the $\mathcal{N}=5$ and $\mathcal{N}=6$ Lagrangians were obtained by truncation [316] from that of the $\mathcal{N}=8$ supergravity. Amplitudes of theories with extended $\mathcal{N}\geqslant 4$ supersymmetry will be given by a double copy involving $\mathcal{N}=4$ SYM theory together with a YM or SYM theory. The possibilities are the following [31]:

$$
\begin{aligned}
\mathcal{N}=8 \text{ supergravity}: &\quad (\mathcal{N}=4\ \text{SYM})\otimes(\mathcal{N}=4\ \text{SYM})\,,\\
\mathcal{N}=6 \text{ supergravity}: &\quad (\mathcal{N}=4\ \text{SYM})\otimes(\mathcal{N}=2\ \text{SYM})\,,\\
\mathcal{N}=5 \text{ supergravity}: &\quad (\mathcal{N}=4\ \text{SYM})\otimes(\mathcal{N}=1\ \text{SYM})\,,\\
\mathcal{N}=4 \text{ supergravity}: &\quad (\mathcal{N}=4\ \text{SYM})\otimes(\mathcal{N}=0\ \text{YM})\,.
\end{aligned}
\tag{177}
$$

All double copies above involve gauge theories with only adjoint fields. These are cases in which the symmetries of the desired supergravity single out the correct construction, without any free parameters. Supergravities with $\mathcal{N}>4$ are unique. For $\mathcal{N}=4$ supergravities one can add matter in the form of $\mathcal{N}=4$ vector multiplets, which correspond to adding adjoint scalars in the non-supersymmetric gauge theory.

Perturbative mass spectra and on-shell superfield structure of the theories listed above can be straightforwardly obtained from the on-shell superfields of the gauge theories. Alternatively, all theories with $\mathcal{N}>4$ and some examples of $\mathcal{N}=4$ theories can be seen as truncations of $\mathcal{N}=8$ supergravity using a field-theory orbifold construction.

$\mathcal{N}=4$ supergravity has an alternative double-copy construction, in terms of two $\mathcal{N}=2$ SYM theories coupled to hypermultiplets in matter representations. Apart from the mass spectra, it has been verified that tree-level and four-point one-loop amplitudes in the two realizations are the same, including anomalous amplitudes [20, 241].

Table 6. List of non-gravitational theories constructed as double copies.

Double copy	Starting theories	References	Variants and notes
DBI theory	• NLSM • (S)YM theory	[125, 126, 285, 298–301]	• $\mathcal{N} \leqslant 4$ possible • Also obtained as $\alpha' \to 0$ limit of abelian Z-theory
Volkov–Akulov theory	• NLSM • SYM theory (external fermions)	[125, 302–308]	• Restriction to external fermions from supersymmetric DBI
Special Galileon theory	• NLSM • NLSM	[125, 285, 301, 306, 309]	• Theory is also characterized by its soft limits
DBI + (S)YM theory	• NLSM + ϕ^3 • (S)YM theory	[125, 126, 156, 285, 298–300, 306, 310]	• $\mathcal{N} \leqslant 4$ possible • Also obtained as $\alpha' \to 0$ limit of semi-abelianized Z-theory
DBI + NLSM theory	• NLSM • YM + ϕ^3 theory	[125, 126, 156, 285, 298–300]	

We now look more in detail at the pure $\mathcal{N} = 4$ supergravity in four dimensions. The theory involves one complex scalar, whose asymptotic states are obtained by taking the double copy of gauge-theory gluons with opposite polarizations. Geometrically, the scalar can be regarded as the complex coordinate of the coset space

$$\boldsymbol{\mathcal{M}}_{4D} = \frac{SU(1,1)}{U(1)}\,. \tag{178}$$

As we will discuss more in detail later, the fact that the scalar lives in an homogeneous space can be confirmed by checking the vanishing of the scalar soft limits at tree level[33]. More explicitly, the bosonic part of the Lagrangian for pure $\mathcal{N} = 4$ supergravity has a relatively simple form,

$$\begin{aligned} e^{-1}\mathcal{L} &= -\frac{R}{2} + \frac{1}{4}\frac{\partial_\mu\tau\partial^\mu\bar{\tau}}{(\mathrm{Im}\,\tau)^2} - \frac{1}{4}\mathrm{Im}\,\tau F^I{}_{\mu\nu}F^{I\mu\nu} - \frac{1}{8}\mathrm{Re}\,\tau e^{-1}\,\epsilon^{\mu\nu\rho\sigma}F^I{}_{\mu\nu}F^I{}_{\rho\sigma} \\ &= -\frac{R}{2} + \frac{1}{4}\left(\frac{\partial_\mu\tau\partial^\mu\bar{\tau}}{(\mathrm{Im}\,\tau)^2} + i\tau\left(F^{+I}_{\mu\nu}\right)^2 - i\bar{\tau}\left(F^{-I}_{\mu\nu}\right)^2\right), \end{aligned} \tag{179}$$

where $\tau = ie^{-\phi} + \chi$ is the dilaton-axion scalar, $\tilde{F}^I{}_{\mu\nu} = (i/2)e\epsilon_{\mu\nu\rho\sigma}F^{I\rho\sigma}$, $F^{\pm I}_{\mu\nu} = (F^I{}_{\mu\nu} \pm \tilde{F}^I{}_{\mu\nu})/2$, and $I = 1,\ldots,6$ is an index running over the vector fields in the theory. Alternatively, the kinetic term for the scalars can be written with a Cayley parameterization of the form (115).

In contrast to more supersymmetric settings, the double-copy construction for $\mathcal{N} = 4$ supergravities can be easily modified by adding extra adjoint scalars in the non-supersymmetric gauge theory. This can be done by considering a YM theory coupled to N scalars, which is

[33] In $\mathcal{N} = 4$ supergravity the single-soft-scalar limit no longer vanishes at one loop [260] due to an anomaly of the $U(1)$ symmetry in equation (178). Finite local counterterms can be used to restore this symmetry (at the expense of the other $SU(1,1)$ generators) [265]. This counterterm also restores the vanishing single-soft-scalar limit.

the reduction to four dimensions of $D = (N+4)$ pure YM theory[34]. This theory is invariant under an $SO(N)$ symmetry, which is the subgroup of the D-dimensional Lorentz group transverse to four dimensions, $SO(1,3+N) \to SO(1,3) \times SO(N)$. Under this symmetry, the vector fields are inert while the scalars transform in the vector (fundamental) representation. Since these scalars transform in the adjoint representation of the gauge group, they can be double-copied with the $\mathcal{N} = 4$ vector multiplet to yield N vector multiplets in the supergravity theory. Their scalars transform in the $(\boldsymbol{N}, \boldsymbol{6})$ representation of $SO(N) \times SO(6)$, where latter factor is the R-symmetry group[35].

In $\mathcal{N} = 4$ supergravity scalars fields outside the graviton multiplet parametrize a homogeneous space of the form

$$\mathcal{M}_{4D} = \frac{G}{H}, \tag{180}$$

where the stabilizer group H is the symmetry which is linearly realized and thus visible in amplitudes involving scalars. Thus, the symmetry which is manifest in the double-copy construction cannot be larger than H. In our case, we have $SO(N) \times SO(6) \subseteq H$. This suggests that the $6N$ vector multiplet scalars parametrize $SO(6,N)/(SO(6) \times SO(N))$ and, together with equation (178), that the $(6N+2)$ real scalars in the four-dimensional $\mathcal{N} = 4$ supergravity theory parametrize the symmetric space

$$\mathcal{M}_{4D} = \frac{SO(6,N)}{SO(6) \times SO(N)} \times \frac{SU(1,1)}{U(1)}. \tag{181}$$

The double-copy construction for $\mathcal{N} \geqslant 4$ supergravities can be used to find expressions for amplitudes at one loop, which are discussed in section 6. Beyond one loop, amplitudes in extended supergravity theories have been the subject of intense investigation, especially on their UV properties. Lore has it that all supergravity must diverge at a sufficiently high loop order. Is this actually true or might there be surprises? A variety of multiloop calculations for $\mathcal{N} \geqslant 4$ supergravity have been carried out to analyze UV properties:

- Four-point amplitudes for pure $\mathcal{N} = 4$ supergravity have been shown to be UV-finite at three loops and UV-divergent at four loops in four dimensions [33, 36, 37, 293]. The four-loop UV-divergence appears to be related to a $U(1)$ anomaly [260, 263, 265, 266]. Full one- and two-loop four-point amplitudes in $\mathcal{N} = 4$ supergravity are given in [31, 32].
- Four-point amplitudes for $\mathcal{N} = 5$ supergravity are finite at least through four loops in four dimensions [292]. Despite various attempts, there is no standard symmetry explanation for the 'enhanced cancellations' that lead to this improved UV behavior [317]. See, however, [318, 319] for arguments suggesting that U-duality invariance may be ultimately responsible. It is of considerable interest to settle the origin of these cancellations, and to know whether they continue to higher orders.
- The complete two-loop four-point amplitude of $\mathcal{N} = 6$ supergravity may be found in [32]. As yet there have not been any direct studies of the critical dimension of this theory at high loop orders, although it follow from the calculations in $\mathcal{N} = 5$ supergravity that divergences

[34] Recall that dimensional reduction is an operation which is known to preserve CK duality. These theories will sometimes be denoted as $\mathrm{YM}_{\mathrm{DR}}$.

[35] Vector fields in this theory are of two types: graviphotons, which are part of the gravition multiplet and transform in $(\boldsymbol{1}, \boldsymbol{6})$ and vectors which are part of the additional vector multiplets, which transform as $(\boldsymbol{N}, \boldsymbol{1})$.

cannot appear before five loops. Standard symmetry considerations imply that divergences are delayed until at least five loops [317, 320].

- UV properties of four-point amplitudes in $\mathcal{N}=8$ supergravity have been analyzed in detail through five loops [38]. In contrast to the case of $\mathcal{N}=5$ supergravity, $\mathcal{N}=8$ supergravity at five loops does not appear have enhanced cancellations, but it is possible that this is an artifact of the fact that the analysis is carried out in the fractional critical dimension $D=24/5$, where from various considerations [321, 322] divergences are first expected to appear. A proper study of this issue in the most interesting dimension $D=4$ requires a seven-loop computation, as suggested by symmetry considerations [274, 321–326]. The complete three-loop four-point and two-loop five-point amplitudes of $\mathcal{N}=8$ supergravity have been obtained [327, 328], starting from integrands constructed via the double copy [2, 4]. The construction of $\mathcal{N}=8$ one-, two- and three-loop integrands via the double copy is described in section 6.

5.3.2. Maxwell–Einstein theories with $\mathcal{N}=2$ supersymmetry. In this section we discuss amplitudes in theories with $\mathcal{N}=2$ supersymmetry in four dimensions (eight supercharges). Theories of this type are no longer specified solely by their matter content. Hence, we need a strategy to conveniently classify the interactions consistent with $\mathcal{N}=2$ supersymmetry. An efficient approach is to focus on theories that can be uplifted to five dimensions. The Lagrangians for these theories have long been known explicitly [329–332]. Here we will write only the bosonic part of the Lagrangian[36]:

$$e^{-1}\mathcal{L}=-\frac{1}{2}R-\frac{1}{4}\overset{\circ}{a}_{IJ}F^I_{\mu\nu}F^{J\mu\nu}+\frac{1}{2}g_{xy}\partial_\mu\phi^x\partial^\mu\phi^y+\frac{e^{-1}}{6\sqrt{6}}C_{IJK}\varepsilon^{\mu\nu\rho\sigma\lambda}F^I_{\mu\nu}F^J_{\rho\sigma}A^K_\lambda. \tag{182}$$

All vectors in the Lagrangian are taken to be abelian (YME theories will be discussed in section 5.3.6). The index $I=0,1,\ldots,n$ runs over the number of vectors in the theory with $I=0$ corresponding to the graviphoton. $F^I_{\mu\nu}$ are the field strengths, while $\overset{\circ}{a}_{IJ}$ and g_{xy} are functions of the physical scalars ϕ^x $(x=1,\ldots,n)$. The key insight is that the symmetric constant tensor C_{IJK} is sufficient to specify the theory completely, i.e. to fix all functions appearing in the two-derivative Lagrangian. The formalism manifesting this feature introduces an auxiliary ambient space with coordinates ξ^I and dimension equal to the number of vectors in the theory which, together with the C_{IJK}-tensor, are used to define a cubic polynomial

$$\mathcal{V}(\xi)\equiv C_{IJK}\xi^I\xi^J\xi^K\,. \tag{183}$$

In turn, this is used to define a metric on the ambient space:

$$a_{IJ}(\xi)\equiv-\frac{1}{3}\frac{\partial}{\partial\xi^I}\frac{\partial}{\partial\xi^J}\ln\mathcal{V}(\xi)\,. \tag{184}$$

The scalar manifold $\boldsymbol{\mathcal{M}}_{5D}$ is defined as the hypersurface obeying the equation

$$\mathcal{V}(h)=C_{IJK}h^Ih^Jh^K=1\,,\qquad h^I=\sqrt{\frac{2}{3}}\xi^I. \tag{185}$$

[36] For consistency with the rest of this review, this Lagrangian is written with a metric of mostly-minus signature, in contrast to most of the supergravity literature.

The functions $\overset{\circ}{a}_{IJ}(\phi)$ and $g_{xy}(\phi)$ which appear in the Lagrangian are given by the restriction of the ambient-space metric to $\boldsymbol{\mathcal{M}}_{5D}$ and the pullback to that surface of the ambient space metric, respectively:

$$\overset{\circ}{a}_{IJ}(\phi) = a_{IJ}\big|_{\mathcal{V}(h)=1}; \qquad g_{xy}(\phi) = \frac{3}{2}\frac{\partial\xi^I}{\partial\phi^x}\frac{\partial\xi^J}{\partial\phi^y}a_{IJ}\bigg|_{\mathcal{V}(h)=1}. \tag{186}$$

The functions appearing in the fermionic part of the Lagrangian can also be expressed in terms of the C_{IJK}-tensor. Since the C_{IJK}-tensor can be obtained by inspecting three-point amplitudes, $\mathcal{N}=2$ Maxwell–Einstein theories in five dimensions are uniquely specified by their three-point interactions. This is in contrast to Maxwell–Einstein theories that only exist in four dimensions as well as theories with hypermultiplets. It is in principle possible to compute amplitudes from the Lagrangian (182) using Feynman rules. To this end one should first expand around a scalar-base point (i.e. some background values for the scalar fields) at which the scalar and vector kinetic terms are positive-definite. The quadratic terms should then be diagonalized in order to find the spectrum, the propagators, and the vertices. For practical calculations, it is often convenient to reduce the theory to four dimensions and use the spinor-helicity formalism.

To identify the simplest supergravity theories we will utilize symmetry considerations together with minimal information on the trilinear interaction terms. A natural starting point is to consider a double copy of the form

$$\mathcal{N}=2 \text{ supergravity}: \quad (\mathcal{N}=2 \text{ SYM}) \otimes (\mathcal{N}=0 \text{ YM}),$$

in which the non-supersymmetric theory is a pure $(4+n)$-dimensional YM theory reduced to four dimensions. Bosonic asymptotic states from the double copy are identified with those from the supergravity Lagrangian as follows [120][37]:

$$\begin{aligned}
A_-^{-1} &= \bar\phi\otimes A_-, & h_- &= A_-\otimes A_-, & A_+^{-1} &= \phi\otimes A_+, & h_+ &= A_+\otimes A_+,\\
A_-^{0} &= \phi\otimes A_-, & i\bar z^0 &= A_+\otimes A_-, & A_+^{0} &= \bar\phi\otimes A_+, & -iz^0 &= A_-\otimes A_+,\\
A_-^{A} &= A_-\otimes\phi^A, & i\bar z^A &= \bar\phi\otimes\phi^A, & A_+^{A} &= A_+\otimes\phi^A, & -iz^A &= \phi\otimes\phi^A.
\end{aligned} \tag{187}$$

The index A above has range $A=1,2,\ldots,n$. Note that, in four dimensions, an extra vector field (A_μ^{-1}) is present. ϕ denotes the single complex scalar in the $\mathcal{N}=2$ SYM theory. Since all gauge-theory fields are in the adjoint representations, each field bilinear is associated to a supergravity state. Overall, the construction produces a supergravity with the following properties:

(i) has $\mathcal{N}=2$ supersymmetry in four dimensions;
(ii) has $(n+1)$ vector multiplets in four dimensions. Scalars obtained as $\phi\otimes\phi^A$ transform under a $U(1)\times SO(n)$ symmetry;
(iii) uplifts to five dimensions whenever $n>0$;
(iv) has vanishing single-soft limits at tree level (this can be checked explicitly);
(v) the supergravity can be seen as a truncation of $\mathcal{N}=4$ supergravity with the same number of vector multiplets.

[37] The phase in the map between the asymptotic states from the supergravity Lagrangian and the ones from the double copy was chosen to match the phase conventions in the supergravity literature, see e.g. [121, 329, 332]. Note that z^0 is the same scalar as τ from the previous subsection.

Putting together all available information, the scalar manifold of the resulting theory turns out to be

$$\mathcal{M}_{4D} = \frac{SO(n,2)}{SO(n) \times SO(2)} \times \frac{SU(1,1)}{U(1)}. \tag{188}$$

This infinite family of theories is known in the supergravity literature as the generic Jordan family of $\mathcal{N} = 2$ Maxwell-Einstein supergravities. The corresponding cubic polynomial in the natural basis is [329][38]:

$$\mathcal{V}(\xi) = \sqrt{2}\left(\xi^0 \left(\xi^1\right)^2 - \xi^0 \left(\xi^i\right)^2\right), \qquad i = 2,3,\ldots,n. \tag{189}$$

Exercise 5.4. Calculate explicitly $\overset{\circ}{a}_{IJ}$ and g_{xy} corresponding to the cubic polynomial above.

Exercise 5.5. Calculate the three point amplitude $\mathcal{M}_3^{\text{tree}}\left(1A_-^0, 2A_-^A, 3\bar{z}^B\right)$ using the double-copy prescription and the map (187).

5.3.3. Homogeneous $\mathcal{N} = 2$ Maxwell–Einstein supergravities. We now want to consider more general theories with $\mathcal{N} = 2$ supersymmetry and homogeneous scalar manifolds. A scalar manifold is said to be homogeneous if it admits a transitive group of isometries. From an amplitude perspective, not all these isometries will linearly realized, i.e. some of them correspond go constant shifts of the scalars which modify the vacuum of the theory. Hence, in the homogeneous case, all coordinates of the manifold are Goldstone bosons and, consequently all single-soft limits of scalar amplitudes vanish (see for example [222]). In short, drawing from the discussion in section 4.5, we have the following criterion:

> A necessary condition for a theory to possess a (locally) homogeneous scalar manifold is that all single-soft limits of scalar amplitudes vanish.

We also note that double-soft limits can be used to identify the particular homogeneous space under consideration (i.e. G in G/H) [222]. More generally, each independent vanishing single-soft scalar limit will correspond to an isometry of the scalar manifold.

We now return to the double-copy construction outlined in section 5.3.2. A natural extension consists of adding some matter fields in both gauge theories. Hypermultiplets are the only available matter that can be coupled to $\mathcal{N} = 2$ SYM theory. One hypermultiplet consists of four real scalars and two Majorana fermions and is an irreducible representation of the $\mathcal{N} = 2$ supersymmetry algebra. For the construction described below, we will need to assign matter representations of the gauge group to hypermultiplets. If the gauge-group representation is pseudo-real, an additional option becomes available: we may consider a half-hypermultiplet instead of a full one. A single half-hypermultiplet is by itself a representation of the supersymmetry algebra, but one is forced to include the Charge-Parity-Time reversal (CPT)-conjugate states unless its gauge-group representation is pseudo-real. This leads to a full hypermultiplet. Taking a single half-hypermultiplet, i.e. choosing the smallest representations of supersymmetry algebra, amounts to introducing the minimal possible number of states and, in principle, allows us to manifest a larger global symmetry in the non-supersymmetric theory.

If the desired supergravity theory is of the Maxwell–Einstein class, the non-supersymmetric gauge theory needs to be a YM-scalar theory with extra fermions, so that additional vector multiplets are obtained as double copies involving one gauge-theory hypermultiplet and one

[38] The detailed form of the C-tensor may be changed by field redefinitions without changing the scattering amplitudes. See [329] for a discussion of the *canonical* and *natural* basis.

fermion. Because of Working Rule 4, we will take the additional fermions to transform in the same pseudo-real representation $\mathcal{R}$ used for the supersymmetric theory. The Lagrangian is then written as

$$\mathcal{L} = -\frac{1}{4}F^{\hat{a}}_{\mu\nu}F^{\hat{a}\mu\nu} + \frac{1}{2}\left(D_\mu\phi^a\right)^{\hat{a}}\left(D^\mu\phi^a\right)^{\hat{a}} + \frac{i}{2}\overline{\lambda}^{\alpha}D_\mu\gamma^\mu\lambda_\alpha$$
$$+\frac{g}{2}\phi^{a\hat{a}}\Gamma^{a\ \beta}_{\alpha}\overline{\lambda}^{\alpha}\gamma_5 t^{\hat{a}}_{\mathcal{R}}\lambda_\beta - \frac{g^2}{4}f^{\hat{a}\hat{b}\hat{e}}f^{\hat{c}\hat{d}\hat{a}}\phi^{a\hat{a}}\phi^{b\hat{b}}\phi^{a\hat{c}}\phi^{b\hat{d}}, \tag{190}$$

where $\hat{a},\hat{b}$ are adjoint indices of the gauge group and $\alpha,\beta = 1,\ldots,n_F$ and $a,b = 1,\ldots,(D-4)$ are global indices. The matrices $\Gamma^{a\ \beta}_{\alpha}$ in the global indices need to be constrained by imposing the duality between color and kinematics at four points.

Imposing the duality on amplitudes between two adjoint scalars and two matter fermions gives the constraint [121]

$$\left\{\Gamma^a,\Gamma^b\right\} = -2\delta^{ab}, \tag{191}$$

that is, the matrices Γ^a are gamma matrices which belong to a $(D-4)$-dimensional Euclidean Clifford algebra. Because of this relation, the non-supersymmetric theory can be regarded as the dimensional reduction of a YM theory coupled to fermions in D dimensions. A second parameter, P, will count the number of irreducible fermions in D dimensions.

Exercise 5.6. Show that imposing CK duality on amplitudes the two-scalars two-fermion amplitudes given by the Lagrangian (190) yields the relation (191).

An important difference with the standard treatment of D-dimensional spinors is the fact that fermions transform in pseudo-real representations of the gauge group. To obtain irreducible spinors, a case-by-case analysis is necessary. Depending on the value of the parameter D, one can impose reality (R) or pseudo-reality (PR) conditions [121]

$$\overline{\lambda} = \lambda^t\mathcal{C}_4 CV, \qquad \mathrm{R}:\ C = \mathcal{C}_{D-4}, \qquad \mathrm{PR}:\ C = \mathcal{C}_{D-4}\Omega\,, \tag{192}$$

where $\mathcal{C}_{D-4}$ and $\mathcal{C}_4$ are the internal and spacetime charge-conjugation matrices, respectively. They obey the relations $\mathcal{C}_{D-4}\Gamma^a\mathcal{C}^{-1}_{D-4} = -\zeta(\Gamma^a)^t$, $\mathcal{C}_4\gamma^\mu\mathcal{C}^{-1}_4 = -\zeta(\gamma^\mu)^t$, $\zeta = \pm 1$. V is the unitary antisymmetric matrix entering the pseudo-reality condition for the gauge-group representation matrices, $Vt^{\hat{a}}_{\mathcal{R}}V^\dagger = -(t^{\hat{a}}_{\mathcal{R}})^*$. Ω is an antisymmetric real matrix acting on indices which run over the number P of irreducible spinors. Alternatively, if D is even, one can impose Weyl conditions. If we have more than one irreducible spinor, an extra flavor symmetry is present (either $U(P)$, $SO(P)$ or $USp(P)$, depending on whether Weyl, Reality or pseudo-Reality conditions were employed). A separate treatment is needed when $D = 6, 10$ (mod 8). In these dimensions, there are two inequivalent irreducible spinors with different chirality and one needs to introduce parameters $P, \dot{P}$ which count the number of each.

Explicit computations reveal that soft-scalar limits vanish for amplitudes constructed by the double copy [121]. Hence, this generalized construction yields supergravities with homogeneous scalar manifolds. The dimension-by-dimension analysis is given in table 7. The number of vector multiplets in the four-dimensional supergravity is equal to $(D-3+n_F)$, where $n_F(D,P,\dot{P})$ is the number of 4D fermions in the non-supersymmetric gauge theory; the supergravity bosonic states are obtained as double copies in the following way [121]:

Table 7. Parameters in the double-copy construction for homogeneous supergravities [121, 240]. $n_F(D,P,\dot{P})$ is the number of $4D$ irreducible spinors in the non-supersymmetric gauge theory, which can obey a reality (R), pseudo-reality (PR) or Weyl (W) conditions. Note that the pattern repeats itself with periodicity 8.

D	$\mathcal{D}_D$	$4D$ fermions $n_F(D,P,\dot{P})$	Conditions	Flavor group
4	1	P	R or W	$SU(P)$
5	1	P	R	$SO(P)$
6	1	$P+\dot{P}$	RW	$SO(P)\times SO(\dot{P})$
7	2	$2P$	R	$SO(P)$
8	4	$4P$	R or W	$U(P)$
9	8	$8P$	PR	$USp(2P)$
10	8	$8P+8\dot{P}$	PRW	$USp(2P)\times USp(2\dot{P})$
11	16	$16P$	PR	$USp(2P)$
12	16	$16P$	R or W	$U(P)$
$k+8$	$16\mathcal{D}_k$	$16\,r(k,P,\dot{P})$	as for k	as for k

$$\begin{aligned}
A_-^{-1} &= \bar{\phi}\otimes A_-\,, & h_- &= A_-\otimes A_-\,, & A_+^{-1} &= \phi\otimes A_+\,, & h_+ &= A_+\otimes A_+\,,\\
A_-^{0} &= \phi\otimes A_-\,, & i\bar{z}^0 &= A_+\otimes A_-\,, & A_+^{0} &= \bar{\phi}\otimes A_+\,, & -iz^0 &= A_-\otimes A_+\,,\\
A_-^{A} &= A_-\otimes \phi^A\,, & i\bar{z}^A &= \bar{\phi}\otimes \phi^A\,, & A_+^{A} &= A_+\otimes \phi^A\,, & -iz^A &= \phi\otimes \phi^A\,,\\
A_{\alpha-} &= \chi_-\otimes \lambda_{\alpha-}\,, & i\bar{z}_\alpha &= \chi_+\otimes \lambda_{\alpha-}\,, & A_+^{\alpha} &= \chi_+\otimes \lambda_+^\alpha\,, & -iz^\alpha &= \chi_+\otimes \lambda_-^\alpha\,.
\end{aligned} \tag{193}$$

Exercise 5.7. Show that the amplitude $\mathcal{M}_5^{\text{tree}}\left(z^0,\bar{z}^0,z^\alpha,\bar{z}_\beta,z^0\right)$ has vanishing single-soft limits for all external scalars.

Exercise 5.8. Show that the amplitude $\mathcal{M}_3^{\text{tree}}\left(1A_-^a,2A_-^\alpha,3\bar{z}^\beta\right)$ can be expressed as

$$\mathcal{M}_3^{\text{tree}}\left(1A_-^a,2A_-^\alpha,3\bar{z}^\beta\right) = \frac{\kappa}{2\sqrt{2}}\langle 12\rangle^2\left(U^t\Gamma^a C^{-1}\right)^{\alpha\beta}. \tag{194}$$

A remarkable result is that the theories listed in table 7 reproduce the complete classification of homogeneous supergravities by de Wit and van Proeyen [282]. Theories obtained with this construction include some classic examples. Specifically, for $P=1$ and $D=7,8,10,14$, we find the so-called Magical Supergravities. These theories exhibit additional symmetry enhancement, which results in the corresponding scalar manifolds being symmetric spaces. Their scalar manifolds are:

$$\begin{aligned}
&\boldsymbol{\mathcal{M}}_{4D}^{\mathbb{R}} = \frac{Sp\left(6,\mathbb{R}\right)}{U\left(3\right)}\,,\quad \boldsymbol{\mathcal{M}}_{4D}^{\mathbb{C}} = \frac{SU\left(3,3\right)}{S\left(U\left(3\right)\times U\left(3\right)\right)}\,,\quad \boldsymbol{\mathcal{M}}_{4D}^{\mathbb{H}} = \frac{SO^*\left(12\right)}{U\left(6\right)}\,,\\
&\boldsymbol{\mathcal{M}}_{4D}^{\mathbb{O}} = \frac{E_{7(-25)}}{E_6\times U\left(1\right)}\,.
\end{aligned} \tag{195}$$

An important property is that Magical theories are unified, that is there exists a symmetry with respect to which all vector fields transform in a single irreducible representation. Physically, this implies that fields from different matter multiplets have the same properties. In contrast, vectors in generic homogeneous theories typically have different interactions according to whether they are obtained as vector-scalar or as fermion-fermion from a double-copy perspective. The construction of these theories from a supergravity perspective relies on degree-three

Jordan algebras which have as elements 3×3 matrices with entries in the four division algebras ($\mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{O}$). Reviewing the supergravity construction is beyond the scope of this review; we refer the reader to [329] for details.

Aside from the magical supergravities, there is another class of examples of unified theory in four dimensions. They are obtained by choosing $D = 4$ and P is arbitrary. It is not difficult to see that the construction exhibits a global $U(P)$ flavor symmetry, as well as that the resulting supergravity theory will have $(P+1)$ complex scalars in its spectrum. Putting together this information results in the scalar manifold [121, 240]

$$\boldsymbol{\mathcal{M}}_{4D} = \frac{U(P+1,1)}{U(P+1) \times U(1)}, \tag{196}$$

which is the complex projective space $\mathbb{CP}^{P+1}$. Theories in this family are also referred to as minimally-coupled or the Luciani model. The analysis can be repeated in dimensions different from four. In five and six dimensions we find exactly one infinite-dimensional family of unified theories:

$$5D: \qquad \text{Generic non-Jordan family} \quad \boldsymbol{\mathcal{M}}_{5D} = \frac{SO(P+1,1)}{SO(P+1)}, \tag{197}$$

$$6D: \qquad \text{Generic Jordan family} \quad \boldsymbol{\mathcal{M}}_{6D} = \frac{SO(P+1,1)}{SO(P+1)}. \tag{198}$$

In both cases, the double-copy construction is similar to the one in four dimensions: the non-supersymmetric gauge theory is a YM theory in the appropriate dimension with an arbitrary number of fermions and no additional scalars. While the parameter P is by construction non-negative, it should be noted that pure supergravities in dimensions $4, 5, 6$ can be obtained as particular cases by setting $P = -1$. This observation will be consequential in formulating double-copy constructions for pure supergravities with $\mathcal{N} = 2$ in various dimensions. The construction outlined in this section has been used to compute one-loop matter amplitudes in these theories and to analyze their UV properties at that order, see [294].

5.3.4. Pure supergravities. Pure supergravities with $\mathcal{N} = 1, 2, 3$ have been originally formulated from a Lagrangian perspective in [287, 333–335] and [336, 337], respectively. Regardless of the number of supercharges which are manifest in the construction, a double-copy gravity theory in four dimensions contains a complex scalar which is obtained from the product of gluons of opposite polarizations. For extended $\mathcal{N} \geqslant 4$ supersymmetry, this scalar still belongs to the gravity multiplet. For $\mathcal{N} < 4$, however, the gravity multiplet does not contain any scalar field so the complex scalar under consideration belongs to a matter multiplet. Hence, to obtain pure supergravities with $\mathcal{N} < 4$, one needs to modify the construction and remove the contributions to amplitudes of the unwanted scalar. At tree level, one can always project out the unwanted scalars from the amplitudes by judiciously choosing the asymptotic states. Special care is however necessary for loops.

A solution to the problem was first outlined in [188]. The first step is to introduce an additional matter representation (the fundamental representation, without any loss of generality) in both gauge theories. The precise map depends on the desired amount of supersymmetry and is listed in table 8. Since the various graphs contributing to the amplitude carry representation information associated to all internal and external lines, we can organize the graphs with no external matter according to the number of matter loops. We can then treat the additional matter

Table 8. Pure gravities constructed as double copies [188]. The construction necessitates ghosts from matter-antimatter double copies. Barred multiplets transform in the anti-fundamental representation. For compactness, graviton and vector supermultiplets $\mathcal{H}, \mathcal{V}_{\mathcal{N}<4}$ include the CPT-conjugate states. Pairs of chiral/antichiral $\mathcal{N}=1$ supermultiplets are grouped in $\mathcal{N}=2$ hypermultiplets, denoted as $\Phi_{\mathcal{N}=2}$.

$\mathcal{N}$	tensoring vector states	ghosts = matter $\otimes$ $\overline{\text{matter}}$
$0+0$	$A_\mu \otimes A_\nu = h_{\mu\nu} \oplus \phi \oplus a$	$(\psi_+ \otimes \psi_-) \oplus (\psi_- \otimes \psi_+) = \phi \oplus a$
$1+0$	$\mathcal{V}_{\mathcal{N}=1} \otimes A_\mu = \mathcal{H}_{\mathcal{N}=1} \oplus \Phi_{\mathcal{N}=2}$	$(\Phi_{\mathcal{N}=1} \otimes \psi_-) \oplus (\bar{\Phi}_{\mathcal{N}=1} \otimes \psi_+) = \Phi_{\mathcal{N}=2}$
$2+0$	$\mathcal{V}_{\mathcal{N}=2} \otimes A_\mu = \mathcal{H}_{\mathcal{N}=2} \oplus \mathcal{V}_{\mathcal{N}=2}$	$(\Phi_{\mathcal{N}=2} \otimes \psi_-) \oplus (\bar{\Phi}_{\mathcal{N}=1} \otimes \psi_+) = \mathcal{V}_{\mathcal{N}=2}$
$1+1$	$\mathcal{V}_{\mathcal{N}=1} \otimes \mathcal{V}_{\mathcal{N}=1} = \mathcal{H}_{\mathcal{N}=2} \oplus 2\Phi_{\mathcal{N}=2}$	$(\Phi_{\mathcal{N}=1} \otimes \bar{\Phi}_{\mathcal{N}=1}) \oplus (\bar{\Phi}_{\mathcal{N}=1} \otimes \Phi_{\mathcal{N}=1}) = 2\Phi_{\mathcal{N}=2}$
$2+1$	$\mathcal{V}_{\mathcal{N}=2} \otimes \mathcal{V}_{\mathcal{N}=1} = \mathcal{H}_{\mathcal{N}=3} \oplus \mathcal{V}_{\mathcal{N}=4}$	$(\Phi_{\mathcal{N}=2} \otimes \bar{\Phi}_{\mathcal{N}=1}) \oplus (\bar{\Phi}_{\mathcal{N}=2} \otimes \Phi_{\mathcal{N}=1}) = \mathcal{V}_{\mathcal{N}=4}$
$2+2$	$\mathcal{V}_{\mathcal{N}=2} \otimes \mathcal{V}_{\mathcal{N}=2} = \mathcal{H}_{\mathcal{N}=4} \oplus 2\mathcal{V}_{\mathcal{N}=4}$	$(\Phi_{\mathcal{N}=2} \otimes \bar{\Phi}_{\mathcal{N}=2}) \oplus (\bar{\Phi}_{\mathcal{N}=2} \otimes \Phi_{\mathcal{N}=2}) = 2\mathcal{V}_{\mathcal{N}=4}$

as a ghost multiplet by associating an extra minus sign to each matter loop (as with Faddeev–Popov ghosts). More explicitly, loop-level pure-supergravity amplitudes are constructed using the prescription

$$\mathcal{M}_m^{(L)} = i^{L-1}\left(\frac{\kappa}{2}\right)^{m+2L-2} \sum_{i\in\text{cubic}} \int \frac{d^{LD}\ell}{(2\pi)^{LD}} \frac{(-1)^{|i|}}{S_i} \frac{n_i \tilde{n}_i}{D_i}, \tag{199}$$

where $|i|$ denotes the number of matter loops in the ith graph. It has been shown by explicit calculation through two loops and argued to all loop orders that amplitudes obtained with this prescription have the same unitarity cuts as the ones of the pure supergravities listed in table 8. It is interesting to note that $\mathcal{N}=2$ ghost multiplets are constructed as double copies involving fermions in the non-supersymmetric theory. One can in principle consider an analogous construction involving scalars, but amplitudes constructed in this way would have unitarity cuts which are different from those of pure supergravities. This observation provides a clue on the meaning of the construction: formally, the prescription above is equivalent to considering one of the unified infinite families of supergravities described at the end of the last subsection and setting $P=-1$.

Explicit calculations show that pure Einstein gravity is UV-divergent at two loops[39] [340, 341], although it is finite at one loop because the candidate counterterm is a total derivative in four dimensions [342]. For $\mathcal{N}=1,2,3$ pure supergravities, UV divergences cannot appear before three loops [343, 344]; the relevant explicit calculations at this loop order, probing the appearance of divergences, have as yet not been carried out.

5.3.5. Theories with hypermultiplets and supergravities with $\mathcal{N}<2$. An alternative option is to couple matter hypermultiplets with $\mathcal{N}=2$ supergravity. In general, supergravities with hypermultiplets are less constrained than theories with vector multiplets. A subset of such theories, however, appears to be closely related to theories with vector multiplets through a procedure known as c-map [345].

Starting from a Maxwell–Einstein theory in four dimensions, one first reduces the theory to three dimensions. After dualization of the vector field, each supermultiplet in the three-dimensional theory contains four real scalars and four Majorana fermions, which is the field

[39] The interpretation of the divergence is rather subtle because of its dependence on evanescent operators and choice of fields [338, 339].

content corresponding to a hypermultiplet. Since the hypermultiplet action is the same in any dimension up to six, the three-dimensional theory obtained with this procedure can be uplifted to higher dimension, leading to the image of the original Maxwell–Einstein theory under the c-map.

The basic double-copy construction for theories with hypermultiplets was mentioned in [121] and further detailed in [240]. It relies on taking as one copy $\mathcal{N}=2$ SYM theory with matter hypermultiplets and, as the second copy, a YM theory with extra matter scalars. The simplest realization of this construction involves a SYM theory with a single hypermultiplet in a real representation and a YM theory with m real scalars. A Lagrangian for the latter theory is[40]:

$$\mathcal{L} = -\frac{1}{4}F^{\hat{a}}_{\mu\nu}F^{\hat{a}\mu\nu} + \frac{1}{2}D_\mu\varphi^I D^\mu\varphi^I + \frac{g^2}{4}\left(\varphi^{[I}t^{\hat{a}}_{\mathcal{R}}\varphi^{J]}\right)\left(\varphi^{[I}t^{\hat{a}}_{\mathcal{R}}\varphi^{J]}\right). \tag{200}$$

Scalar fields φ^I are labeled by flavor indices $I,J=1,\dots,m$, which refer to the global $SO(m)$ symmetry, and gauge-group representation indices for some real representation $\mathcal{R}$, which we do not display; $t^{\hat{a}}_{\mathcal{R}}$ are the gauge-group generators in this representation and $\hat{a},\hat{b}$ are adjoint indices. One can verify that this theory obeys CK duality at four points. The computation is identical to the one for a higher-dimensional YM theory reduced to four dimensions, with the only difference being related to representation of the scalar fields.

Based on the symmetry $SO(m)\times SO(4)$ which is manifest in the construction, the $4m$ real hypermultiplet scalars in the theory together with the universal dilaton-axion parametrize the scalar manifold

$$\boldsymbol{\mathcal{M}}_{4D} = \frac{SU(1,1)}{U(1)}\times\frac{SO(m,4)}{SO(m)\times SO(4)}. \tag{201}$$

The second term in the product manifold is the special quaternionic-Kähler manifold which is the image of the generic Jordan family scalar manifold under the c-map. From the point of view of scattering amplitudes, the relation between the two classes of theories is a consequence of the fact that the kinematic numerator factors from the non-supersymmetric gauge theory are identical in the two constructions. The differences relate to the pairing between the kinematic numerators of the two gauge theories, which is now different because of the different gauge-group representations and color factors.

Several additional constructions for ungauged supergravities with various matter contents deserve mention:

- Various $(\mathcal{N}=1)\times(\mathcal{N}=1)$ double copies were studied in [239, 241, 295, 346]. In this case, at least one hypermultiplet is present. Since $\mathcal{N}=1$ gauge theories do not generically uplift to higher dimensions, the construction does not manifestly give a supergravity which can be written in five dimensions and, hence, three-point amplitudes cannot be used to specify the theory completely. Instead, the identification relies on symmetry consideration and on the possibility of embedding the theory into a supergravity with extended supersymmetry.
- Several examples of $\mathcal{N}=1$ supergravities constructed as double copies are known. The known examples can often be seen as truncations of theories with a larger number of supersymmetries [239, 241, 295, 346].

[40] In principle, it is possible to choose a different coefficient for the quartic scalar coupling while preserving CK duality. Indeed the scalar sector of this theory is the same as the theory discussed at the end of section 5.2. The theory given here can also be constructed as a field-theory orbifold of an adjoint YM theory in higher dimension. See also [122] for a similar discussion.

- Various examples of non-supersymmetric gravities constructed as double copies are known [1, 188]. Among these, the simplest example is Einstein gravity with a scalar and antisymmetric tensor, which we have already encountered in section 2. This theory is constructed as the square of YM theory. In four-dimensions, the antisymmetric tensor is dual to an axion. The scalar-sector Lagrangian is then identical to the one in (179).
- An interesting version of the construction applies to the so-called twin supergravities [347]. These are pairs of supergravities with different amounts of supersymmetries which share the same bosonic Lagrangian but have different fermionic field content and interactions.

Exercise 5.9. Consider two supergravities constructed as field theory orbifolds of $\mathcal{N}=8$ supergravity with the following generators:

$$\begin{aligned}&\text{Theory 1}: R=\operatorname{diag}\left(1,e^{\frac{2\pi i}{3}},e^{\frac{2\pi i}{3}},e^{\frac{2\pi i}{3}},1,1,1,1\right), \quad R'=\operatorname{diag}(1,1,1,1,-1,-1,-1,-1)\,,\\&\text{Theory 2}: R=\operatorname{diag}(1,i,i,i,-1,-1,-1,-i)\,.\end{aligned}$$

Find the corresponding spectra and, using the manifest symmetries of the construction, find a candidate for the scalar manifolds.

5.3.6. YME theories. YME theories are supergravities that involve nonabelian gauge interactions among (some of) the vector fields. Surprisingly, they admit a very simple double-copy realization, which relies on the following principle [120]:

> To introduce nonabelian gauge interactions in a gravitational theory from the double copy, it is sufficient to add trilinear couplings among adjoint scalar fields in one of the gauge theories entering the construction.

The relevant Lagrangian was introduced in (150). The effect of the trilinear coupling is to introduce nonzero supergravity amplitudes of the form

$$\begin{aligned}\mathcal{M}_3\left(1A_-^A,2A_-^B,3A_+^C\right)&=iA_3\left(1A_-,2A_-,3A_+\right)A_3\left(1\phi^A,2\phi^B,3\phi^C\right)\\&=-\frac{\kappa}{2\sqrt{2}}\lambda F^{ABC}\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 31\rangle}=i\frac{\kappa}{4}\lambda\tilde{F}^{ABC}\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 31\rangle}\,,\end{aligned}\tag{202}$$

i.e. amplitudes between three spin-1 fields which are proportional to an antisymmetric tensor obeying Jacobi relations and have the same momentum dependence as the three-gluon amplitudes from the supergravity Lagrangian. In particular, the supergravity gauge coupling constant g_s is related to the parameter λ in (150) as

$$g_s=\left(\frac{\kappa}{4}\right)\lambda\,,\tag{203}$$

where we have temporarily re-introduced κ. In this construction, the global-symmetry tensor F^{ABC}, which obeys the Jacobi identity (153), is identified with the structure constants of the supergravity gauge group. Hence, a global symmetry in one of the two gauge theories becomes a local symmetry in the resulting double-copy gravity theory.

We note that this approach gives, by construction, gauge groups which are subgroups of the manifest isometry group of the corresponding Maxwell-Einstein theory. These groups are necessarily compact. Gauging a subgroup of the R symmetry, a construction which results in the so-called gauged supergravities, requires a more involved procedure which will be outlined in section 5.3.8. The double-copy construction for YME theories can be adapted to supergravities with various amounts of supersymmetry, which are listed in table 9 [120].

Table 9. Amplitudes in YME gravity theories for different number of supersymmetries, corresponding to different choices for the left gauge-theory factor entering the double copy [120]. YM_{DR} denotes the YM-scalar theory obtained from dimensional reduction.

Gravity coupled to YM	Gauge theory 1	Gauge theory 2
$\mathcal{N}=4$ YME supergravity	$\mathcal{N}=4$ SYM	$\text{YM}+\phi^3$
$\mathcal{N}=2$ YME supergravity (gen.Jordan)	$\mathcal{N}=2$ SYM	$\text{YM}+\phi^3$
$\mathcal{N}=1$ YME supergravity	$\mathcal{N}=1$ SYM	$\text{YM}+\phi^3$
$\mathcal{N}=0$ YME + dilaton + $B^{\mu\nu}$	YM	$\text{YM}+\phi^3$
$\mathcal{N}=0$ YM_{DR}-E + dilaton + $B^{\mu\nu}$	YM_{DR}	$\text{YM}+\phi^3$

Aside from spelling out the construction at the level of the gauge-theory Lagrangian, it is interesting to consider the implications of the double-copy structure on YME amplitudes. We start from the double copy

$$\left(\text{YM}+\phi^3\right) = \left(\text{YM}+\phi^3\right) \otimes \left(\phi^3 \text{ theory}\right), \tag{204}$$

i.e. we note that the double copy between the YM+ϕ^3 theory and the bi-adjoint ϕ^3 theory gives amplitudes from the YM+ϕ^3 theory itself. By choosing numerator factors corresponding to the DDM basis [180], we then write a color-ordered tree amplitude between k gluons and $m \geqslant 2$ scalars in the YM+ϕ^3 theory as follows

$$\mathcal{A}_{k,m}^{\text{YM+}\phi^3}(1,\ldots,k|k+1,\ldots,k+m) = -i \sum_{w\in\sigma_{12\ldots k}} N_k(w) A_{k+m}^{\phi^3}(w) \ + \ \text{Perm}(1,\ldots,k)\,. \tag{205}$$

$A_{k+m}^{\phi^3}(w)$ are amplitudes in bi-adjoint ϕ^3 theory that are color-ordered only with respect to one of the two colors. In the above formula we are summing over all color orderings w that belong to the set $\sigma_{123\cdots k}$; which is explicitly constructed using a shuffle product $\sqcup\!\sqcup$, as

$$\begin{aligned}\sigma_{123\cdots k} = &\left\{\{k+1,\gamma,k+m\}\,\middle|\,\gamma\in\alpha\sqcup\!\sqcup\beta\right\}, \quad \text{where}\\ &\alpha = \{1,2,3,\ldots,k\} \quad \text{and} \quad \beta = \{k+2,\ldots,k+m-1\}\,.\end{aligned} \tag{206}$$

The set $\sigma_{123\cdots k}$ contains all shuffles of the gluon (α) and scalar (β) sets that respect the ordering within each set, with the additional constraint that the first and last scalars are held fixed. We will refer to the elements of this set as 'words' w. A remarkable observation is that gauge invariance is sufficient to fix the numerators $N_k(w)$ in the expression above.

Color-ordered single-trace YME amplitudes are obtained by replacing the ϕ^3 partial amplitudes with partial amplitudes belonging to pure YM theory (or its supersymmetric relatives, depending on the target gravitational theory) [156],

$$M_{k,m}^{\text{YME(SG)}}(1,\ldots,k|k+1,\ldots,k+m) = \sum_{w\in\sigma_{12\ldots k}} N_k(w) A_{k+m}^{\text{(S)YM}}(w) + \text{Perm}(1,\ldots,k)\,. \tag{207}$$

Since the partial amplitudes $A_{k+m}^{\text{(S)YM}}(w)$ obey the same relations as $A_{k+m}^{\phi^3}(w)$ (including in particular the BCJ relations), the YME amplitudes given by this formula will be by construction gauge invariant. The numerators $N_k(w)$ are obtained by imposing gauge invariance on (205), that is by imposing that the amplitude vanishes after the replacement

$$\varepsilon_i \to p_i\,. \tag{208}$$

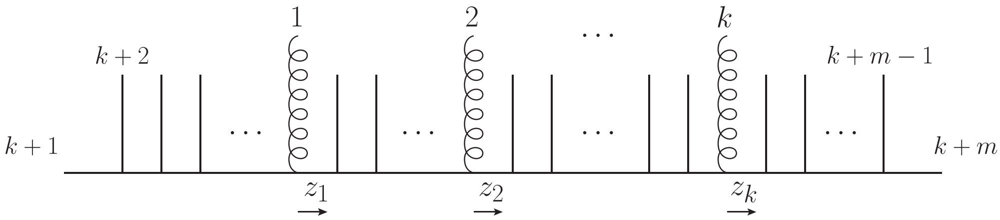


Figure 19. Half-ladder graph for the YM+ϕ^3 theory. The gluons are denoted with $1, 2, \ldots, k$ and the remaining particles are scalars. z_i denote the momentum of the internal scalar to the right of gluon i.

Along the same lines of section 3.2, we construct the numerators $N_k(w)$ from the following independent Lorentz invariants:

$$\{(\varepsilon_i \cdot z_i), \ (p_i \cdot z_i), \ (\varepsilon_i \cdot \varepsilon_j), \ (\varepsilon_i \cdot p_j), \ (p_i \cdot p_j)\}, \qquad (i,j = 1,\ldots,k) \tag{209}$$

where p_i denote only the momenta of the gluons. The momenta of the scalars will only appear implicitly through the region momenta $z_i = z_i(w)$ that we define as

$$z_i(w) = \sum_{\substack{1 \leqslant j \leqslant l \\ w_l = i}} p_{w_j}, \tag{210}$$

which give the sum of the momenta of all the particles to the left of the ith gluon in the multiperipheral graph corresponding to the word w (including the gluon momentum p_i, see figure 19).

We consider the case of one external gluon ($k = 1$). By dimensional analysis, each term in $N(w)$ will need to contain a single factor of momentum. σ_1 is the set of external-leg orderings in which the order of the scalars is preserved and the single external gluon is inserted in different positions (leaving a scalar as the first and last entry). A natural guess for the numerator is given by

$$N_1 = 2(\varepsilon_1 \cdot z_1). \tag{211}$$

We can check gauge invariance with the replacement (208); the gauge variation of the amplitude becomes

$$\sum_{w \in \sigma_1} (p_1 \cdot z_1) A^{\text{(S)YM}}_{m+1}(w) = 0, \tag{212}$$

which is zero as a consequence of the BCJ relations. Indeed, (212) is precisely the fundamental BCJ relation (60) [1, 113]. That this BCJ relation can be obtained from YME amplitudes with a single graviton, using the numerator (211), was first shown in [133].

The next-simplest example has two external gluons. Now σ_{12} will be the set of external-leg orderings in which the two gluons are inserted in different positions while leaving the order of scalars and gluons unchanged (and keeping scalars as the first and last entries). The last term in (207) is obtained by exchanging the two external gluons. The numerators are

$$N_2 = 4(\varepsilon_1 \cdot z_1)(\varepsilon_2 \cdot z_2) + 2(p_2 \cdot z_2)(\varepsilon_1 \cdot \varepsilon_2), \tag{213}$$

where the last (contact) term is fixed by imposing gauge invariance. Taking a gauge variation $\varepsilon_1 \to p_1$ of the amplitude we obtain:

$$4\left(\sum_{w\in\sigma_{12}}+\sum_{w\in\sigma_{21}}\right)(p_1\cdot z_1)(\varepsilon_2\cdot z_2)A^{\text{(S)YM}}_{m+2}(w)$$
$$+2(\varepsilon_2\cdot p_1)\left(\sum_{w\in\sigma_{12}}(p_2\cdot z_2)A^{\text{(S)YM}}_{m+2}(w)+\sum_{w\in\sigma_{21}}(p_1\cdot z_1)A^{\text{(S)YM}}_{m+2}(w)\right). \tag{214}$$

Using the fundamental BCJ relation (60), the reader can verify that the above gauge variation reduces to

$$(\varepsilon_2\cdot p_1)\sum_{w\in\sigma_{12}}((p_1\cdot z_1)+(p_2\cdot z_2))A^{\text{(S)YM}}_{m+2}(w)=0. \tag{215}$$

This is equivalent to the sum of two BCJ relations and thus vanishes for all amplitudes that satisfy CK duality.

Exercise 5.10. Verify (215) starting from (214) and using the fundamental BCJ relation.

Semi-recursive expressions for YME amplitudes with up to five external gravitons were given in [156]. General expressions for any multiplicity based on BCFW recursion were given in [214] in the single-trace case and in [216] in the multi-trace case. The reader may also consult [212, 289] for alternative expressions obtained through the CHY formalism and [307, 348] for loop-level amplitudes in YME theory.

5.3.7. Higgsed supergravities. A key feature of the double-copy construction for YME theories is that it can be generalized to cases in which the nonabelian gauge supersymmetry of the supergravity theory is spontaneously broken. Since YME theories possess a moduli space in which the unbroken-gauge phase is given by a single isolated point, the fact that the double-copy construction admits an extension of this sort gives a strong hint of its applicability for generic gravity theories. At the same time, the construction we review here is one of the simplest examples in which some of the fields are massive. The double-copy construction for a Higgsed supergravity has the schematic form

$$(\text{Higgsed YME SG})=(\text{Coulomb-branch SYM theory})\otimes(\text{YM}+\text{massive scalars}). \tag{216}$$

Schematically, amplitudes for the first gauge-theory factor can be obtained with a two-step process: (1) break the gauge-group down to a subgroup (see section 5.2.1); (2) assign masses (seen as compact momenta) to fields transforming in matter representations of the unbroken subgroup (see section 5.2.3). We have seen in section 5.3.6 that, with the appropriate choices of gauge theories, the global symmetry in one of the gauge-theory factors becomes a nonabelian gauge symmetry in (super)gravity. The scenario discussed here extends this property by showing that an explicitly-broken symmetry in one of the two gauge theories becomes, through the double copy, a spontaneously-broken gauge symmetry in (super)gravity.

To avoid a notationally-heavy discussion, we will review here the simplest example of the Higgsed double-copy construction. We start from a $\mathcal{N}=2$ SYM theory with $SU(N+M)$ gauge group and decompose it with respect to the $SU(M)\times SU(N)\times U(1)$ subgroup. The direct sum of the adjoint representations of the unbroken gauge-group factors is denoted as $\mathcal{G}$; the corresponding fields are left massless. In addition, there will be two vector multiplets transforming in the bifundamental $\mathcal{R}=(\boldsymbol{M},\overline{\boldsymbol{N}})$ and anti-bifundamental $\overline{\mathcal{R}}=(\overline{\boldsymbol{M}},\boldsymbol{N})$ representations. These

Table 10. Bosonic fields in gauge-theory factors for the example of double-copy construction for Higgsed supergravities.

Fields	Representation	Mass	Fields	Representation	Mass
$(A_\mu,\phi,\bar{\phi})$	$\mathcal{G}$	0	$(A_\mu,\phi,\bar{\phi})$	$\mathcal{G}$	0
(W_μ,φ)	$\mathcal{R}$	m	φ	$\overline{\mathcal{R}}$	m
$(\overline{W}_\mu,\overline{\varphi})$	$\overline{\mathcal{R}}$	$-m$	$\overline{\varphi}$	$\mathcal{R}$	$-m$

fields are made massive by a suitable assignment of momenta along one single compact dimension. This relies, implicitly, on the fact that the theory can be uplifted to higher dimension. The resulting bosonic states are given in table 10 (while fermionic states are not displayed). As discussed in section 5.2.3, this is equivalent to giving a scalar VEV

$$\langle\phi\rangle = Vt^{\hat{0}}\,, \tag{217}$$

where $t^{\hat{0}}=\text{diag}\left(\frac{1}{M}I_M,-\frac{1}{N}I_N\right)$ and V is a real parameter (since our assignment of compact momenta only involves a single compact dimension).

At this stage, we need to examine the constraints coming from the duality between color and kinematics at four points. CK duality for amplitudes between two adjoint and two matter fields is automatically satisfied. This is a consequence of the fact that the theory can be obtained by assigning compact momenta to a higher-dimensional massless theory, as explained in section 5.2.3. Alternatively, one could adopt a bottom-up approach and start from a Lagrangian involving massive vectors and scalars and leave free parameters in the interaction terms. Imposing CK duality would fix the interaction terms to be the ones of the Coulomb-branch theory.

Amplitudes involving four matter fields require a more detailed analysis. In particular, scattering amplitude of four massive scalars can be cast in the form

$$\mathcal{A}_4\left(1\varphi_{\hat{\alpha}},2\varphi_{\hat{\beta}},3\overline{\varphi}^{\hat{\gamma}},4\overline{\varphi}^{\hat{\delta}}\right) = -i\,g^2\left(\frac{n_t c_t}{D_t}+\frac{n_u c_u}{D_u}+\frac{n_s c_s}{D_s}\right), \tag{218}$$

where the color factors are[41]

$$c_t=\tilde{f}^{\hat{a}}{}_{\hat{\alpha}}{}^{\hat{\delta}}\tilde{f}^{\hat{a}}{}_{\hat{\beta}}{}^{\hat{\gamma}}\,,\qquad c_u=\tilde{f}^{\hat{a}}{}_{\hat{\alpha}}{}^{\hat{\gamma}}\tilde{f}^{\hat{a}}{}_{\hat{\beta}}{}^{\hat{\delta}}\,,\qquad c_s=\tilde{f}^{\hat{\gamma}\hat{\delta}}{}_{\hat{\epsilon}}\tilde{f}^{\hat{\epsilon}}{}_{\hat{\alpha}\hat{\beta}}\,, \tag{219}$$

while the inverse propagators are

$$D_t=(p_1+p_4)^2\,,\qquad D_u=(p_1+p_3)^2\,,\qquad D_s=(p_1+p_2)^2-(2m)^2\,. \tag{220}$$

To understand the mass $(2m)$ in the massive channel it is useful to recall that masses have been assigned as momenta in some additional dimensions. Because of this, masses are conserved at each vertex. Since the color factor c_s contains two fields of the same complex representation of masses m meeting at a vertex, the third field must necessarily have mass $2m$. The kinematic numerators are given by:

$$\begin{aligned}&n_t=-p_1\cdot p_2+p_1\cdot p_3+2m^2\,,\qquad n_u=-2p_1\cdot p_2+m^2-p_1\cdot p_3\,,\\&n_s=p_1\cdot p_2+2p_1\cdot p_3+m^2\,.\end{aligned} \tag{221}$$

These numerators can be obtained from (12) by assigning momenta along a single compact dimension or, alternatively, from the YM-scalar Lagrangian (154) with $a=0$.

[41] As discussed in section 2, it is convenient to write the color factors in terms of $\tilde{F}^{ABC}=\sqrt{2}i\,F^{ABC}$.

Exercise 5.11. Modify the example discussed above by introducing a VEV that corresponds to compact momenta along two compact dimensions. Write the numerators for four-scalar amplitudes.

We note that the s-channel color factor is zero because there does not exist an invariant gauge-group object with two bifundamental and one anti-bifundamental indices[42]. However, the corresponding numerator factor is nonzero. Alternatively stated, the color factors obey two-term algebraic relations, while numerator factors obey three-term relations. This observation affects the choice of the second gauge-theory factor entering the double-copy construction, which must have an identically-vanishing s-channel numerator.

We choose a non-supersymmetric theory with one complex massive scalar transforming in the representation conjugate to the one of the Coulomb-branch theory (see table 10). Its Lagrangian is

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}\tilde{F}^{\hat{a}}_{\mu\nu}\tilde{F}^{\hat{a}\mu\nu} + \frac{1}{2}D_\mu\phi^{a\hat{a}}D^\mu\phi^{a\hat{a}} + D_\mu\overline{\varphi}D^\mu\varphi - m^2\overline{\varphi}\varphi - \frac{g^2}{4}f^{\hat{a}\hat{b}\hat{e}}f^{\hat{c}\hat{d}\hat{e}}\phi^{a\hat{a}}\phi^{b\hat{b}}\phi^{a\hat{c}}\phi^{b\hat{d}} \\ &-\frac{g^2}{2}\left(\overline{\varphi}t^{\hat{a}}_{\mathcal{R}}\varphi\right)\left(\overline{\varphi}t^{\hat{a}}_{\mathcal{R}}\varphi\right) + g^2\phi^{a\hat{a}}\phi^{a\hat{b}}\overline{\varphi}t^{\hat{a}}_{\mathcal{R}}t^{\hat{b}}_{\mathcal{R}}\varphi + g\lambda\phi^{2\hat{a}}\overline{\varphi}t^{\hat{a}}_{\mathcal{R}}\varphi\,,\end{aligned} \tag{222}$$

where $a, b = 1, 2$, $t_{\mathcal{R}}$ are gauge-group representation matrices for the massive scalars, and only $\phi^{2\hat{a}}$ enters the trilinear scalar couplings. The reason for this latter choice is that we want a construction which manifestly uplifts to five dimensions. Without any loss of generality we can rotate the other scalars which appear in the trilinear couplings into $\phi^{2\hat{a}}$. One can check that numerators of this theory obey a two-term relation.

Putting all together, the spectrum of the resulting supergravity theory is given by [122]:

$$\begin{aligned} &A_-^{-1} = \bar{\phi}\otimes A_-\,, \quad &&h_- = A_-\otimes A_-\,, \quad &&A_+^{-1} = \phi\otimes A_+\,, \quad &&h_+ = A_+\otimes A_+\,, \\ &A_-^{0} = \phi\otimes A_-\,, \quad &&i\bar{z}^0 = A_+\otimes A_-\,, \quad &&A_+^{0} = \bar{\phi}\otimes A_+\,, \quad &&-iz^0 = A_-\otimes A_+\,, \\ &A_-^{A} = A_-\otimes \phi^A\,, \quad &&i\bar{z}^A = \bar{\phi}\otimes \phi^A\,, \quad &&A_+^{A} = A_+\otimes \phi^A\,, \quad &&-iz^A = \phi\otimes \phi^A\,, \\ &W_i = W_i\otimes\varphi\,, \quad &&\varphi = \varphi\otimes\varphi\,. \end{aligned} \tag{223}$$

with massive fields given by $\mathcal{R}\otimes\overline{\mathcal{R}}$ bilinears (the index i runs over the massive-vector three physical polarizations). Note that this construction has two free parameters: the mass m and the constant λ in the trilinear scalar couplings. Comparison with amplitudes from the Higgsed supergravity Lagrangian leads to the identification (203)[43]. The masses of supergravity fields are the same as those of the gauge-theory fields from which they are constructed. In turn, this determines the choice of scalar base-point for the supergravity perturbative expansion that matches the result of the double copy. Given the presence of two massive W bosons in the supergravity spectrum, the supergravity gauge-symmetry breaking is $SU(2)\to U(1)$.

This is arguably the most straightforward example of Higgsed supergravity constructed as double copy. In the general case, we need to consider a generic breaking of the Coulomb-branch theory. The structure constants, generators and Clebsch–Gordan coefficients obey relations inherited from the Jacobi relations of the original gauge group. A first set of relations is

[42] In a standard formulation of the Higgs mechanism, this channel does not appear in the amplitude because the necessary vertices are absent from the Lagrangian.

[43] To obtain a Higgsed supergravity, we take $\lambda > 0$ strictly.

Table 11. New double-copy constructions corresponding to spontaneously-broken YME gravity theories for different amounts of supersymmetry [122]. The dimensionally-reduced $\mathrm{YM_{DR}}$ theory must have at least one scalar to provide the VEV responsible for spontaneous symmetry breaking.

Gravity coupled to $\cancel{\mathrm{YM}}$	Left gauge theory	Right gauge theory
$\mathcal{N}=4$ $\cancel{\mathrm{YM}}$E supergravity	$\mathcal{N}=4$ S$\cancel{\mathrm{YM}}$	$\mathrm{YM}+\cancel{\phi^3}$
$\mathcal{N}=2$ $\cancel{\mathrm{YM}}$E supergravity (gen. Jordan)	$\mathcal{N}=2$ S$\cancel{\mathrm{YM}}$	$\mathrm{YM}+\cancel{\phi^3}$
$\mathcal{N}=0$ $\cancel{\mathrm{YM}}_{\mathrm{DR}}$-E + dilaton + $B^{\mu\nu}$	$\cancel{\mathrm{YM}}_{\mathrm{DR}}$	$\mathrm{YM}+\cancel{\phi^3}$

$$
\begin{aligned}
&f^{\hat d\hat a\hat c}f^{\hat c\hat b\hat e}-f^{\hat d\hat b\hat c}f^{\hat c\hat a\hat e}=f^{\hat a\hat b\hat c}f^{\hat d\hat c\hat e},\\
&f^{\hat a}{}_{\hat\gamma}{}^{\hat\beta}f^{\hat b}{}_{\hat\alpha}{}^{\hat\gamma}-f^{\hat b}{}_{\hat\gamma}{}^{\hat\beta}f^{\hat a}{}_{\hat\alpha}{}^{\hat\gamma}=f^{\hat a\hat b\hat c}f^{\hat c}{}_{\hat\alpha}{}^{\hat\beta},\\
&f^{\hat a}{}_{\hat\epsilon}{}^{\hat\gamma}f^{\hat\epsilon}{}_{\hat\delta}{}^{\hat\beta}-f^{\hat a}{}_{\hat\epsilon}{}^{\hat\beta}f^{\hat\epsilon}{}_{\hat\delta}{}^{\hat\gamma}=f^{\hat a}{}_{\hat\delta}{}^{\hat\epsilon}f^{\hat\gamma}{}_{\hat\epsilon}{}^{\hat\beta}.
\end{aligned}
\tag{224}
$$

These relations are necessary to ensure gauge invariance. The Clebsch–Gordan coefficients $f^{\hat\gamma}{}_{\hat\epsilon}{}^{\hat\beta}$ need to obey additional identities:

$$
\begin{aligned}
f^{\hat\alpha}{}_{\hat\epsilon}{}^{\hat\gamma}f^{\hat\epsilon}{}_{\hat\delta}{}^{\hat\beta}-f^{\hat\alpha}{}_{\hat\epsilon}{}^{\hat\beta}f^{\hat\epsilon}{}_{\hat\delta}{}^{\hat\gamma}&=f^{\hat\alpha}{}_{\hat\delta}{}^{\hat\epsilon}f^{\hat\gamma}{}_{\hat\epsilon}{}^{\hat\beta},\\
\left(f^{\hat\beta}{}_{\hat\gamma}{}^{\hat\epsilon}f_{\hat\epsilon}{}^{\hat\alpha}{}_{\hat\delta}+f^{\hat\alpha}{}_{\hat\delta}{}^{\hat\epsilon}f_{\hat\epsilon}{}^{\hat\beta}{}_{\hat\gamma}+f^{\hat a}{}_{\hat\gamma}{}^{\hat\beta}f^{\hat a}{}_{\hat\delta}{}^{\hat\alpha}\right)-\left(\hat\alpha\leftrightarrow\hat\beta\right)&=f^{\hat\alpha}{}_{\hat\epsilon}{}^{\hat\beta}f_{\hat\delta}{}^{\hat\epsilon}{}_{\hat\gamma}.
\end{aligned}
\tag{225}
$$

The seven-term identity is to be thought of as a compact way of writing a set of three- and two-term identities. The general construction for Higgsed supergravities proceeds as follows [122]:

- One introduces a non-supersymmetric gauge theory with massive scalars and imposes the identities (224) and (225) on its numerator factors. Note that the numerators of the Coulomb-branch theory need not obey the same identities.
- Masses need to be matched on both gauge-theory factors. For gaugings that uplift to five dimensions, the Higgs mechanism requires that the Coulomb-branch theory masses be proportional to a preferred $U(1)$ gauge generator (given by the direction of the VEV). Imposing CK duality results in demanding than the masses in the explicitly-broken massive-scalar theory also be proportional to a preferred $U(1)$ global generator (in our example, the $U(1)$ acting as a phase rotation on the complex scalars).
- In general, the symmetry-breaking pattern (number of factors in the gauge group, number of matter representation, existence of Clebsch–Gordan coefficients corresponding to a given triplet of representations) from the Coulomb-branch gauge theory matches both that of the explicitly-broken theory and that of the supergravity theory.
- Identification of the supergravity relies on the unbroken limit (setting all masses to zero), as well as on the symmetry breaking information encoded in the trilinear scalar couplings.

A list of constructions for Higgsed supergravities with various amounts of supersymmetry can be found in table 11.

Exercise 5.12. What would happen if we attempted to double copy two Coulomb-branch theories realized both in terms of compact momenta? Find out as much information as possible on the resulting gravity theory.

5.3.8. Gauged supergravities. In this subsection, we consider an important variant of the construction for Higgsed supergravities. In a sense, the construction outlined in the previous

Table 12. Fields in gauge-theory factors for a simple example of a double-copy construction for $\mathcal{N}=2$ gauged supergravities.

Fields	Representation	Mass	Fields	Representation	Mass
$(A_\mu, \bar{\phi}^a)$	$\mathcal{G}$	0	(A_μ, φ^α)	$\mathcal{G}$	0
(W_μ, φ^s)	$\mathcal{R}$	m	χ	$\overline{\mathcal{R}}$	m
$(\overline{W}_\mu, \overline{\varphi}^s)$	$\overline{\mathcal{R}}$	$-m$	$\overline{\chi}$	$\mathcal{R}$	$-m$

subsection can be regarded as the simplest double-copy prescription which produces a gravity with massive vector fields. Various details of the construction can then be traced back to the requirement that such massive vectors obey the relevant Ward identities corresponding to spontaneous symmetry breaking.

Along similar lines, we might want to consider double copies leading to massive spin-$3/2$ fields. It turns out that the construction will lead to gauged supergravities—supergravities in which a subgroup of the R symmetry is promoted to a gauge symmetry under which gravitini are charged. In a gauged supergravity with a Minkowski vacuum, minimal coupling between gravitini and gauge vector produces a nonzero amplitude of the form

$$\mathcal{M}_3\left(1\overline{\psi}_i, 2\psi_j, 3A^a\right) = i\, g_R t^a_{ij} \bar{v}^\mu_1 \not{\varepsilon}_3 v_{2\mu} + \mathcal{O}\left(\left(g_R\right)^0\right). \tag{226}$$

g_R is the coupling constant and $v_{l\mu}$ $(l=1,2)$ are the gravitini's polarization spinor-vectors. The matrices t^a_{ij} generate the gauged R-symmetry subgroup acting nontrivially on the gravitini. The above amplitude does not vanish with the replacement

$$v_{l\mu} \to v_{l\mu} + k_{l\mu}\epsilon\,, \qquad \not{k}_l\epsilon = 0\,. \tag{227}$$

Since this replacement correspond to an linearized supersymmetry transformation, the presence of a nonzero amplitude of the form (226) signifies that supersymmetry is spontaneously broken. Indeed, R-symmetry gauging and spontaneous supersymmetry breaking go hand in hand for supergravities which admit Minkowski vacua. In turn, the fact that supersymmetry is spontaneously broken results in (some) massive gravitini. This can be understood by comparing the number of physical polarizations of our gravitini; because some of the supersymmetry generators are broken, they cannot be used to eliminate components of gravitini. Some of the gravitini will have four physical polarizations and must therefore become massive.

This observation provides a hint on how to find a double-copy construction for amplitudes of gauged supergravities with Minkowski vacua. In analogy with the previous subsection, we start by seeking a construction that has the following two properties:

(i) contains massive spin-$3/2$ fields, realized as the double copies of a massive W bosons in one gauge theory with massive fermions in the other;
(ii) reduces to the construction of the corresponding ungauged supergravity in the massless limit.

The simplest realization with these properties has the schematic form

$$(\text{Gauged Supergravity}) = (\text{Coulomb-branch YM}) \otimes \left(\cancel{\text{super}}\text{ YM}\right), \tag{228}$$

where the second factor stands for a theory with explicit supersymmetry breaking and massive fermions.

Next, we discuss the two gauge theories separately, focusing on the particular case of $\mathcal{N}=2$ supersymmetry and using the toolbox introduced in sections 5.2.2 and 5.2.3. Unlike the case of Higgsed YME theories, the Coulomb-branch theory is non-supersymmetric; we will take it to be a pure YM theory coupled with n scalars, obtained from dimensional reduction from $D=(n+4)$ dimensions. The corresponding VEV will be parameterized by a n-dimensional vector which we will denote as V^a. The theory with supersymmetry explicitly broken by fermion masses is obtained by starting with four-dimensional $SU(N+M)$ $\mathcal{N}=2$ SYM theory and spontaneously breaking the gauge group to $G=SU(N)\times SU(M)\times U(1)$ by introducing a VEV

$$\langle\varphi_\alpha\rangle=\widetilde{V}_\alpha\times\mathrm{Diag}\left(\frac{1}{N}I_N,-\frac{1}{M}I_M\right). \tag{229}$$

We then orbifold the theory by a $\mathbb{Z}_2$ generated by $\gamma=\mathrm{diag}(I_N,-I_M)$:

$$A_\mu\mapsto\gamma A_\mu\gamma^{-1},\qquad \chi\mapsto-\gamma\chi\gamma^{-1},\qquad \varphi\mapsto\gamma\varphi\gamma^{-1}. \tag{230}$$

Note that, as explained in section 5.2.2, this operation preserves CK duality. The VEVs in both theories are chosen to have the same magnitude $(V^a)^2=(\widetilde{V}_\alpha)^2$, so that the two theories have common mass spectra. The explicitly-broken theory has Lagrangian

$$\mathcal{L}_{\not{\mathcal{N}}=2}=\frac{1}{g^2}\mathrm{Tr}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}-\frac{1}{2}D_\mu\varphi_\alpha D^\mu\varphi_\alpha+\frac{1}{4}[\varphi_\alpha,\varphi_\beta]^2+\frac{i}{2}\overline{\chi}\Gamma^\mu D_\mu\chi+\frac{1}{2}\overline{\chi}\Gamma^\alpha[\varphi_\alpha+\langle\varphi_\alpha\rangle,\chi]\right], \tag{231}$$

where χ is a six-dimensional Weyl fermion and $\alpha,\beta=5,6$. The fields in the gauge-theory factors are listed in table 12. Denoting with ξ_μ the massive gravitino field on the supergravity side, the fermionic states have the following double-copy origin:

$$\begin{aligned}\xi_\mu&=W_\mu\otimes\chi\;-\;W_\nu\otimes\left(\frac{\gamma_\mu}{3}-\frac{ip_\mu}{3m}\right)\gamma^\nu\chi,\\ \xi&=W_\nu\otimes\gamma^\nu\chi,\qquad (U\lambda)^s=\varphi^s\otimes\chi.\end{aligned} \tag{232}$$

The combination on the first line is manifestly transverse and γ-traceless. U is a unitary matrix diagonalizing the spin-$1/2$ mass terms. Last, the $U(1)_R$ gauge vector is:

$$A_+^{U(1)_R}=-A_+\otimes\varphi^6\;\pm\;\phi^2\otimes A_+. \tag{233}$$

We note that the massless limit leads an ungauged theory belonging to the Generic Jordan family discussed in section 5.3.2. The freedom of choosing the $U(1)_R$ gauge group corresponds to the choice of VEVs in the two gauge theories entering the construction. As for Higgsed supergravities, this is the simplest example of the double-copy construction. However, it is immediate to generalize the construction reviewed here to $U(1)_R$ gaugings of $\mathcal{N}=4,6,8$ supergravity by adjusting the supersymmetry of the gauge-theory factors.

Exercise 5.13. Introduce massive spinor-helicity notation by splitting massive momenta as $p_i=p_i^\perp-\frac{m^2}{2p_i\cdot q}q$. Here q is a reference momentum and $p_i^\perp,q$ are both massless. Write massive spinor polarizations as $v_+^t=\left(|i^\perp],\ m|q\rangle/\langle i^\perp q\rangle\right)$ and $v_-^t=\left(m|q]/[i^\perp q],\ |i^\perp\rangle\right)$. Show that an amplitudes involving massive gravitini with $\pm$ polarizations and the A^{-1} vector field can be written as

$$\mathcal{M}_3^{\text{tree}}\left(1\bar{\xi}_+,2\xi_-,3A_+^{-1}\right)=-\sqrt{2}im\,\Omega\frac{\langle 2^\perp q\rangle}{\langle 1^\perp q\rangle},\qquad \Omega=\frac{[3^\perp1^\perp]^3}{[1^\perp2^\perp][2^\perp3^\perp]}. \tag{234}$$

The construction outlined above can be generalized to allow gauging of nonabelian subgroups of the R symmetry. To do so, we need to consider double copies that [284]:

(i) contain massive spin-3/2 fields;
(ii) give the suitable ungauged supergravity in the massless limit;
(iii) involve gauge theories with trilinear scalar couplings.

The first two requirements parallel the abelian example discussed at the beginning of this section, while the last property reflects the fact that cubic couplings involving gauge-theory scalars result in nonabelian interactions in the theory from the double copy, as seen in the example of the construction for YME theories. As before, the gauge theories entering the double copy are obtained from higher dimension with a combination of Higgsing and orbifolding. We specialize to the case of gaugings of $\mathcal{N}=8$ supergravity and start by writing both copies of $\mathcal{N}=4$ SYM as the dimensional reduction of SYM theories in ten dimensions. For the left gauge-theory factor, we choose undeformed $\mathcal{N}=4$ SYM theory on the Coulomb branch. In the right gauge-theory factor, we introduce a massive deformation which involves trilinear scalar couplings,

$$\begin{aligned}\mathcal{L}=&-\frac{1}{4}\left(F^{\hat{a}}_{\mu\nu}\right)^2+\frac{1}{2}\left(D_\mu\phi^{\hat{a}I}\right)^2-\frac{1}{2}m^2_{IJ}\phi^{\hat{a}I}\phi^{\hat{a}J}-\frac{g^2}{4}f^{\hat{a}\hat{b}\hat{e}}f^{\hat{c}\hat{d}\hat{e}}\phi^{\hat{a}I}\phi^{\hat{b}J}\phi^{\hat{c}I}\phi^{\hat{d}J}-\frac{g\lambda}{3!}f^{\hat{a}\hat{b}\hat{c}}F^{IJK}\phi^{\hat{a}I}\phi^{\hat{b}J}\phi^{\hat{c}K}\\&+\frac{i}{2}\bar{\psi}\not{D}\psi-\frac{1}{2}\bar{\psi}M\psi+\frac{g}{2}\phi^{\hat{a}I}\bar{\psi}\Gamma^I t^{\hat{a}}_{\mathcal{R}}\psi\,.\end{aligned}\tag{235}$$

CK duality of the two-scalar-two-fermion four-point amplitude demands that the fermionic mass matrix M obey the relation

$$\left[\Gamma^I,\left\{\Gamma^J,M\right\}\right]+i\lambda F^{IJK}\Gamma^K=0\,,\tag{236}$$

where Γ^I are the Dirac matrices in higher dimensions and F^{IJK} are related to the structure constants of the supergravity gauge group. The right gauge theory is then Higgsed and orbifolded, following the same strategy outlined in the $\mathcal{N}=2$ example. Double copies involving theories obtained with this prescription need however to satisfy additional consistency requirements.

Referring to the literature for the general construction [284], we consider the action (235) with an $SU(3N)$ gauge group and the deformation

$$M=i\frac{g}{4}\Gamma^{789}\,,\qquad \lambda F^{789}=g\,.\tag{237}$$

This deformation breaks ten-dimensional Lorentz invariance to $SO(3)\times SO(6,1)$ and can be uplifted to seven dimensions. Starting from $D=7$, we take a $\mathbb{Z}_5$ orbifold which acts as

$$\psi\to e^{\frac{2\pi}{5}\Gamma_{56}}g^\dagger\psi\, g\,,\qquad \phi^I\to R^{IJ}\left(\frac{4\pi}{5}\right)g^\dagger\phi^J g\,,\qquad g=\mathrm{diag}\left(I_N,e^{i\frac{2\pi}{5}}I_N,e^{i\frac{4\pi}{5}}I_N\right)\tag{238}$$

where $I,J=5,6$ and R_{56} generates a rotation in the 5–6 plane. We also take the scalar mass-matrix to be

$$m_{55}=m=m_{66}\,,\qquad m_{IJ}=0\ \text{ otherwise}\,.\tag{239}$$

After the projection, the fields of the theory are organized schematically as:

$$\begin{pmatrix} A_\mu,\phi^i & \psi^r & \phi^+\\ \tilde{\psi}^{r'} & A_\mu,\phi^i & \psi^r\\ \phi^- & \tilde{\psi}^{r'} & A_\mu,\phi^i\end{pmatrix},\tag{240}$$

where $i=4,7,8,9$, $r=1,2$, $r'=3,4$, and $\phi^\pm=\phi^5\pm i\phi^6$. In the above equation, we represent the fields surviving the projection as entries the in $3N\times 3N$ matrices originating from the parent theory; each entry is an $N\times N$ block. To obtain a number of states that reproduces the spectrum

Table 13. Fields and mass spectra for gauging of $\mathcal{N}=8$ supergravity with $\mathcal{N}=4$ residual supersymmetry [284].

Rep.	R	L	Supergravity fields	mass^2
$\mathcal{G}$	$\mathcal{V}^0_{\mathcal{N}=4}$	$A_\mu \oplus \phi^i$	$\mathcal{H}_{\mathcal{N}=4} \oplus 4\mathcal{V}^0_{\mathcal{N}=4}$	0
$\mathcal{R}_1$	$\mathcal{V}^m_{\mathcal{N}=4}$	ψ^r	$2\Psi^m_{\mathcal{N}=4}$	u_1^2
$\bar{\mathcal{R}}_1$	$\mathcal{V}^m_{\mathcal{N}=4}$	$\tilde{\psi}^{r'}$	$2\Psi^m_{\mathcal{N}=4}$	u_1^2
$\mathcal{R}_2$	$\mathcal{V}^m_{\mathcal{N}=4}$	ϕ^+	$\mathcal{V}^m_{\mathcal{N}=4}$	$4u_1^2$
$\bar{\mathcal{R}}_2$	$\mathcal{V}^m_{\mathcal{N}=4}$	ϕ^-	$\mathcal{V}^m_{\mathcal{N}=4}$	$4u_1^2$

of $\mathcal{N}=8$ supergravity, we need to combine the representations $(\boldsymbol{N},\bar{\boldsymbol{N}},\mathbf{1})$ with $(\mathbf{1},\boldsymbol{N},\bar{\boldsymbol{N}})$ and the representation $(\bar{\boldsymbol{N}},\boldsymbol{N},\mathbf{1})$ with $(\mathbf{1},\bar{\boldsymbol{N}},\boldsymbol{N})$ into a (reducible) representation which is denoted as $\mathcal{R}_1$. This can be realized by rewriting the Lagrangian in a way that only representation matrices for $\mathcal{R}_1$ appear explicitly.

In the left theory, we take a a VEV of the form

$$\langle\phi^4\rangle = \mathrm{diag}\left(u_1 I_N, u_2 I_N, u_3 I_N\right), \qquad u_1+u_2+u_3=0\,. \tag{241}$$

Since the two irreducible representations that have been combined into $\mathcal{R}_1$ need to have the same mass, we get a condition involving the VEV parameters,

$$u_1-u_2=u_2-u_3 \quad\rightarrow\quad u_2=\frac{u_1+u_3}{2}=0\,. \tag{242}$$

In addition, we get the following conditions by matching the mass spectra of the two theories:

$$M^2=-u_1^2, \qquad\qquad m^2=4u_1^2\,. \tag{243}$$

We list the fields from the double copy with their respective mass spectra in table 13.

The vacuum of this theory has an unbroken $SU(2)\times U(1)$ gauge group which is reflected by the F^{IJK} tensors in (237). $\mathcal{N}=4$ unbroken supersymmetry is inherited from the Coulomb-branch gauge-theory factor. Many additional examples can be worked out along similar lines. A complete classification of double-copy-constructible gaugings is currently an open problem.

5.3.9. Conformal supergravity. A double-copy construction for conformal gravity was set forth in [152] and further investigated in [153]. Before we get into the details of that construction, let us review some general properties of conformal gravity. The simplest model is that of Weyl gravity, which has the four-derivative action

$$S=-\frac{1}{\varkappa^2}\int d^4x\sqrt{-g}\,\left(W_{\mu\nu\rho\sigma}\right)^2\,, \tag{244}$$

where $W_{\mu\nu\rho\sigma}$ is the Weyl curvature tensor, and $\varkappa$ is a dimensionless coupling. The action is invariant under local rescaling of the metric, $g_{\mu\nu}\to\Omega(x)g_{\mu\nu}$; more generally the theory possesses local conformal symmetry at the classical level. The symmetry can be extended to local superconformal symmetry by considering supergravity formulations of the Weyl theory. It is believed that $\mathcal{N}=4$ is the maximum allowed supersymmetry. In contrast to expectations from SYM and ordinary two-derivative supergravity, the maximally supersymmetric theory is not unique, in fact it has an infinite number of free parameters [349]. The free parameters are encoded in a free holomorphic function that multiplies the square of the Weyl tensor[44],

$$-\frac{\kappa^2}{\sqrt{-g}}\mathcal{L}_{\mathcal{N}=4} = f(\tau)\left(W^+_{\mu\nu\rho\sigma}\right)^2 + \overline{f(\tau)}\left(W^-_{\mu\nu\rho\sigma}\right)^2 + \dots, \tag{245}$$

where $W^{\pm}_{\mu\nu\rho\sigma} = W_{\mu\nu\rho\sigma}/2 \pm (i\sqrt{-g}/4)W_{\mu\nu}{}^{\lambda\kappa}\epsilon_{\lambda\kappa\rho\sigma}$ is the (anti-)selfdual Weyl tensor and the complex scalar $\tau = ie^{-\phi} + \chi$ is the dilaton-axion field. The ellipsis denotes additional terms that are fully constrained by the superconformal symmetry. The choice $f(\tau) = 1$ corresponds to the supersymmetrization of the Weyl theory, and it is usually called *minimal* conformal supergravity. When $f(\tau)$ is not constant, the theory corresponds to *non-minimal* conformal supergravity. The double-copy constructions that we will consider corresponds to the two cases [153]:

$$\begin{aligned}
f(\tau) &= -i\tau \qquad (\mathcal{N}=4 \text{ Berkovits-Witten theory})\,,\\
f(\tau) &= 1 \qquad\quad (\mathcal{N}=4 \text{ minimal conformal supergravity})\,.
\end{aligned} \tag{246}$$

These two cases are special. The Berkovits-Witten theory [350] corresponds to the unique conformal supergravity theory that has an uplift to 10 dimensions [145, 152, 153, 351]. At tree level, the minimal theory has the same $SU(1,1)$ electromagnetic duality symmetry as $\mathcal{N}=4$ supergravity, and certain all-multiplicity tree-level amplitudes are the same as in that theory. All $\mathcal{N}=4$ conformal supergravities are expected to be anomalous at loop level unless they are coupled to four vector multiplets [352, 353].

For reasons of conciseness, we will restrict the discussion in this section to scattering amplitudes where the external states are plane waves. As is well known, the four-derivative action of conformal gravity also permits other types of asymptotic states, see e.g. [153, 354] for further details. The double copy that gives amplitudes in the Berkovits–Witten conformal supergravity theory has the schematic form

$$(\text{Berkovits–Witten CSG}) = (\text{SYM}) \otimes \Big((DF)^2 - \text{theory}\Big)\,, \tag{247}$$

where SYM is the maximally supersymmetric YM theory, and the $(DF)^2$ theory is a bosonic gauge theory with dimension-six operators which has the following Lagrangian [152]:

$$\mathcal{L}_{(DF)^2} = \frac{1}{2}(D_\mu F^{a\mu\nu})^2 - \frac{g}{3}F^3 + \frac{1}{2}(D_\mu\varphi^\alpha)^2 + \frac{g}{2}C^{\alpha ab}\varphi^\alpha F^{\,a}_{\mu\nu}F^{b\mu\nu} + \frac{g}{3!}d^{\alpha\beta\gamma}\varphi^\alpha\varphi^\beta\varphi^\gamma\,.$$

The vector A^a_μ transforms in the adjoint representation of a gauge group G with indices a, b, c. φ^α are additional scalars transforming in a real representation for which $C^{\alpha\,a\,b}$ and $d^{\alpha\beta\gamma}$ are invariant tensors. We have used the short-hand notation $F^3 = f^{abc}F^{a\nu}_{\mu}F^{b\lambda}_{\nu}F^{c\mu}_{\lambda}$. It should be noted that $C^{\alpha ab}, T^a_{\mathcal{R}}$ and $d^{\alpha\beta\gamma}$ are implicitly defined through the two relations:

$$\begin{aligned}
C^{\alpha ab}C^{\alpha cd} &= f^{ace}f^{edb} + f^{ade}f^{ecb}\,,\\
C^{\alpha ab}d^{\alpha\beta\gamma} &= (T^a_{\mathcal{R}})^{\beta\alpha}\left(T^b_{\mathcal{R}}\right)^{\alpha\gamma} + C^{\beta ac}C^{\gamma cb} + (a \leftrightarrow b)\,,
\end{aligned} \tag{248}$$

which are sufficient relations for expressing tree-level gluon amplitudes in terms only f^{abc} tensors.

[44] Note that compared to [152] we are using a convention where we have swapped $i\bar{\tau}$ with $-i\tau$. This changes the sign of the axion field, which is physically unobservable.

A massive deformation of this theory was also introduced in [152] and is defined by the Lagrangian:

$$\mathcal{L}_{(DF)^2+\mathrm{YM}} = \frac{1}{2}\left(D_\mu F^{a\mu\nu}\right)^2 - \frac{g}{3}F^3 + \frac{1}{2}\left(D_\mu\varphi^\alpha\right)^2 + \frac{g}{2}C^{\alpha ab}\varphi^\alpha F^a_{\ \mu\nu}F^{b\mu\nu} + \frac{g}{3!}d^{\alpha\beta\gamma}\varphi^\alpha\varphi^\beta\varphi^\gamma$$
$$- \frac{1}{2}m^2\left(\varphi^\alpha\right)^2 - \frac{1}{4}m^2\left(F^a_{\ \mu\nu}\right)^2 . \tag{249}$$

This theory interpolates between the $(DF)^2$ theory and a pure YM theory and has the mass as a free parameter. Along similar lines, the theory (249) can be further augmented by introducing adjoint scalars ϕ^{aA} which are also charged under a global group and appear in trilinear couplings which are analogous to the ones introduced for YME theories and nonabelian gauged supergravities:

$$\mathcal{L}_{(DF)^2+\mathrm{YM}+\phi^3} = \frac{1}{2}\left(D_\mu F^{a\mu\nu}\right)^2 - \frac{g}{3}F^3 + \frac{1}{2}\left(D_\mu\varphi^\alpha\right)^2 + \frac{g}{2}C^{\alpha ab}\varphi^\alpha F^a_{\ \mu\nu}F^{b\mu\nu} + \frac{g}{3!}d^{\alpha\beta\gamma}\varphi^\alpha\varphi^\beta\varphi^\gamma$$
$$- \frac{1}{2}m^2\left(\varphi^\alpha\right)^2 - \frac{1}{4}m^2\left(F^a_{\ \mu\nu}\right)^2 + \frac{1}{2}\left(D_\mu\phi^{aA}\right)^2 + \frac{g}{2}C^{\alpha ab}\varphi^\alpha\phi^{aA}\phi^{bA}$$
$$+ \frac{g\lambda}{3!}f^{abc}F^{ABC}\phi^{aA}\phi^{bB}\phi^{cC} . \tag{250}$$

These deformations of the $(DF)^2$ theory will also play an important role for double-copy constructions involving various string theories, which are reviewed in the next subsections. We also note that the $(DF)^2$ theory is just a representative of a large class of gauge theories with higher-dimension operators. An investigation of their amplitudes in the general case is an open problem; we refer the reader to [296] for a similar construction of supergravities with higher-dimension operators and to [145] for a study of the $(DF)^2$ theory from the point of view of ambitwistor strings.

Exercise 5.14. Show that three- and four-gluon color-ordered amplitudes in the $(DF)^2$ theories have the expressions

$$A_{(DF)^2}(1,2,3) = -4\left(\varepsilon_1\cdot p_2\right)\left(\varepsilon_2\cdot p_3\right)\left(\varepsilon_3\cdot p_1\right) ,$$
$$A_{(DF)^2}(1,2,3,4) = 4\frac{s_{12}^2 s_{23}^2}{s_{13}}\left(\frac{p_4\cdot\varepsilon_1}{s_{14}} - \frac{p_2\cdot\varepsilon_1}{s_{12}}\right)\left(\frac{p_1\cdot\varepsilon_2}{s_{12}} - \frac{p_3\cdot\varepsilon_2}{s_{23}}\right)\left(\frac{p_2\cdot\varepsilon_3}{s_{23}} - \frac{p_4\cdot\varepsilon_3}{s_{12}}\right)$$
$$\times\left(\frac{p_3\cdot\varepsilon_4}{s_{12}} - \frac{p_1\cdot\varepsilon_4}{s_{23}}\right) , \tag{251}$$

and that they obey color-kinematics duality. Note that the products between polarization vectors, $\varepsilon_i\cdot\varepsilon_j$, always cancel out (this is a special property of the $(DF)^2$ theory).

Finally, we will consider amplitudes in the minimal $\mathcal{N}=4$ conformal supergravity theory. For external plane waves at tree level, the relation is

$$(\text{minimal CSG}) = (\text{SYM}) \otimes \left(\text{minimal } (DF)^2\text{-theory}\right) , \tag{252}$$

where we have truncated the bosonic gauge theory to a ‘minimal’ version,

$$\mathcal{L}_{\text{min.}\,(DF)^2} = \frac{1}{2}\left(D_\mu F^{a\mu\nu}\right)^2 . \tag{253}$$

However, as the reader may confirm, the all tree-level plane-wave amplitudes in this theory vanish—a property that is also true of minimal conformal supergravity. In order to have something nonvanishing to compare with, we must deform the two theories by a mass term,

$$\mathcal{L}_{\text{min.}\,(DF)^2+\text{YM}} = \frac{1}{2}\left(D_\mu F^{a\mu\nu}\right)^2 - \frac{1}{4}m^2\left(F^a{}_{\mu\nu}\right)^2 . \tag{254}$$

The resulting double copy

$$(\text{mass-deformed minimal CSG}) = (\text{SYM}) \otimes \left(\text{minimal } (DF)^2 + YM\right), \tag{255}$$

gives amplitudes in a mass-deformed minimal $\mathcal{N}=4$ theory that interpolates between (Weyl)^2 and a Ricci scalar term

$$-\varkappa^2\sqrt{-g}^{-1}\mathcal{L}_{\mathcal{N}=4} = \left(W_{\mu\nu\rho\sigma}\right)^2 - 2m^2 R + \ldots \tag{256}$$

where the ellipsis are additional terms fixed by supersymmetry. The tree amplitudes, for external plane waves, in the mass-deformed theories, are proportional to the corresponding amplitudes in ordinary YM and supergravity [153],

$$\begin{aligned} A_{\text{min.}\,(DF)^2+\text{YM}} &= m^2 A_{\text{YM}}\,,\\ M_{\text{mass-def. min. CSG}} &= m^2 M_{\text{SG}}\,. \end{aligned} \tag{257}$$

In addition to considering $\mathcal{N}=4$ conformal supergravity, the corresponding theories with reduced supersymmetry $\mathcal{N}=0,1,2$ can be obtained by replacing the $\mathcal{N}=4$ SYM factor in the double copies (247), (252) and (255) by $\mathcal{N}=0,1,2$ (S)YM. The $\mathcal{N}=0,1,2$ conformal (super)gravity theories will not be pure, as they will inherit a dilaton-axion multiplet from the $\mathcal{N}=4$ theory, in close analogy to the case of ordinary two-derivative supergravity theories.

5.3.10. Perturbative string theories. In references [109, 355], disk integrals that appear in open-string amplitudes were organized in terms of building blocks

$$Z_\sigma\left(\rho(1,2,\ldots,n)\right) = \left(2\alpha'\right)^{n-3} \int_{\sigma\{-\infty\leqslant z_1\leqslant z_2\leqslant\ldots\leqslant z_n\leqslant\infty\}} \frac{\mathrm{d}z_1\,\mathrm{d}z_2\ldots\mathrm{d}z_n}{\text{vol}\left(\text{SL}\left(2,\mathbb{R}\right)\right)} \frac{\prod_{i<j}^n |z_{ij}|^{\alpha' s_{ij}}}{\rho\left\{z_{12}z_{23}\cdots z_{n-1,n}z_{n,1}\right\}}. \tag{258}$$

Here we use the notation $z_{ij}=z_i-z_j$ and vol(SL$(2,\mathbb{R})$) refers to fixing three punctures on the disk to $z_i,z_j,z_k\to(0,1,\infty)$ while introducing a Jacobian $|z_{ij}z_{ik}z_{jk}|$. Such building blocks depend on two permutations $\sigma,\rho\in S_n$ and obey field-theory BCJ relations with respect to ρ,

$$\sum_{j=2}^{n-1}\left(p_1\cdot p_{23\ldots j}\right) Z_\sigma\left(2,3,\ldots,j,1,j+1,\ldots,n\right) = 0\,, \tag{259}$$

and string-theory monodromy relations [104, 105] with respect to σ,

$$\sum_{j=1}^{n-1} e^{2i\pi\alpha' p_1\cdot p_{23\ldots j}} Z_{(2,3,\ldots,j,1,j+1,\ldots,n)}\left(\rho\right) = 0\,. \tag{260}$$

One may therefore think of $Z_\sigma(1,\rho(2,\ldots,n-2),n-1,n)$ as partial amplitudes ordered with respect to two symmetry groups. One of them corresponds to dressing Z with traces built out of Chan–Paton factors following the permutation σ and the other corresponds to dressing it with trace color factors, unrelated to the Chan–Paton factors, following the permutation ρ.

With these building blocks, the open-superstring amplitudes with massless external states color-ordered, with respect to the Chan–Paton factors, can be expressed directly in terms of

YM scattering amplitudes [109], and written in terms of a field theoretic double-copy factorization in [355],

$$A_{\mathrm{OS}}^{\mathrm{tree}}\left(\sigma\left(1,2,3,\ldots,n\right)\right) = \sum_{\tau,\rho\in S_{n-3}(2,\ldots,n-2)} Z_\sigma\left(1,\tau,n,n-1\right)S[\tau|\rho]A_{\mathrm{SYM}}\left(1,\rho,n-1,n\right), \tag{261}$$

where the $(n-3)!\times(n-3)!$ matrix $S[\tau|\rho]=S[\tau(2,\ldots,n-2)|\rho(2,\ldots,n-2)]$ is the field-theory KLT kernel[45] introduced in section 2.3.1. It is fascinating to note that a suggestive hint of this type of field-theoretic double-copy factorization was identified in [357].

Rather than focusing on the partially-ordered open string amplitudes (261), consider instead the content of the full Chan–Paton-dressed open supersymmetric string amplitude. Dressing Z_σ with all relevant $(n-1)!$ traces built out of Chan–Paton factors, yields a singly ordered function,

$$\mathbf{Z}^{\mathrm{tree}}\left(1,\ldots,n\right) \equiv \sum_{\sigma\in S_{n-1}(2,\ldots,n)} \mathrm{Tr}\left[T^{a_1}T^{a_{\sigma(2)}}\cdots T^{a_{\sigma(n-1)}}T^{a_{\sigma(n)}}\right]\sum_{\rho\in S_{n-3}} Z_\sigma\left(1,\rho,n-1,n\right), \tag{262}$$

which obeys only the field-theory amplitude relations (i.e. equation (259) with the replacement $Z_\sigma\to\mathbf{Z}^{\mathrm{tree}}$). The full Chan–Paton-dressed open superstring amplitude,

$$\mathcal{A}_{\mathrm{OS}} = \sum_{\sigma\in S_{n-1}} \mathrm{Tr}\left[T^{a_1}T^{a_{\sigma(2)}}\cdots T^{a_{\sigma(n-1)}}T^{a_{\sigma(n)}}\right]A_{\mathrm{OS}}^{\mathrm{tree}}\left(1,\sigma\right), \tag{263}$$

can also be written entirely as a field-theory double copy

$$\mathcal{A}_{\mathrm{OS}} = \sum_{\tau,\rho\in S_{n-3}} \mathbf{Z}^{\mathrm{tree}}\left(1,\tau,n,n-1\right)S[\tau|\rho]A_{\mathrm{SYM}}\left(1,\rho,n-1,n\right). \tag{264}$$

An interesting open problem is the physical interpretation of the above building blocks. Given the adjoint field-theory relations obeyed by the ordered $\mathbf{Z}(\rho)$, it is natural to consider the orderless-functions resulting by dressing the ρ ordering with adjoint f^{abc} structure constants as per a DDM basis. This yields a fully dressed function that can be expressed in terms of cubic graphs dressed with two factors that both satisfy Jacobi identities and antisymmetry:

$$\mathcal{Z} = \sum_i \frac{\mathbf{z}_i c_i}{D_i} = \sum_{\rho\in S_{n-2}} c_{1|\rho|n}\mathbf{Z}^{\mathrm{tree}}\left(1,\rho,2\right), \tag{265}$$

such that:

$$\mathbf{Z}^{\mathrm{tree}}\left(1,\rho\left(2\right),\ldots\rho\left(m-1\right),m\right) = -i\sum_{i\in\mathrm{planar}} b_{i\,\rho}\frac{\mathbf{z}_i}{D_i}, \tag{266}$$

where $\mathbf{z}_i$ are Jacobi satisfying functions of both higher-derivative scalar kinematics and string Chan–Paton factors, D_i are the propagators of the graph, and $b_{i\,\rho}\in\{0,\pm1\}$ are integer coefficients that depend on the ordering ρ. Both $\mathcal{Z}$ and $\mathbf{Z}$ can be derived as the tree-level amplitudes, color-dressed and ordered respectively, of an effective field theory of double-colored scalar fields in which the scalars obey an equation of motion of the schematic form [171]

$$\Box\Phi = \Phi^2+\alpha'^2\zeta_2\left(\partial^2\Phi^3+\Phi^4\right)+\alpha'^3\zeta_3\left(\partial^4\Phi^3+\partial^2\Phi^4+\Phi^5\right)+\mathcal{O}\left(\alpha'^4\right). \tag{267}$$

This theory was named Z-theory in references [169, 171, 310]. It is worth noting that the color structure of the leading term in the equation of motion is the same as the bi-adjoint ϕ^3 theory.

[45] Note that the α'-dependent KLT kernel, given in [24] (and its inverse in [356]), needs not feature in the factorization of tree-level string amplitudes, cf equation (270).

The entire tower of higher derivative operators relevant to the open-string are encoded in this effective scalar theory, whose double copy with the supersymmetric gauge theory yields the supersymmetric open string. Schematically, the formula (264) can be rewritten with the short-hand notation

$$(\text{massless open superstring}) = (\text{Z-theory}) \otimes (\text{SYM}) \,. \tag{268}$$

The simplest set of Z-theory amplitudes arise when one trivializes the string Chan-Paton factors, taking all the generators to be the identity, corresponding to a $U(1)$ group. This operation on the Chan-Paton dressed open string results in a symmetrization over all orders referred to as the abelian or photonic open-string whose low-energy limit yields amplitudes in maximally supersymmetric DBI theory, where the fermionic sector is of Volkov–Akulov type [302–305, 358–362]. Abelian $\mathcal{Z}$ amplitudes yield in the low-energy limit NLSM amplitudes[46] in the $\alpha' \to 0$ [169]. This is consistent with the realization that the NLSM double-copies with $\mathcal{N}=4$ SYM in four dimensions to generate DBI-VA amplitudes [125, 307].

A closed-string version of the Z-theory amplitudes involves integrals over the moduli space of punctured Riemann spheres [155, 363–365]

$$\mathrm{sv}\, Z(\tau|\sigma) = \left(\frac{2\alpha'}{\pi}\right)^{n-3} \int \frac{\mathrm{d}^2 z_1\, \mathrm{d}^2 z_2 \ldots \mathrm{d}^2 z_n}{\mathrm{vol}\,(\mathrm{SL}\,(2,\mathbb{C}))} \, \frac{\prod_{i<j}^n |z_{ij}|^{2\alpha' s_{ij}}}{\tau\left\{\bar{z}_{12}\bar{z}_{23}\cdots\bar{z}_{n-1,n}\bar{z}_{n,1}\right\} \sigma\left\{z_{12}z_{23}\cdots z_{n-1,n}z_{n,1}\right\}} \,. \tag{269}$$

The notation sv$Z$ refers to the so-called single-valued projection of multiple zeta values (MZVs), which can be regarded as a formal operation acting on the building blocks which arise in the construction for tree amplitudes of massless open-superstring states in the low-energy expansion [366, 367]. Here, we use (269) as the definition of $\mathrm{sv}\, Z(\tau|\sigma)$. Amplitudes in the Z-theory, together with their closed-string counterparts $\mathrm{sv}\, Z(\tau|\sigma)$, enter a particular class of tree-level double-copy constructions which combine the amplitudes of a string theory with the amplitudes of a gauge theory. For example, amplitudes in the closed superstring with massless asymptotic states can be obtained with the construction [368, 369]

$$(\text{closed superstring}) = (\text{SYM}) \otimes \mathrm{sv}\,(\text{open superstring}) \,. \tag{270}$$

Remarkably, various incarnations of the $(DF)^2$ theory introduced in the previous subsection in a completely different context enter these double-copy constructions for string amplitudes [370]:

$$(\text{open bosonic string}) = (\mathrm{Z-theory}) \otimes \left((DF)^2 + \mathrm{YM}\right) , \tag{271}$$

$$(\text{closed bosonic string}) = \left((DF)^2 + \mathrm{YM}\right) \otimes \mathrm{sv}\,(\text{open bosonic string}) \,, \tag{272}$$

$$(\text{heterotic string}) = \left((DF)^2 + \mathrm{YM} + \phi^3\right) \otimes \mathrm{sv}\,(\text{open superstring}) \,. \tag{273}$$

We should note that these constructions are of the generic form (261), i.e. they involve the field-theory KLT kernel, and apply at tree level and with massless external states. Remarkably, the free mass parameter in the $(DF)^2$+YM theory is related to the inverse string tension α' as

$$m^2 = -\frac{1}{\alpha'} \,. \tag{274}$$

[46] See equation (115) for one representation of the action.

Table 14. Various known double-copy constructions of string amplitudes [370]. The single-valued projection sv(•) converts disk to sphere integrals.

string $\otimes$ QFT	SYM	$(DF)^2 + \mathrm{YM}$	$(DF)^2 + \mathrm{YM} + \phi^3$
Z − theory	open superstring	open bosonic string	compactified open bosonic string
sv (open superstring)	closed superstring	heterotic (gravity)	heterotic (gauge/gravity)
sv (open bosonic string)	heterotic (gravity)	closed bosonic string	compactified closed bosonic string

Various relations between Z-theory and string amplitudes are summarized in table 14. Some extensions to loop level can be found in references [371–377]. Additionally, a double-copy construction for string amplitude in terms of field-theory amplitudes in the CHY formalism was obtained in [378, 379].

5.3.11. Other theories. We conclude the section by listing further examples of double-copy constructions.

- The non-gravitational (supersymmetric) DBI theory was constructed in [125] using the scattering Equation formalism (see also [285, 301]). It can be regarded as the double copy of (S)YM theory and the NLSM. It should be noted that the NLSM can be obtained in the $\alpha' \to 0$ limit of abelian Z-theory [169]. A further interesting feature of the NLSM is that it admits a Lagrangian in which the duality between color and kinematics is manifest [309].
- Similarly, the (supersymmetric) DBI theory coupled to (S)YM theory can be constructed as a double copy involving (S)YM theory and the NLSM coupled to bi-adjoint ϕ^3 theory [306]. The latter gauge-theory factor can be obtained from the $\alpha' \to 0$ limit of partially-Abelianized Z-theory [310].
- The DBI theory coupled to the NLSM can be constructed as a double copy involving YM coupled to bi-adjoint ϕ^3 theory and the NLSM [156].
- Volkov–Akulov theory has tree-level amplitudes that can be obtained from supersymmetric DBI by restricting the external states to be fermions. Since DBI only has nonvanishing even-point amplitudes, and internal bosons would require tree-level factorization with an odd number of particles (2×fermions + 1 boson), this restriction gives a consistent truncation of the theory. The double copy for Volkov–Akulov theory can thus be inferred to be a product between NLSM and SYM with only external fermions.
- Two copies of the NLSM give the so-called special-Galileon theory [125, 301].
- In three dimensions, two copies of the BLG theory [245, 246] yield an alternative construction for maximal three-dimensional supergravity [119, 243, 244, 297, 380]. The three-dimensional version of CK duality relevant to this construction is based on a so-called three-algebra. The three-algebra for BLG theory is introduced formally using a totally antisymmetric triple product $[X, Y, Z]$. Using a basis of generators the triple product can be expressed using rank-four structure constants,

$$\left[T^a, T^b, T^c\right] = f^{abc}{}_d T^d \,. \tag{275}$$

Consistency of the algebra requires that the structure constants satisfy the four-term identity

$$0 = f^{abc}{}_l f^{dleg} + f^{bae}{}_l f^{dlcg} + f^{ceb}{}_l f^{dalg} + f^{eca}{}_l f^{dblg} \,, \tag{276}$$

which plays the same role as the standard Jacobi identity for a Lie two-algebra. It turns out that the only nontrivial compact three-algebra is SO(4) [381], where $f^{abc}{}_d = \epsilon^{abcd}$. However, for color-kinematics duality to work, it is sufficient to impose the four-term identity, whereas identities specific to SO(4) should be ignored. Finally, we may note that the closely-related ABJM theory [382] appears to not have similarly nice properties under color-kinematics duality. While the tree-level ABJM amplitudes up to six points obey the duality and their double copy gives three-dimensional maximal supergravity, at eight points the double copy does not reproduce the corresponding amplitude in maximal supergravity [243, 244]. Since there is no dynamical graviton in three dimension, this mismatch is not forbidden by the diffeomorphism symmetry argument in section 2.5.

Additional theories for which a double-copy construction has been proposed involve massive higher-spin $\mathcal{N}=7$ W-supergravity theories [383, 384]; this amount of supersymmetry has not been accessible through different constructions. Chiral higher-spin theories have been shown to obey generalized BCJ relations in [385]. Theories with gravitationally-coupled fermions have been discussed in [386]. A construction of the free spectrum of $D=3$ supergravities in terms of SYM theories with fields valued in the four division algebras was given in [387]. Further examples of constructions in higher dimensions include half-maximal supergravity in six dimensions [388] and the so-called $(4,0)$ theory in six dimensions [268, 389], which can be seen as the double copy of two $(2,0)$ theories, at least at the level of the free spectrum [390, 391].

6. BCJ duality at loop level

In this section, we describe loop-level examples of BCJ duality and the associated double-copy construction. Whenever a gauge-theory integrand can be found in a form that manifests the duality between color and kinematics, corresponding gravity integrands can be immediately written down via the double-copy procedure. This procedure enormously simplifies the construction of gravity loop integrands and has been successful for carrying out a variety of loop-level studies in perturbative quantum gravity theories (see e.g. [15, 17, 18, 23, 31–33, 36, 292, 293]). As explained in section 5, the precise gravity theory to which the integrands belong depends on the choice of input gauge theories. We start by briefly recalling the definition and the main points of the duality and of the double-copy construction, discussed at length in section 2. With the appropriate separation of diagrams' symmetry factors and judicious choice of loop momenta, they are essentially the same as at tree level.

Similarly to tree-level amplitudes, loop-level amplitudes in a gauge theory coupled to matter fields can be organized as a sum over diagrams with only cubic (trivalent) vertices by multiplying and dividing by appropriate propagators to absorb contact diagrams into diagrams with only cubic vertices. If all fields are in the adjoint representation of the gauge group, this rearrangement puts the amplitude in a form equivalent to equation (35),

$$\mathcal{A}_m^{L\text{-loop}} = i^{L-1} g^{m-2+2L} \sum_{\mathcal{S}_m} \sum_j \int \prod_{l=1}^{L} \frac{d^D \ell_l}{(2\pi)^D} \frac{1}{S_j} \frac{c_j n_j(\ell)}{D_j}\,, \tag{277}$$

where the c_i are color factors obtained by assigning structure constant factors $\tilde{f}^{abc} = i\sqrt{2} f^{abc}$ to each cubic vertex. The first sum runs over the set $\mathcal{S}_m$ of $m!$ permutations of the external legs. The second sum runs over the distinct L-loop m-point diagrams with only cubic vertices. As at tree level, by multiplying and dividing by propagators, it is trivial to absorb contribution from higher-than-three-point vertices into numerators of diagrams with only cubic vertices.

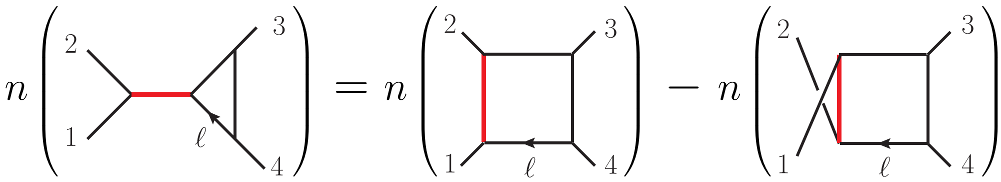


Figure 20. A BCJ kinematic numerator relation at one loop. When the external particles are gluons this holds just as well for adjoint or fundamental representation particles circulating in the loop. The shaded (red) line differs between the diagrams, but the others are identical.

The symmetry factor S_j counts the number of automorphisms of the labeled diagram j from both the permutation sum and from any internal automorphism symmetries[47]. This symmetry factor should not be included in the kinematic numerator.

The nontrivial conjecture is that, as at tree level, for every loop-level color Jacobi identity there is a matching kinematic numerator identity (41).

$$c_i - c_j = c_k \quad \Leftrightarrow \quad n_i(\ell) - n_j(\ell) = n_k(\ell)\,. \tag{278}$$

However, unlike at tree level, one has to be cautious with the treatment of degrees of freedom that are not fixed by the external states. This includes proper accounting of the loop momenta of the numerators, generically called ℓ, as well as being careful to not set to zero color factors that vanish when summing over internal indices.

We can change the signs of the color factors using the antisymmetry of the f^{abc}s, but any relative signs between color factors in the Jacobi relation are then inherited by the corresponding relation between the kinematic numerator factors. A simple example of such loop-level relations is illustrated in figure 20 for the case of a one-loop amplitude. At loop-level, the duality between color and kinematics (41) remains a conjecture [2], although evidence in its favor continues to accumulate [4–6, 9–23, 156].

Just as for tree-level numerators, once gauge-theory numerator factors which satisfy the duality are available, replacing the color factors by the corresponding numerator factors, $c_i \to n_i$ yields the double-copy form of gravity loop integrands (45),

$$\mathcal{M}_m^{L\text{-loop}} = i^{L-1}\left(\frac{\kappa}{2}\right)^{m-2+2L}\sum_{\mathcal{S}_m}\sum_j\int\prod_{l=1}^{L}\frac{d^D\ell_l}{(2\pi)^D}\frac{1}{S_j}\frac{\tilde{n}_j(\ell)\,n_j(\ell)}{D_j}\,, \tag{279}$$

where $\tilde{n}_j$ and n_j are gauge-theory numerator factors, which can come from distinct gauge theories and κ is the gravitational coupling defined below equation (5). The duality needs to be manifest in only one of the two gauge-theory amplitudes for the double-copy formula to hold.

6.1. One-loop examples of BCJ duality: $\mathcal{N} = 4$ SYM theory

The simplest example that illustrates CK duality at loop level is the one-loop four-point superamplitude of $\mathcal{N} = 4$ SYM theory. These amplitudes are remarkably simple, making them very useful for this purpose.

[47] Note that this symmetry factor is different from the symmetry factor in equation (35), where S_j counts the automorphisms of graphs with fixed external legs.

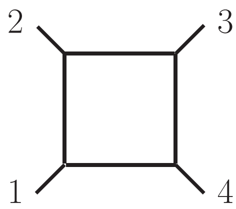


Figure 21. A one-loop box integral, $I_4(s,t)$, appearing in the one-loop four-point $\mathcal{N}=4$ SYM and $\mathcal{N}=8$ supergravity amplitudes. The three independent relabelings of external legs appear in the amplitudes.

The Jacobi identity obeyed by the structure constants of any Lie algebra guarantees that, in any gauge theory with all fields in the adjoint representation of the gauge group, any one-loop four-point amplitude can be organized as

$$\begin{aligned}\mathcal{A}_4^{1-\text{loop}}(1,2,3,4)\\ =g^4\Big(c_{1234}A_4^{1-\text{loop}}(1,2,3,4)+c_{1243}A_4^{1-\text{loop}}(1,2,3,4)+c_{1423}A_4^{1-\text{loop}}(1,2,3,4)\Big),\end{aligned} \tag{280}$$

where the color factor c_{1234} in equation (280) corresponds to the one of the box diagram in figure 21 and is given by dressing each three-point vertex with an f^{abc} structure constant, and summing over all repeated indices,

$$c_{1234}=4f^{ba_1c}f^{ca_2d}f^{da_3e}f^{ea_4b}\,. \tag{281}$$

The other two color factors are obtained by relabeling and we normalized c_{1234} following standard conventions [88]. Passing to a trace basis for the color factors identifies $A^{1-\text{loop}}(1,2,3,4)$ with the one-loop four-point color-ordered amplitudes. The form (280) can be obtained by applying the color Jacobi identity to the color factors of any valid representation (e.g. Feynman diagrams) of the amplitude to trade other color factors in favor of the three box ones [180]. Similar manipulations, together with use of the defining commutation relations of the Lie algebra, can be used to map the color factors of all one-loop four-point amplitudes in a theory with fields in any representation to the color factors of a box diagram; in this subsection we will however restrict ourselves to theories with fields in the adjoint representation.

Exercise 6.1. Prove equation (280) by starting from standard Feynman diagrams and then applying color Jacobi identities to express all color factors in terms of the color factors of the box diagrams.

Consider now the one-loop four-point superamplitude of $\mathcal{N}=4$ SYM theory. Each color-ordered superamplitude in equation (280) is especially simple and given by [392]

$$A_{\mathcal{N}=4}^{1-\text{loop}}(1,2,3,4)=i\,stA_{\mathcal{N}=4}^{\text{tree}}(1,2,3,4)I_4(s,t)\,, \tag{282}$$

where $I_4(s,t)$ is the box integral illustrated in figure 21, $s=(p_1+p_2)^2$ and $t=(p_2+p_3)^2$ are the usual Mandelstam invariants and $A_{\mathcal{N}=4}^{\text{tree}}(1,2,3,4)$, standing for the $n=4$ case of equation (465), is the color-ordered four-point tree superamplitude.

Since the diagram structure of the kinematic propagators in the three color-ordered amplitudes entering equation (280) matches that of their color factors, the kinematic numerators of the representation (277) of the one-loop amplitude can be straightforwardly identified. The combination $stA^{\text{tree}}_{\mathcal{N}=4}(1,2,3,4)$ is fully crossing-symmetric, as a consequence of the BCJ four-point tree-level amplitude relations (28), so all three numerators are the same,

$$n_{1234} = n_{1243} = n_{1423} = i\,stA^{\text{tree}}_{\mathcal{N}=4}(1,2,3,4) = \frac{[12]\,[34]}{\langle 12\rangle\langle 34\rangle}\delta^{(8)}\left(\sum_{i=1}^{4}\lambda_i\eta_i^I\right), \tag{283}$$

where we have specialized to four-dimensional external kinematics in the last equality.

Because triangle and bubble diagrams do not appear in the $\mathcal{N}=4$ SYM amplitude (282) (or, alternatively, they enter with vanishing numerators), it is straightforward to check, using equation (283), that the BCJ duality relation illustrated in figure 20 holds. The remaining kinematic Jacobi relations are also satisfied for similar reasons.

The corresponding $\mathcal{N}=8$ supergravity amplitude follows immediately from the basic double-copy substitution (44), replacing color factors by numerators and compensating for the change in coupling. This gives

$$\mathcal{M}^{1-\text{loop}}_{\mathcal{N}=8}(1,2,3,4) = -i\,stu\mathcal{M}^{\text{tree}}_{\mathcal{N}=8}(1,2,3,4)\left(I_4(s,t)+I_4(s,u)+I_4(t,u)\right), \tag{284}$$

where we used (4),

$$\left(\frac{\kappa}{2}\right)^4\left(stA^{\text{tree}}_{\mathcal{N}=4\,\text{SYM}}(1,2,3,4)\right)^2 = stu\mathcal{M}^{\text{tree}}_{\mathcal{N}=8}(1,2,3,4), \tag{285}$$

to replace the square of the $\mathcal{N}=4$ SYM four-point tree-level amplitude with the $\mathcal{N}=8$ supergravity four-point tree-level amplitude. This is a consequence of the KLT relations (31) and the BCJ amplitude relation (28). The amplitude in equation (284) reproduces the known $\mathcal{N}=8$ supergravity four-point tree-level amplitude [194, 392].

The explicit value of the massless scalar box integral $I_4(s,t)$ appearing in both the $\mathcal{N}=4$ SYM and $\mathcal{N}=8$ supergravity one-loop four-point amplitudes is

$$I_4(s,t) = \int\frac{d^D\ell}{(2\pi)^D}\frac{1}{\ell^2\,(\ell-p_1)^2\,(\ell-p_1-p_2)^2\,(\ell+p_4)^2}, \tag{286}$$

where the p_i's are the external momenta and the Feynman $i\varepsilon$ prescription, not included explicitly, is used to define the propagators. In dimensional regularization, we take $D=4-2\epsilon$ with ϵ small. The explicit functional form of $I_4(s,t)$ is (see e.g. [393, 394])

$$I_4(s,t) = i\frac{c_\Gamma}{st}\left[\frac{2}{\epsilon^2}\left((-s)^{-\epsilon}+(-t)^{-\epsilon}\right)-\ln^2\left(\frac{-s}{-t}\right)-\pi^2\right]+\mathcal{O}(\epsilon), \tag{287}$$

with

$$c_\Gamma = \frac{(4\pi)^\epsilon}{16\pi^2}\frac{\Gamma(1+\epsilon)\Gamma(1-\epsilon)^2}{\Gamma(1-2\epsilon)}. \tag{288}$$

The other box integrals can be obtained from this one by relabeling. Using these explicit expressions one can verify general properties of (super)gravity amplitudes, such as existence of only soft infrared (IR) divergences.

We can use equation (280), together with the $\mathcal{N}=4$ SYM numerators (283), to immediately obtain the four-point amplitudes of any $4\leqslant\mathcal{N}\leqslant 8$ supergravity after integration. Because the duality satisfying $\mathcal{N}=4$ four-point SYM kinematic numerators (283) are independent of the loop momentum, they come out of the integral as in equation (284) and behave essentially the same way as color factors. Thus, to obtain results for $\mathcal{N}\geqslant 4$ supergravity, we can start with equation (280) evaluated for $\mathcal{N}\leqslant 4$ (S)YM theory and replace the color factors with the $\mathcal{N}=4$ SYM numerators in equation (283). This gives us a general representation of the four-point amplitudes of all $\mathcal{N}\geqslant 4$ supergravities:

$$\mathcal{M}^{1-\text{loop}}_{\mathcal{N}+4\,\text{susy}}(1,2,3,4)$$
$$= \left(\frac{\kappa}{2}\right)^4 i\, st A^{\text{tree}}_4(1,2,3,4)\left(A^{1-\text{loop}}_{\mathcal{N}\,\text{susy}}(1,2,3,4) + A^{1-\text{loop}}_{\mathcal{N}\,\text{susy}}(1,2,4,3) + A^{1-\text{loop}}_{\mathcal{N}\,\text{susy}}(1,4,2,3)\right). \tag{289}$$

As explained above, $A^{\text{1-loop}}_{\mathcal{N}\,\text{susy}}$ are one-loop color-ordered gauge-theory amplitudes after loop integration for a theory with $\mathcal{N}$ (including zero) supersymmetries (cf equation (280)). This expression applies just as well for external matter multiplets in $\mathcal{N}=4$ supergravity. The needed integrated gauge-theory amplitudes may be found in [270, 393].

The double copy of amplitudes of gauge theories with $\mathcal{N}<4$ supersymmetry is less straightforward because the required gauge-theory numerators are in general not independent of loop momenta. Because of this, although the double-copy construction holds at the integrand level, one cannot simply carry over the integrated results from gauge to gravity theories. It is nevertheless remarkable that there is such a simple relation between these two different theories.

To illustrate CK duality and the double-copy construction at one loop, we consider the one-loop identical-helicity four-gluon amplitude in QCD with N_f quark flavors in the fundamental representation, originally constructed in [270]. It is[48]

$$\mathcal{A}^{1-\text{loop}}_{\text{QCD}}(1^+,2^+,3^+,4^+) = 2g^4\frac{[12]\,[34]}{\langle 12\rangle\langle 34\rangle}\left(\left(c_{1234} - N_f\, c^f_{1234}\right) I_4(s,t)\left[\mu^4\right]\right.$$
$$\left. + \left(c_{1243} - N_f\, c^f_{1234}\right) I_4(s,u)\left[\mu^4\right] + \left(c_{1423} - N_f\, c^f_{1234}\right) I_4(t,u)\left[\mu^4\right]\right), \tag{290}$$

where the color factor associated with the quark loop is

$$c^f_{1234} = \text{Tr}\left[T^{a_1}T^{a_2}T^{a_3}T^{a_4}\right] + \text{Tr}\left[T^{a_4}T^{a_3}T^{a_2}T^{a_1}\right]. \tag{291}$$

For simplicity, we have assumed that the quarks are massless. Here μ is the (-2ϵ)-dimensional component of loop momentum, so

$$\ell = \ell^{(4)} + \mu, \qquad \ell^2 = \left(\ell^{(4)}\right)^2 - \mu^2, \tag{292}$$

and $I_4(s,t)[\mu^4]$ is the integral corresponding to the diagram in figure 21 with a μ^4 numerator factor. As required by Bose symmetry, the prefactor is fully cross symmetric, i.e.

$$\frac{[12]\,[34]}{\langle 12\rangle\langle 34\rangle} = \frac{[23]\,[41]}{\langle 23\rangle\langle 41\rangle} = \frac{[13]\,[24]}{\langle 13\rangle\langle 24\rangle}, \tag{293}$$

and, up to the supermomentum conservation delta function, it is the same as in equation (283).

Exercise 6.2. Show the prefactor in equation (293) is crossing symmetric. Spinor properties may be found in appendix B and in various reviews [88, 89, 91].

It is not difficult to check that the amplitude in equation (290) obeys CK duality. Consider the duality relation in figure 20: because the triangle diagrams have vanishing numerators in equation (290), the duality requires the different box integrals to have an identical numerator, which follows from equation (293) and the integrals' numerators being crossing symmetric.

[48] It may also be obtained via the dimension-shifting relation [395] from the four-gluon superamplitude in $\mathcal{N}=4$ SYM theory (280),(282).

Exercise 6.3. Make the quarks massive. For the identical helicity case, the integral numerator is obtained with the replacement $\mu^4 \to (\mu^2 + m_q)^2$ [270] while the loop propagators become massive with mass m_q. Do the BCJ relations hold? What does the double-copy theory correspond to?

Consider now the double-copy construction with one of the two amplitude factors being equation (290) with $N_f = 0$. Taking the second amplitude to be the four-gluon superamplitude of $\mathcal{N} = 4$ SYM theory given in equations (280),(282) leads to an anomalous superamplitude in $\mathcal{N} = 4$ supergravity [260]:

$$\mathcal{M}^{1-\mathrm{loop}}(1,2,3,4)_{\mathcal{N}=4} = 2\left(\frac{\kappa}{2}\right)^4 \left(\frac{[12]\,[34]}{\langle 12\rangle\langle 34\rangle}\right)^2 \delta^{(8)}\left(\sum_{i=1}^{4} \lambda_i \eta_i^I\right) \times \left(I_4(s,t)\left[\mu^4\right] + I_4(s,u)\left[\mu^4\right] + I_4(t,u)\left[\mu^4\right]\right). \tag{294}$$

As outlined in section 4, this amplitude breaks the $U(1)$ duality symmetry of this theory [260] and is the amplitude-level manifestation of the duality anomaly identified in [263] from a Lagrangian perspective.

Another example is the double copy in which both amplitudes are given by equation (290). In D dimensions, the double copy of a gluon has a total for $(D-2)^2$ states, corresponding to a graviton ($D(D-3)/2$ states), antisymmetric tensor ($(D-2)(D-3)/2$ states) and dilaton (1 state). Taking both amplitudes in the double copy to be given by equation (290) with $N_f = 0$ leads to the four-graviton amplitude in a theory with a dilaton and antisymmetric tensor,

$$\mathcal{M}^{1-\mathrm{loop}}(1^+,2^+,3^+,4^+) = 4\left(\frac{\kappa}{2}\right)^4 \left(\frac{[12]\,[34]}{\langle 12\rangle\langle 34\rangle}\right)^2 \Big(I_4(s,t)\left[\mu^8\right] + I_4(s,u)\left[\mu^8\right] + I_4(t,u)\left[\mu^8\right]\Big). \tag{295}$$

The polarization vectors in the spinor-helicity basis used in equation (290) project out the dilaton and antisymmetric tensor asymptotic states from this amplitude. BCJ duality and the double copy for general helicity have been described in [12, 396]. For a theory with only gravitons and no anti-symmetric tensor or dilaton, the result for the identical helicity four-graviton amplitude is the same as in equation (295), except that the overall factor of 4 becomes a factor of 2. This can be proven by inserting graviton physical-state projectors into the unitarity cuts, as described in appendix C.

The integrals in the gauge-theory and gravity amplitudes in equations (290), (294) and (295),

$$I_4(s,t)\left[\mu^{4k}\right] = \int \frac{d^D\ell}{(2\pi)^D} \frac{\mu^{4k}}{\ell^2(\ell-p_1)^2(\ell-p_1-p_2)^2(\ell+p_4)^2}, \tag{296}$$

evaluate to

$$I_4(s,t)\left[\mu^4\right] = -\frac{i}{(4\pi)^2}\frac{1}{6} + \mathcal{O}(\epsilon), \quad I_4(s,t)\left[\mu^8\right] = -\frac{i}{(4\pi)^2}\frac{1}{840}\left(2s^2+2t^2+st\right). \tag{297}$$

Their finite values arise due to a cancellation of the $\mathcal{O}(\epsilon)$ numerator factors and $\mathcal{O}(\epsilon^{-1})$ IR divergences. From this perspective, the nonvanishing amplitude (290) may be interpreted as a self-duality anomaly [264]. The integrals (296) may also be interpreted in terms of higher-dimensional integrals [395].

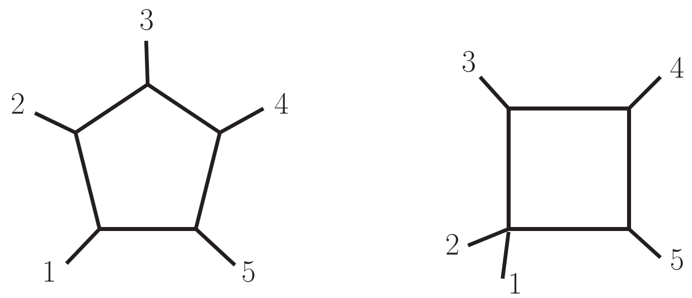


Figure 22. Pentagon and box integrals appearing in the $\mathcal{N}=4$ SYM five-point one-loop amplitudes. The complete set of such integrals is generated by permuting external legs and removing overcounts.

Exercise 6.4. Consider the double copy of amplitudes in QCD with $N_f>0$ flavors of quarks in the adjoint representation. Write down the spectrum of the resulting gravity theory. Construct the corresponding four-graviton amplitude. Would you expect that this theory is consistent quantum mechanically for any value of N_f? Answer the same questions if the $N_f>0$ flavors of quarks are in the fundamental representation.

As a more sophisticated example, consider the one-loop five-gluon amplitude. We will eventually restrict to the $\mathcal{N}=4$ SYM theory, but for now the discussion is quite general. We only need to discuss the maximally-helicity-violating (MHV) amplitude, as the only other nonvanishing one, the $\overline{\text{MHV}}$ amplitude, can be obtained by hermitian conjugation. Five-point amplitudes with other external states can be obtained through a suitable sequence of supersymmetry transformations. This amplitude was constructed in [163, 397] in a color-trace basis. Here we rearrange it slightly and write it in the structure-constant basis,

$$\mathcal{A}_5^{1-\text{loop}}(1,2,3,4,5)=g^5\sum_{S_5/(\mathbb{Z}_5\times\mathbb{Z}_2)}c_{12345}A_5^{1-\text{loop}}(1,2,3,4,5)\,, \tag{298}$$

where $A_5^{1-\text{loop}}$ on the right-hand side are the five-point color-ordered partial amplitudes. The sum runs over the distinct permutations of the external legs: this is the set of all 5! permutations, S_5, but with cyclic, $\mathbb{Z}_5$, and reflection symmetries, $\mathbb{Z}_2$, removed, leaving 12 distinct permutations. The color factor c_{12345} is the one obtained from the pentagon diagram shown in figure 22, with legs following the cyclic ordering, by dressing each vertex with an $\tilde{f}^{abc}$. This color decomposition holds for any gauge-theory amplitude with only adjoint-representation particles and can be reached by starting from a generic color decomposition in terms of products of structure constants and repeatedly using the Jacobi identity to favor structure constants with a single external color index.

Exercise 6.5. By starting from Feynman diagrams, apply color Jacobi identities to express all color factors in terms of those of pentagon diagrams. What is the generalization for an arbitrary number of external legs? (Feynman diagrams can helpful proving various properties, even if not useful for high-multiplicity explicit calculations.)

Exercise 6.6. Generalize equation (298) to include quarks in the fundamental representation in the loop. (See equation (290) at four points).

For $\mathcal{N}=4$ SYM theory, the color-ordered one-loop five-point amplitudes in equation (298) are [163, 397],

$$A^{1-\text{loop}}_{\mathcal{N}=4}(1,2,3,4,5) = \frac{1}{2}A^{\text{tree}}_5(1,2,3,4,5)\Big(s_{34}s_{45}I_4^{(12)345} + s_{45}s_{15}I_4^{1(23)45} + s_{12}s_{15}I_4^{12(34)5} + s_{12}s_{23}I_4^{123(45)} + s_{23}s_{34}I_4^{234(51)}\Big) + \mathcal{O}(\epsilon)\,, \tag{299}$$

where $A^{\text{tree}}_5(1,2,3,4,5)$ is the color-ordered MHV tree-level amplitude. We may obtain the entire one-loop five-point MHV superamplitude by replacing $A^{\text{tree}}_5(1,2,3,4,5)$ with the five-point tree-level MHV superamplitude in equation (465). The external kinematic invariants are $s_{ij}=(p_i+p_j)^2$. The $I_4^{abc(de)}$ are scalar box integrals where the legs in parenthesis connect to the same vertex, e.g. $I_4^{(12)345}$ is the box diagram in figure 22. This representation (299) of the amplitude does not manifestly satisfy the duality. An alternative representation of the MHV superamplitude, which manifests the duality between color and kinematics, is [4]:

$$\mathcal{A}^{1-\text{loop}}_{\mathcal{N}=4}(1,2,3,4,5) = g^5\left(\sum_{S_5/(\mathbb{Z}_5\times\mathbb{Z}_2)} c_{12345}n_{12345}I_5^{12345} + \sum_{S_5/\mathbb{Z}_2^2} c_{[12]345}n_{[12]345}\frac{1}{s_{12}}I_4^{(12)345}\right). \tag{300}$$

Each of the two sums runs over the distinct permutations of the external legs of the integrals. For I_5^{12345}, the set $S_5/(\mathbb{Z}_5\times\mathbb{Z}_2)$ denotes all permutations but with cyclic and reflection symmetries removed, leaving 12 distinct permutations. For $I_4^{(12)345}$ the set $S_5/\mathbb{Z}_2^2$ denotes all permutations but with the two symmetries of the one-mass box removed, leaving 30 distinct permutations. The pentagon numerator for this representation of the superamplitude is

$$n_{12345} = -\delta^{(8)}(Q)\frac{[12][23][34][45][51]}{4i\epsilon(1,2,3,4)}\,, \tag{301}$$

where $4i\epsilon(1,2,3,4) = 4i\epsilon_{\mu\nu\rho\sigma}k_1^\mu k_2^\nu k_3^\rho k_4^\sigma = [12]\langle 23\rangle[34]\langle 41\rangle - \langle 12\rangle[23]\langle 34\rangle[41]$. With this pentagon numerator, the box numerators that manifest the kinematic Jacobi relations illustrated in figure 23 are

$$n_{[12]345} = n_{12345} - n_{21345}\,. \tag{302}$$

Other box numerators are obtained by relabeling. It is not difficult to see that the diagrams with triangle or bubble integrals have vanishing numerators. For example, the numerator of the triangle diagram with momenta p_1+p_2 at one vertex and p_4+p_5 at another is

$$\begin{aligned} n_{[12]345} - n_{[12]354} &= n_{12345} - n_{21345} - n_{12354} + n_{21354} \\ &= -\frac{\delta^{(8)}(Q)}{4i\epsilon(1,2,3,4)}\{[12][23][34][45][51] + [21][13][34][45][52] \\ &\quad + [12][23][35][54][41] + [21][13][35][54][42]\}\,, \end{aligned} \tag{303}$$

where we used momentum conservation to relate all Levi-Civita symbols contracted with four external momenta. Upon use of the Schouten identities,

$$[51][23] - [52][13] = [12][35]\ , \quad [23][41] - [12][34] = [13][42]\,, \tag{304}$$

the term in brackets in equation (303) vanishes, so

$$n_{[12]345} - n_{[12]354} = n_{12345} - n_{21345} - n_{12354} + n_{21354} = 0\,. \tag{305}$$

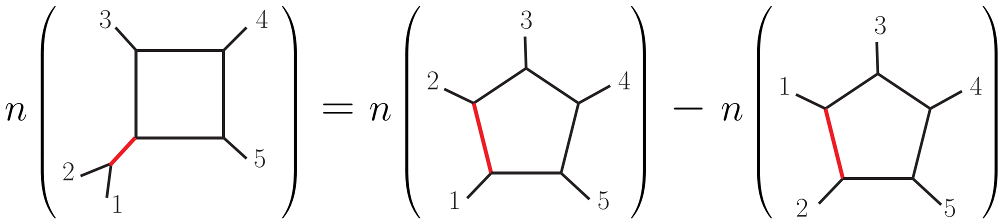


Figure 23. A BCJ kinematic numerator relation between a diagram containing a box integral and two pentagon diagrams. The shaded (red) line differs between the diagrams, but the others are identical.

The first of the identities (304) is used to combine the first two terms in equation (303) and the second identity shows that the remaining term cancel.

Exercise 6.7. Show that all kinematic numerator relations hold for the amplitude given in equation (300).

A nice feature of this representation is that the numerator factors of both the pentagon and box integrals do not depend on loop momentum. This greatly simplifies the construction of the corresponding supergravity amplitudes.

Given that the duality holds for the representation (300) of the five-point one-loop MHV $\mathcal{N}=4$ SYM superamplitude, we can immediately obtain the corresponding $\mathcal{N}=8$ amplitude. We replace the color factors with a numerator factor (44),

$$c_{12345} \to n_{12345}\,, \qquad\qquad c_{[12]345} \to n_{[12]345}\,, \tag{306}$$

as well as the gauge coupling with the gravitational one. The resulting five-graviton one-loop MHV superamplitude in $\mathcal{N}=8$ supergravity reads (45)

$$\mathcal{M}^{\text{1-loop}}_{\mathcal{N}=8}(1,2,3,4,5) = \left(\frac{\kappa}{2}\right)^5 \left(\sum_{S_5/(\mathbb{Z}_5\times\mathbb{Z}_2)} (n_{12345})^2 I_5^{12345} + \sum_{S_5/\mathbb{Z}_2^2} \left(n_{[12]345}\right)^2 \frac{1}{s_{12}} I_4^{(12)345}\right), \tag{307}$$

where the sums run over the same permutations as in equation (300) and, as discussed in section 4, the $\delta^{(16)}(Q)$ should be understood as containing eight different η parameters for each external particle.

The scalar pentagon integral and the one external-mass box integral have been computed in [394]. We include them here for convenience:

$$\begin{aligned} I_4^{(12)345} = &-\frac{2ic_\Gamma}{s_{34}s_{45}}\left\{-\frac{1}{\epsilon^2}\left[(-s_{34})^{-\epsilon} + (-s_{45})^{-\epsilon} - \left(-s_{12}^2\right)^{-\epsilon}\right]\right. \\ &\left.+\mathrm{Li}_2\left(1-\frac{s_{12}}{s_{34}}\right) + \mathrm{Li}_2\left(1-\frac{s_{12}}{s_{45}}\right) + \frac{1}{2}\ln^2\left(\frac{s_{34}}{s_{45}}\right) + \frac{\pi^2}{6}\right\} + \mathcal{O}(\epsilon)\,, \end{aligned} \tag{308}$$

$$\begin{aligned} I_5^{12345} = &\sum_{\mathbb{Z}_5} \frac{-ic_\Gamma(-s_{51})^\epsilon(-s_{12})^\epsilon}{(-s_{23})^{1+\epsilon}(-s_{34})^{1+\epsilon}(-s_{45})^{1+\epsilon}}\left[\frac{1}{\epsilon^2} + 2\,\mathrm{Li}_2\left(1-\frac{s_{23}}{s_{51}}\right) + 2\,\mathrm{Li}_2\left(1-\frac{s_{45}}{s_{12}}\right) - \frac{\pi^2}{6}\right] \\ &+\mathcal{O}(\epsilon)\,, \end{aligned} \tag{309}$$

where $\mathrm{Li}_2(x)$ is the dilogarithm function and c_Γ is defined in equation (288).

Exercise 6.8. Show that the double copy of the one-loop five-point amplitude where one copy is of an MHV amplitude and the second an $\overline{\mathrm{MHV}}$ amplitude vanishes. The $\overline{\mathrm{MHV}}$ amplitude is obtained from the MHV one by parity which amounts to replacing $\langle ab\rangle \leftrightarrow [ab]$ and flipping the overall sign of the amplitude. Do you expect a similar property to hold at n points or at higher loops?

As discussed above, the double-copy construction works even if the duality is manifest in only one gauge-theory factor. Starting with the color decomposition in equation (298) and using the fact that for the one-loop five-point $\mathcal{N}=4$ SYM amplitude the pentagon numerators are independent of loop momentum, we immediately obtain five-point superamplitudes for $(\mathcal{N}+4)$-extended supergravities. By taking the second copy to be any pure SYM theory, with color-ordered one-loop five-point amplitudes $A_{\mathcal{N}}^{1-\mathrm{loop}}(1,2,3,4,5)$, we find

$$\mathcal{M}_{\mathcal{N}}^{1-\mathrm{loop}}(1,2,3,4,5) = \left(\frac{\kappa}{2}\right)^5 \sum_{S_5/(\mathbb{Z}_5\times\mathbb{Z}_2)} n_{12345} A_{\mathcal{N}}^{1-\mathrm{loop}}(1,2,3,4,5)\,. \tag{310}$$

Here n_{12345} is given in equation (301) and the sums run, as in the case of the $\mathcal{N}=4$ amplitude, over all the permutations which are not related to each other by cyclic permutations or reflections.

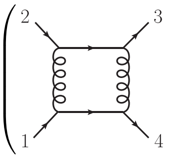
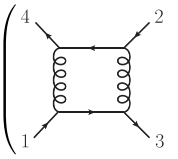
6.2. One-loop examples of BCJ duality: SYM theories with reduced supersymmetry

Gauge theories with reduced supersymmetry provide an opportunity to discuss the construction of duality-satisfying (loop-level) scattering amplitudes with fields in representations other than the adjoint. A simple example, which we will review here in some detail, is the one-loop four-matter-field superamplitude in $\mathcal{N}=2$ SYM theory with a single hypermultiplet in a complex representation $\mathcal{R}$ [30, 294]. The color factors c_j in equation (277) are now constructed by dressing every vertex of every diagram with a gauge-group generator in the appropriate representation. This more complicated color structure is a consequence of reduced supersymmetry, which allows for matter fields in non-adjoint representations. To keep supersymmetry manifest, we organize the hypermultiplet asymptotic states as on-shell superfields and their CPT-conjugates, which are treated as distinct:

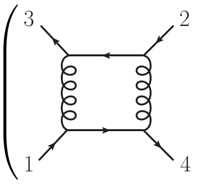
$$\Phi_{\mathcal{N}=2\hat{\alpha}} = \chi_{+\hat{\alpha}} + \eta^i\varphi_{i\hat{\alpha}} + \eta^1\eta^2\tilde{\chi}_{-\hat{\alpha}} \qquad \overline{\Phi}_{\mathcal{N}=2}{}^{\hat{\alpha}} = \tilde{\chi}_+^{\hat{\alpha}} + \eta^i\overline{\varphi}_i^{\hat{\alpha}} + \eta^1\eta^2\chi_-^{\hat{\alpha}}. \tag{311}$$

The lower and upper $\hat{\alpha}$ is the $\mathcal{R}$ and $\bar{\mathcal{R}}$ representation indices, respectively. As outlined in section 5.2.2, such superfields with reduced supersymmetry can in principle be obtained from the ones of $\mathcal{N}=4$ SYM theory by an orbifold truncation.

At one loop, a duality-satisfying representation of the four-hypermultiplet superamplitude can be constructed in terms of two master numerators, which can be chosen to belong to two box diagrams. Adopting the standard notation for theories with matter (super)fields, we denote the adjoint vector multiplet with a curly line and the complex-representation hypermultiplet with a solid line with an arrow. The master numerator factors and the corresponding diagrams are [30, 294]:

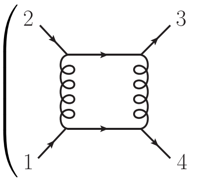
$n\Bigg($  $\Bigg) = \dfrac{s^2}{\langle 12\rangle\langle 34\rangle}\delta^{(4)}\big(Q\big)\,,\quad n\Bigg($  $\Bigg) = -\dfrac{st}{\langle 12\rangle\langle 34\rangle}\delta^{(4)}\big(Q\big)\,,$ (312)

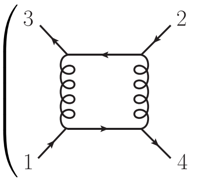
where $\delta^{(4)}\big(Q\big) = \delta^{(4)}\big(\sum_n \eta_n^i \lambda_n\big)$. The other box integrals can then be obtained by permutation, keeping in mind that the overall superamplitude possesses a $\mathbb{Z}_2 \times \mathbb{Z}_2$ Fermi symmetry under the exchange of hypermultiplet superfields. For example, a third box-integral numerator is

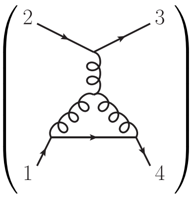
$n\Bigg($ 

$\Bigg) = -\dfrac{su}{\langle 12\rangle\langle 34\rangle}\delta^{(4)}\big(Q\big)\,.$ (313)

Numerators for the triangle and bubble diagrams can be obtained via the kinematic numerator relations. They can be organized in two distinct sets: (1) those that mirror relations between color factors which are a consequence of the defining commutation relations of the color Lie algebra and (2) those that corresponding to color relations that hold only for certain groups and representations but are nonetheless required for the consistency of the double copy of a hypermultiplet with a vector multiplet. An example of numerator relations from the first group is

$n\Bigg($ 

$\Bigg) - n\Bigg($ 

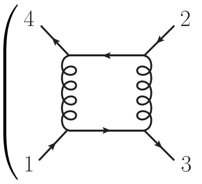
$\Bigg) = n\Bigg($ 

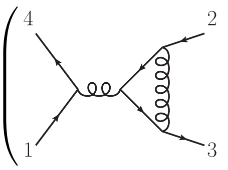
$\Bigg)\,.$ (314)

From the double-copy perspective, following the argument presented in section 2.5, these relations are required for obtaining gravity amplitudes invariant under linearized diffeomorphisms. An example of color relations that hold only for certain groups and representations is

$$T^{\hat a}{}_{\hat\alpha}{}^{\hat\gamma} T^{\hat a}{}_{\hat\beta}{}^{\hat\delta} = T^{\hat a}{}_{\hat\alpha}{}^{\hat\delta} T^{\hat a}{}_{\hat\beta}{}^{\hat\gamma}\,. \tag{315}$$

The corresponding numerator relations include, for example,

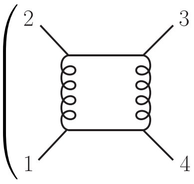
$n\Bigg($ 

$\Bigg) = n\Bigg($ 

$\Bigg)\,.$ (316)

While these color relations are satisfied only for certain choices of gauge group and representations, the fact that the form (276) is independent of such choices suggests that one may choose, as we do here, to always impose the corresponding numerator relations. One may easily convince oneself that these numerator relations are required by consistency of the double-copy construction in case the hypermultiplet fields are combined with spin one fields in the conjugate matter representation. In other cases they may be regarded as 'bonus' relations; it is not clear *a priori* that there exist solutions to the numerator relations in the second group even when solutions to the numerator relations in the first group do.

In section 5.3.3, we have reviewed the double-copy construction for homogeneous supergravities, and showed that it reproduces the existing classification of such theories. An important ingredient of the construction are matter fields in pseudo-real representations. It is therefore instructive to see how our one-loop numerators described above are modified in this case. To enforce the pseudo-reality of the gauge group representation (i.e. the equivalence of the upper and lower $\hat{\alpha}$ indices in equation (310)), the on-shell superfields $\Phi_{\mathcal{N}=2}$ and $\overline{\Phi}_{\mathcal{N}=2}$ are identified. Consequently, the superamplitude needs to acquire a complete Fermi symmetry for all of its external legs. Drawing from this observation, the numerators for a theory with pseudo-real half-hypermultiplets can be obtained as the unique set of numerators which are both invariant under the permutation of all external legs and reduce to the numerators for the complex case whenever the corresponding color factors are nonzero. More concretely, in the pseudo-real case we have only one master numerator:

$n\Bigg($  $\Bigg) = \frac{s^2}{\langle 12\rangle\langle 34\rangle}\delta^{(4)}\big(Q\big)\,,$ (317)

and all the other numerators are obtained either from permutation symmetry or from the numerator relations (313) and (315). For half-hypermultiplets in pseudo-real representations, solid lines no longer carry an arrow since the matter half-hypermultiplets are CPT-self-conjugate.

Exercise 6.9. Given the master numerator (316), use numerator relations to generate all nonzero numerators (up to permutation symmetry).

We emphasize that, as in the case of the one-loop four- and five-point superamplitudes of $\mathcal{N}=4$ SYM theory, the duality-satisfying kinematic numerators of the superamplitude reviewed here are independent of the loop momentum. Consequently, the physical properties of the double-copy supergravity theory can be directly related to properties of the other gauge theory entering the construction. Consider, for example, the construction for homogeneous Maxwell-Einstein supergravities explained in section 5.3.3. We can relate the one-loop divergences of supergravity amplitudes with four vector superfields constructed as hypermultiplet $\times$ fermion to a linear combination of various parts of the one-loop beta function of the non-supersymmetric gauge theory,

$$\mathcal{M}^{\text{1-loop}}\Big|_{\text{div}} = \frac{-i}{(4\pi)^2}\frac{s\,\delta^{(4)}(Q)}{\langle 12\rangle\langle 34\rangle}\left(\frac{\kappa}{2}\right)^4\left\{sA^{\text{tree}}_{s,\phi}\left(\beta_\phi\big|_{T(G)} - \frac{\beta_\phi}{2}\Big|_{T(R)}\right)\right.$$
$$\left. + sA^{\text{tree}}_{s,A}\left(\beta_A\big|_{T(G)} - \frac{\beta_A}{2}\Big|_{T(R)}\right)\right\}\frac{1}{\epsilon} + \text{perms}\,. \tag{318}$$

Here β_ϕ, β_A are the beta-functions for the gauge coupling and the Yukawa interactions. We use the notation $\beta\big|_{T(G),T(R)}$ to label the parts of the relevant beta functions that are proportional to the index of the adjoint, $T(G)$,and pseudo-real, $T(R)$, matter representations. $A^{\text{tree}}_{s,A}$ and $A^{\text{tree}}_{s,\phi}$ are, respectively, the s-channel gluon and scalar exchange parts of the gauge theory tree-level amplitudes. Finally, it should be noted that compact expressions for two-loop amplitudes for $\mathcal{N}=2$ gauge theories with matter can be found in [398].

Exercise 6.10. Use the result of Exercise 7.9 to verify equation (318).

6.3. Two-loop examples

If the duality between color and kinematics holds at tree-level in D dimensions, then it also holds on all D-dimensional generalized cuts that decompose a loop amplitude into a sum of

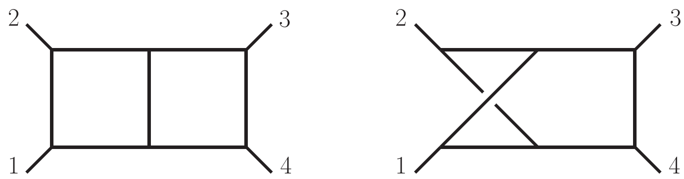


Figure 24. Diagrams for the two-loop integrals appearing in the two-loop four-point $\mathcal{N}=4$ and $\mathcal{N}=8$ supergravity amplitudes.

products of tree amplitudes. Thus, barring anomalies, it is expected to hold beyond one-loop level. As an illustrative example, consider the two-loop four-point amplitude of $\mathcal{N}=4$ SYM theory. This amplitude, originally constructed in [194, 399], is

$$\mathcal{A}_4^{2-\text{loop}}(1,2,3,4) = -g^6 st A_4^{\text{tree}}(1,2,3,4)\Big(c^{\text{P}}_{1234}\, s I_4^{2-\text{loop,P}}(s,t) + c^{\text{P}}_{3421}\, s I_4^{2-\text{loop,P}}(s,u) + c^{\text{NP}}_{1234}\, s I_4^{2-\text{loop,NP}}(s,t) + c^{\text{NP}}_{3421}\, s I_4^{2-\text{loop,NP}}(s,u) + \text{cyclic}\Big)\,, \quad (319)$$

where '+ cyclic' indicates that one should add the two cyclic permutations of $(2,3,4)$. The integrals correspond to the scalar planar and nonplanar double-box diagrams shown in figure 24. As at one loop, the color factor of each diagram is obtained by dressing each cubic vertex with an $\tilde{f}^{abc}$ factor.

As the diagrams appearing in the amplitude are already cubic, we can read off the kinematic numerators for each diagram. They are:

$$\begin{aligned}
n^{\text{P}}_{1234} &= n^{\text{NP}}_{1234} = i s^2 t A_4^{\text{tree}}(1,2,3,4)\,, & n^{\text{P}}_{3412} &= n^{\text{NP}}_{3412} = i s^2 t A_4^{\text{tree}}(1,2,3,4)\,,\\
n^{\text{P}}_{1342} &= n^{\text{NP}}_{1342} = i ust A_4^{\text{tree}}(1,2,3,4)\,, & n^{\text{P}}_{4213} &= n^{\text{NP}}_{4213} = i ust A_4^{\text{tree}}(1,2,3,4)\,,\\
n^{\text{P}}_{1423} &= n^{\text{NP}}_{1423} = i s t^2 A_4^{\text{tree}}(1,2,3,4)\,, & n^{\text{P}}_{2314} &= n^{\text{NP}}_{2314} = i s t^2 A_4^{\text{tree}}(1,2,3,4)\,.
\end{aligned} \quad (320)$$

The factor $st A_4^{\text{tree}}(1,2,3,4)$, being crossing symmetric, remains as overall factor for the complete amplitude, after all cyclic permutations of $(2,3,4)$ are added.

Because of the limited set of nonvanishing diagrams, it is straightforward to check that this amplitude satisfies all duality relations. Three of them are shown in figure 25. The complete set may be obtained by starting with the diagrams in figure 24 and systematically generating the duality relations.

Following the double-copy prescription (44), we obtain the corresponding $\mathcal{N}=8$ supergravity amplitude by replacing the color factor with a numerator factor,

$$c^{\text{P}}_{1234} \to i s^2 t A^{\text{tree}}(1,2,3,4)\,, \qquad c^{\text{NP}}_{1234} \to i s^2 t A^{\text{tree}}(1,2,3,4)\,, \quad (321)$$

including relabelings and then swapping the gauge coupling for the gravitational one. Indeed, this gives the correct $\mathcal{N}=8$ supergravity amplitude, as first noted in [194] which also verified it against the direct construction from unitarity cuts.

As mentioned in section 2, generalized gauge invariance implies that only one of the two copies must be in a form manifestly satisfying the duality (41); for the second copy, such a form should exist but its use is not required. The color Jacobi identity allows us to express any four-point color factor of an adjoint representation in terms of the ones in figure 24 [180]. If the duality and double-copy properties hold and because of the independence of the momentum of the $\mathcal{N}=4$ SYM numerator factors (320), it is possible to obtain integrated $\mathcal{N}\geqslant 4$ supergravity amplitudes starting from $\mathcal{N}\leqslant 4$ SYM theory and applying the replacement rule (321) [32].

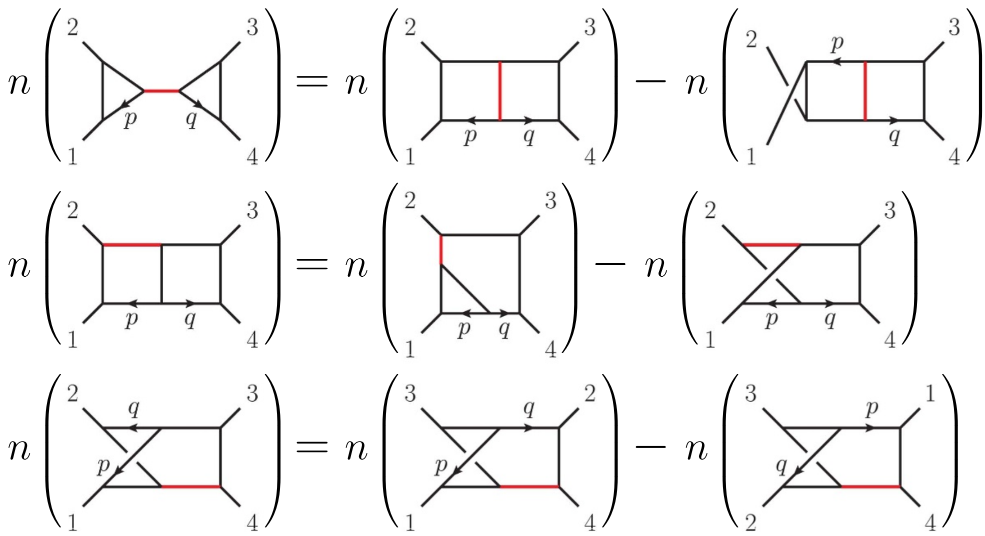


Figure 25. Examples of BCJ kinematic numerator relations at two loops.

A further interesting and nontrivial example is the five-point amplitude in $\mathcal{N} = 4$ SYM [4]. Due to the high degree of supersymmetry, one can express the amplitude in terms of only six contributing diagrams shown in figure 26. To be concise, we will only quote the duality-satisfying numerator of diagram (a); it is

$$n^{(\mathrm{a})}_{12345}(p,q) = \frac{1}{4}\left(\gamma_{12}\left(2s_{45} - s_{12} + \tau_{2p} - \tau_{1p}\right) + \gamma_{23}\left(s_{45} + 2s_{12} - \tau_{2p} + \tau_{3p}\right)\right.$$
$$\left. + 2\gamma_{45}\left(\tau_{5p} - \tau_{4p}\right) + \gamma_{13}\left(s_{12} + s_{45} - \tau_{1p} + \tau_{3p}\right)\right), \qquad (322)$$

where the two independent loop momenta are called p and q. The Lorentz invariants are $\tau_{ip} = 2p_i \cdot p$, $\tau_{iq} = 2p_i \cdot q$ and $s_{ij} = (p_i + p_j)^2$. The external state dependence for the MHV amplitude is captured by the γ_{ij}, where

$$\gamma_{12} \equiv n_{[12]345} = \delta^{(8)}(Q)\,\frac{[12]^2\,[34]\,[45]\,[53]}{4i\,\epsilon(1,2,3,4)} \qquad (323)$$

is the one-loop box numerator given in equation (302), and the other γ_{ij} are given by S_5 permutations of this expression. Note that the γ_{ij} satisfy the relations

$$\gamma_{ij} = -\gamma_{ji}\,, \qquad \sum_{i=1}^{5} \gamma_{ij} = 0\,, \qquad (324)$$

from which it follows that there are only six independent variables of this type. The diagram numerators of the $\overline{\mathrm{MHV}}$ amplitude are obtained by replacing γ_{ij} by their CPT conjugates.

Exercise 6.11. Show that the kinematic numerators corresponding to diagrams (b)–(f) in figure 26 can be obtained from $n^{(\mathrm{a})}_{12345}(p,q)$ using kinematic Jacobi relations. Which numerators happens to be independent of loop momenta? Which numerators are identical to each other (due to the Jacobi relation collapsing to a two-term identity)?

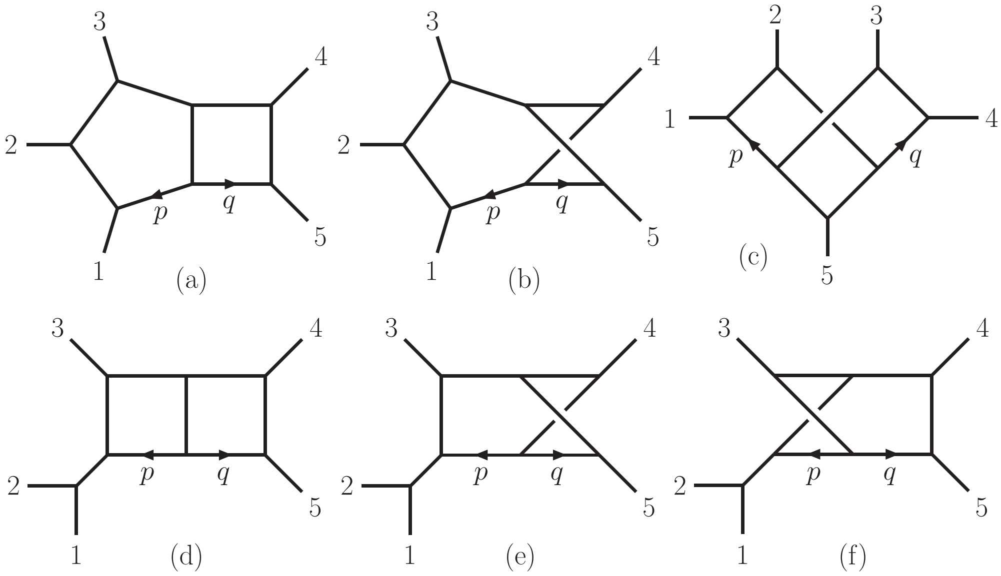


Figure 26. The six nonzero diagrams that contribute to two-loop five-point amplitude in $\mathcal{N}=4$ SYM and $\mathcal{N}=8$ supergravity.

The two-loop $\mathcal{N}=4$ SYM amplitude is given by the sum over the six diagrams (a)–(f) in figure 26, together with the sum over the $\mathcal{S}_5$ permutations over the external legs,

$$\mathcal{A}_5^{2-\text{loop}} = i g^7 \sum_{\mathcal{S}_5} \sum_{j\in\{\text{a},\ldots\text{f}\}} \int \frac{d^D p\, d^D q}{(2\pi)^{2D}} \frac{1}{S_j} \frac{c_{12345}^{(j)} n_{12345}^{(j)}(p,q)}{\prod_{\alpha_j} p_{\alpha_j}^2}\,. \tag{325}$$

The corresponding $\mathcal{N}=8$ supergravity amplitude is obtained by the double-copy replacements $c_{12345}^{(j)} \to n_{12345}^{(j)}(p,q)$ and $g \to \kappa/2$.

Exercise 6.12. By inspecting the diagrams in figure 26, compute the symmetry factors S_j that appear in equation (325).

The two-loop five-point amplitudes of both $\mathcal{N}=4$ SYM and $\mathcal{N}=8$ supergravity, as presented above, were integrated in $D=4-2\epsilon$ dimensions in references [327, 400–402].

6.4. Three-loop example

So far, we illustrated various one- and two-loop amplitudes that manifest the color-kinematics duality. To be more concrete, in this subsection we go through in detail how to construct duality-satisfying amplitudes when the system of numerators is quite large. As a sophisticated example—though still quite manageable—consider the three-loop four-point amplitude of the $\mathcal{N}=4$ SYM and $\mathcal{N}=8$ supergravity theories [2].

Apart from the duality and unitarity constraints, it is beneficial to systematically impose various other constraints which become more important as the complexity of the problem increases. Although not required, such auxiliary constraints, when appropriately chosen, can greatly facilitate the construction. If a constraint is too strong and leads to an inconsistency with unitarity, then one may relax or modify it as needed. This strategy is especially effective for theories with high degrees of supersymmetry, because of their restricted power counting. For the three-loop four-point $\mathcal{N}=4$ SYM amplitudes, a natural set of constraints is as follows.

(i) One-loop tadpole, bubble and triangle subdiagrams do not appear in any diagram [222, 403, 404].
(ii) A one-loop n-gon subdiagram carries no more than $n-4$ powers of loop momentum for that loop.
(iii) After extracting an overall factor of stA_4^{tree}, the numerators are polynomials in D-dimensional Lorentz scalar products of the independent loop and external momenta.
(iv) Numerators carry the same relabeling symmetries as the diagrams (cf discussion in section 3).

In general, the choice of auxiliary constraints depends on the problem at hand. For example, the third constraint above is specific to the four-point amplitude, and should be modified for higher-point amplitudes because of their more complicated external-state structure. As described in the previous subsection, a relatively simple generalization has been found for the five-point (super)amplitude [4], involving prefactors that are proportional [161, 405, 406] to linear combinations of five-point color-ordered tree-amplitudes. For amplitudes in less supersymmetric theories, all but the fourth condition must also be relaxed, because their power counting is such that one-loop triangle and bubble subdiagrams do appear; this is related to e.g. the running of their couplings. The above constraints also work well for the four-loop four-point amplitudes of $\mathcal{N}=4$ SYM [6], but fail at five loops. A procedure which works for this case is described in section 8.

Because the duality imposes stringent relations between diagrams' numerators, a remarkably small subset of generalized unitarity cuts is then sufficient to completely determine the integrand. Of course, to confirm that it is correct, it is necessary to verify that it reproduces correctly a spanning set of unitarity cuts that fully determine the amplitude. Quite generally, one expects that a problem with a generalized cut can be addressed by relaxing some of the auxiliary constraints.

Let us return now to the three-loop four-point amplitudes of $\mathcal{N}=4$ SYM theory and illustrate these ideas. A straightforward enumeration shows that there are 17 distinct cubic diagrams with three loops and four external legs, which do not have one-loop triangle, bubble or tadpole subdiagrams. It turns out that the twelve diagrams shown in figure 27 are sufficient for finding a solution to the duality and unitarity cut constraints, as shown in [2]. Had we kept all 17 diagrams, the construction would be slightly more involved, with the result that the numerators of the additional diagrams vanish identically.

The four-point amplitudes of $\mathcal{N}=4$ SYM theory are special. Applying the third condition above we write the numerator as

$$n^{(x)} = -i\, stA_4^{\text{tree}}(1,2,3,4)\, N^{(x)}\,, \tag{326}$$

where (x) refers to the label for each diagram in figure 27 and $N^{(x)}$ are scalar functions which depend on three independent external momenta, labeled by p_1, p_2, p_3, and on (at most) three independent loop momenta, labeled by ℓ_5, ℓ_6, ℓ_7,

$$N^{(x)} \equiv N^{(x)}(p_1, p_2, p_3, \ell_5, \ell_6, \ell_7)\,. \tag{327}$$

The coefficient $stA_4^{\text{tree}}(1,2,3,4)$ is fully crossing symmetric, as noted in equation (283).

Next, consider the duality relations. We need to discuss first those that allow us to express the complete set of numerators $N^{(x)}$ in terms of a small subset—the master numerators. Some of them are shown in figure 28. The remaining relations are subsequently verified once the

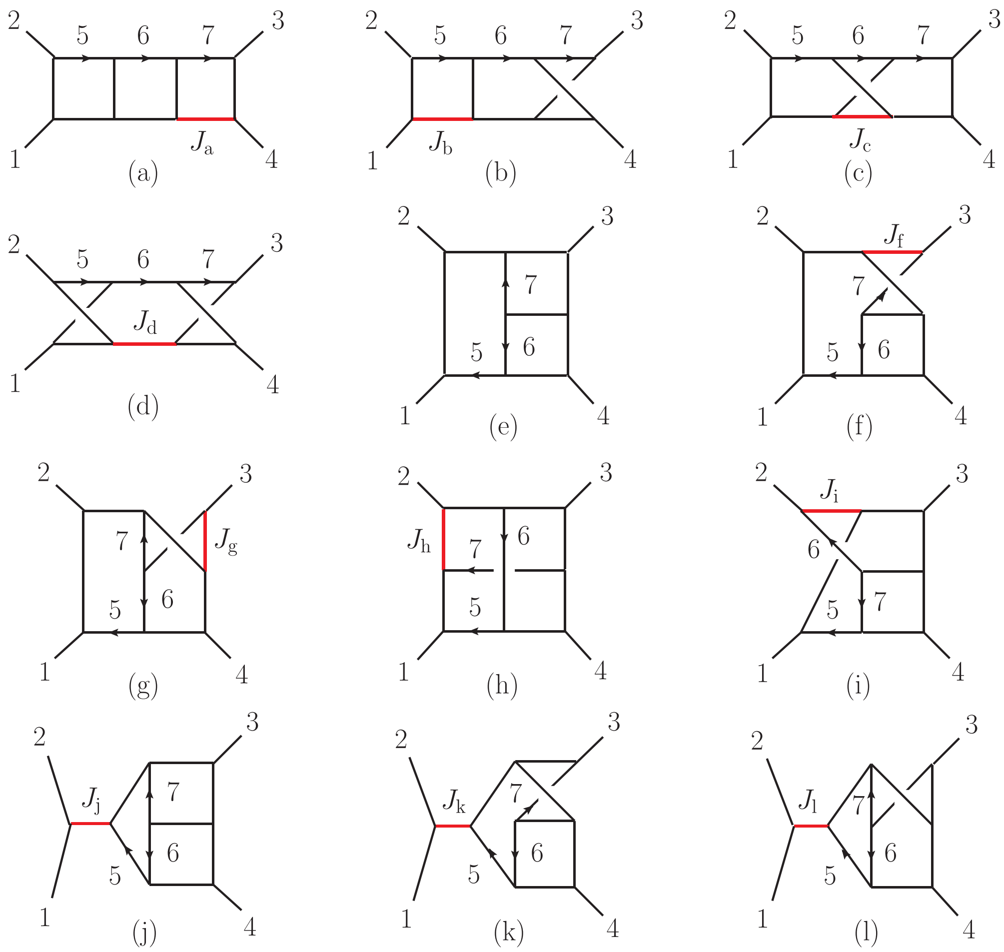


Figure 27. The diagrams for constructing the $\mathcal{N}=4$ SYM and $\mathcal{N}=8$ supergravity three-loop four-point amplitudes. The shaded (red) lines indicate the application of the duality relation. The external momenta are outgoing and the arrows indicate the directions of the labeled loop momenta. Diagram (e) is the master diagram.

former are solved together with the constraints imposed by the unitarity cuts. For the three-loop four-point $\mathcal{N}=4$ SYM amplitude, a simple restricted set of duality relations is [6, 196]:

$$
\begin{aligned}
N^{(a)} &= N^{(b)}\left(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7\right) ,\\
N^{(b)} &= N^{(d)}\left(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7\right) ,\\
N^{(c)} &= N^{(a)}\left(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7\right) ,\\
N^{(d)} &= N^{(h)}\left(p_3,p_1,p_2,\ell_7,\ell_6,p_{1,3}-\ell_5+\ell_6-\ell_7\right)\\
&\quad + N^{(h)}\left(p_3,p_2,p_1,\ell_7,\ell_6,p_{2,3}+\ell_5-\ell_7\right) ,\\
N^{(f)} &= N^{(e)}\left(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7\right) ,\\
N^{(g)} &= N^{(e)}\left(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7\right) ,
\end{aligned}
$$

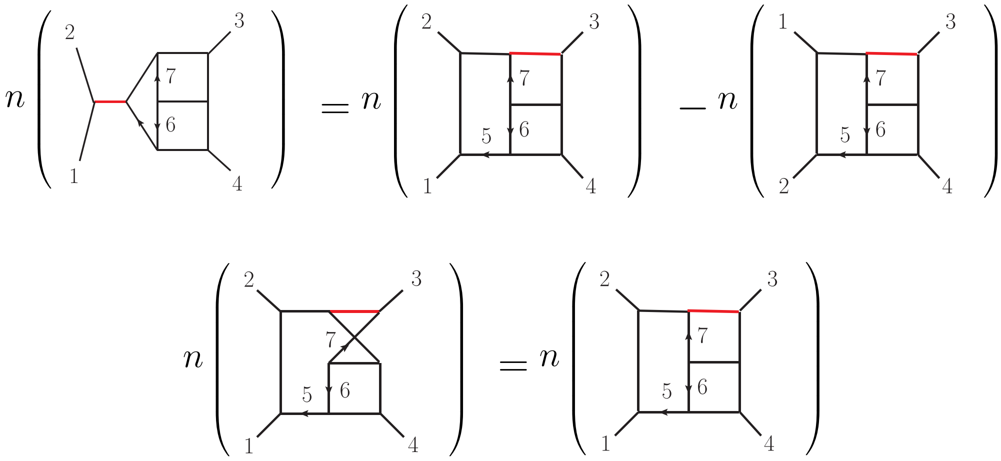


Figure 28. Examples of a BCJ kinematic numerator relation at three loops for $\mathcal{N}=4$ SYM theory. In the two term relations one of the three numerators in a Jacobi triplet of diagrams vanishes.

$$\begin{aligned}
N^{(\mathrm{h})} &= -N^{(\mathrm{g})}\,(p_1,p_2,p_3,\ell_5,\ell_6,p_{1,2}-\ell_5-\ell_7)\\
&\quad -N^{(\mathrm{i})}\,(p_4,p_3,p_2,\ell_6-\ell_5,\ell_5-\ell_6+\ell_7-p_{1,2},\ell_6)\,,\\
N^{(\mathrm{i})} &= N^{(\mathrm{e})}\,(p_1,p_2,p_3,\ell_5,\ell_7,\ell_6)-N^{(\mathrm{e})}\,(p_3,p_2,p_1,-p_4-\ell_5-\ell_6,-\ell_6-\ell_7,\ell_6)\,,\\
N^{(\mathrm{j})} &= N^{(\mathrm{e})}\,(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7)-N^{(\mathrm{e})}\,(p_2,p_1,p_3,\ell_5,\ell_6,\ell_7)\,,\\
N^{(\mathrm{k})} &= N^{(\mathrm{f})}\,(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7)-N^{(\mathrm{f})}\,(p_2,p_1,p_3,\ell_5,\ell_6,\ell_7)\,,\\
N^{(\mathrm{l})} &= N^{(\mathrm{g})}\,(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7)-N^{(\mathrm{g})}\,(p_2,p_1,p_3,\ell_5,\ell_6,\ell_7)\,,
\end{aligned} \tag{328}$$

where $p_{i,j}\equiv p_i+p_j$. To simplify the notation, we have suppressed the canonical arguments $(p_1,p_2,p_3,\ell_5,\ell_6,\ell_7)$ of the numerators on the left-hand side of the equation (328). Each relation specifying an $N^{(x)}$ is generated by considering the kinematic Jacobi relations dual to the color Jacobi relations corresponding to the shaded (red) line and labeled J_x in figure 27. In general, duality relations relate triplets of numerators; if however one of the diagrams is not present in figure 27, e.g. because it has a one-loop triangle subdiagram, then we obtain a two-term relation. Five of the equations above are of this type and they result in pairs of numerators being equal.

The system (328) can be used to express any kinematic numerator factor as a combination of the numerator $N^{(\mathrm{e})}$ with various different arguments. Thus, diagram (e) can be taken as the sole master diagram. This is a convenient choice, but not the only possible one; for example, either diagram (f) or (g) can also be used as a single master diagram. None of the remaining nine diagrams, however, can act alone as a master diagram.

The numerator factor of diagram (e) is constructed such that the unitarity cuts are satisfied simultaneously with the duality constraints. An expression that satisfies the maximal cuts is given by the so-called 'rung-rule' numerator [399],

$$N^{(\mathrm{e})}_{\mathrm{rr}} = s\,(\ell_5+p_4)^2\,, \tag{329}$$

which follows from the general features of iterated two-particle cuts.

We wish to find a modification $N^{(e)}_{rr} \to N^{(e)}$ such that all the other numerators determined from it via equation (328) are consistent with the unitarity cuts. We start by requiring that the maximal-cut of diagram (e) is correct (see appendix C.3 for a description of the maximal cuts), and that the auxiliary constraints above are satisfied. That is, the departure from $N^{(e)}_{rr}$ vanishes on the maximal cut, the numerator $N^{(e)}$ has mass dimension four and possesses the symmetry of the diagram; no loop momentum for any box subdiagram in (e) appears in it (ruling out ℓ_6 and ℓ_7), and $N^{(e)}$ is at most quadratic in the pentagon loop momenta ℓ_5. The last condition is a little weaker than the second auxiliary condition listed earlier, which demands linearity in ℓ_5; we relax it slightly to make it easier to find deformations that vanish on maximal cuts, and impose later that the ℓ_5^2 terms cancel out. The symmetry condition implies that $N^{(e)}$ is invariant under

$$\{p_1 \leftrightarrow p_2,\, p_3 \leftrightarrow p_4,\, \ell_5 \to p_1 + p_2 - \ell_5\}\,. \tag{330}$$

The most general polynomial consistent with these constraints is

$$\begin{aligned} N^{(e)} &= s\,(\ell_5 + p_4)^2 + (\alpha s + \beta t)\,\ell_5^2 + (\gamma s + \delta t)\,(\ell_5 - p_1)^2 \\ &\quad + (\alpha s + \beta t)\,(\ell_5 - p_1 - p_2)^2\,, \end{aligned} \tag{331}$$

where the four parameters $\alpha, \beta, \gamma, \delta$ are to be determined by further constraints. All added terms are proportional to inverse propagators and therefore vanish on the maximal cut. Thus, given that equation (329) is consistent with the maximal cuts, so is equation (331).

The second auxiliary constraint above demands that the numerator of a pentagon subdiagram be at most linear in the corresponding loop momentum, ℓ_5, not quadratic as assumed above. Therefore we impose that the coefficient of ℓ_5^2 in equation (331) vanishes. This yields the relation $\gamma = -1 - 2\alpha$ and $\delta = -2\beta$, which simplifies equation (331) to

$$N^{(e)} = s\,(\tau_{45} + \tau_{15}) + (\alpha s + \beta t)\,(s + \tau_{15} - \tau_{25})\,, \tag{332}$$

where we use the notation,

$$\tau_{ij} \equiv 2p_i \cdot \ell_j\,, \quad (i \leqslant 4, j \geqslant 5)\,. \tag{333}$$

We are therefore left with two undetermined parameters, α and β.

We determine the remaining parameters by imposing that the numerators of other diagrams determined through equation (328) are consistent with the auxiliary constraints and unitarity cuts. A convenient starting point is the numerator of diagram (j), $N^{(j)}$, which is determined in terms of $N^{(e)}$ by the 9th duality constraint in equation (328). Inserting equation (332) into this relation leads to

$$N^{(j)} = s\,(1 + 2\alpha - \beta)\,(\tau_{15} - \tau_{25}) + \beta s\,(t - u)\,. \tag{334}$$

Because the smallest loop in diagram (j) carrying ℓ_5 is a box subdiagram, our auxiliary constraints require that this momentum be absent from $N^{(e)}$. Setting the first term in equation (334) to zero implies that $\beta = 1 + 2\alpha$, which in turn leads to

$$N^{(e)} = s\,(\tau_{45} + \tau_{15}) + (\alpha\,(t - u) + t)\,(s + \tau_{15} - \tau_{25})\,, \tag{335}$$

$$N^{(j)} = (1 + 2\alpha)\,(t - u)\,s\,, \tag{336}$$

leaving undetermined a single parameter α.

To obtain the value of the final parameter we use the numerator of diagram (a) expressed in terms of $N^{(e)}$ by equation (328). Because every loop in diagram (a) is part of a box, the auxiliary constraint that a one-loop box subdiagram cannot carry loop momentum then implies that $N^{(a)}$

cannot contain loop momentum. By solving the duality relations (328), the numerator $N^{(\mathrm{a})}$ is given by

$$\begin{aligned} N^{(\mathrm{a})} =& N^{(\mathrm{e})}\left(p_1,p_2,p_4,-p_3+\ell_5-\ell_6+\ell_7,\ell_5-\ell_6,-\ell_5\right) \\ &+N^{(\mathrm{e})}\left(p_2,p_1,p_4,-p_3-\ell_5+\ell_7,-\ell_5,\ell_5-\ell_6\right) \\ &-N^{(\mathrm{e})}\left(p_4,p_1,p_2,\ell_6-\ell_7,\ell_6,\ell_5-\ell_6\right)-N^{(\mathrm{e})}\left(p_4,p_2,p_1,\ell_6-\ell_7,\ell_6,-\ell_5\right) \\ &-N^{(\mathrm{e})}\left(p_3,p_1,p_2,\ell_7,\ell_6,\ell_5-\ell_6\right)-N^{(\mathrm{e})}\left(p_3,p_2,p_1,\ell_7,\ell_6,-\ell_5\right). \end{aligned} \tag{337}$$

Plugging in the value of the numerator factor $N^{(\mathrm{e})}$ in equation (335), and simplifying we obtain

$$\begin{aligned} N^{(\mathrm{a})} = s^2 + (1+3\alpha)\big(&(\tau_{16}-\tau_{46})\,s - 2\,(\tau_{17}+\tau_{37})\,s \\ &+ (\tau_{16}-2\tau_{17}-\tau_{26}+2\tau_{27})\,t + 4ut\big)\,. \end{aligned} \tag{338}$$

Demanding that this expression is independent of loop momenta, fixes the final parameter to be $\alpha = -1/3$ and completely determines the numerator of diagram (e) to be

$$N^{(\mathrm{e})} = s\,(\tau_{45}+\tau_{15}) + \frac{1}{3}\,(t-s)\,(s+\tau_{15}-\tau_{25})\,. \tag{339}$$

With a proposed expression for $N^{(\mathrm{e})}$ in hand, equation (328) then determines all other numerators and thus the complete amplitude. The resulting numerators are collected in table 15. To confirm that this is indeed the correct amplitude, it is necessary to verify a complete set of unitarity cuts. The three-loop four-gluon amplitude in $\mathcal{N}=4$ SYM theory is determined only by its maximal and next-to-maximal cuts, so it is relatively straightforward to check them all. As a highly-nontrivial test, one can also check the next-to-next-to-maximal cuts. The resulting cuts match those of previous expressions of the amplitude [407, 408] on all D-dimensional unitarity cuts. Thus, the amplitude is complete. We stress again that it is highly-nontrivial that there exists a solution to *all* duality relations which is consistent with all unitarity cuts and exhibits all the diagram symmetries.

Squaring the numerators $n^{(x)} = stA_4^{\mathrm{tree}}(1,2,3,4)n^{(x)}$, using equation (279), yields the numerators for the three-loop four-point $\mathcal{N}=8$ supergravity superamplitude. This form has been confirmed against previous expressions [407, 408] on a spanning set of D-dimensional unitarity cuts [2]. Using as the second copy the three-loop four-point numerator factors of $\mathcal{N}<4$ SYM theories yields the three-loop four-graviton amplitudes in $(4+\mathcal{N})$-extended supergravity theories. The case $\mathcal{N}=0$ was discussed at length in [33, 36, 37], where it was used to explore the UV properties of half-maximal supergravities and demonstrate the absence of UV divergences at this loop order in four dimensions.

Exercise 6.13. Work through the entries in table 15 to explicitly confirm that they do indeed satisfy BCJ duality.

The strategy followed above generalizes straightforwardly to the four-loop four-point [6] and two-loop five-point amplitudes of $\mathcal{N}=4$ SYM and supergravity. It has also been tested in a variety of other cases, including the one- and two-loop amplitudes in various theories with fewer supersymmetries [241], and nonsupersymmetric gauge and gravity theories [12, 396] as well as to the construction of form factors in $\mathcal{N}=4$ SYM through five loops [9, 17]. While the application of the double-copy construction to gauge-theory form factors yields quantities consistent with the linearized diffeomorphism invariance of a gravity theory, their precise physical interpretation is currently an open question.

Table 15. The numerator factors for diagrams in figure 27 [2]. The first column labels the diagram, the second column the relative numerator factor for $\mathcal{N}=4$ SYM theory. The square of this is the relative numerator factor for $\mathcal{N}=8$ supergravity. The momenta are labeled as in figure 27 and the τ_{ij} are defined equation (333).

diagram	$\mathcal{N}=4$ SYM ($\sqrt{\mathcal{N}=8 \text{ supergravity}}$) numerator
(a)–(d)	s^2
(e)–(g)	$\big(s(-\tau_{35}+\tau_{45}+t)-t(\tau_{25}+\tau_{45})+u(\tau_{25}+\tau_{35})-s^2\big)/3$
(h)	$\big(s(2\tau_{15}-\tau_{16}+2\tau_{26}-\tau_{27}+2\tau_{35}+\tau_{36}+\tau_{37}-u)$
	$+t(\tau_{16}+\tau_{26}-\tau_{37}+2\tau_{36}-2\tau_{15}-2\tau_{27}-2\tau_{35}-3\tau_{17})+s^2\big)/3$
(i)	$\big(s(-\tau_{25}-\tau_{26}-\tau_{35}+\tau_{36}+\tau_{45}+2t)$
	$+t(\tau_{26}+\tau_{35}+2\tau_{36}+2\tau_{45}+3\tau_{46})+u\,\tau_{25}+s^2\big)/3$
(j)–(l)	$s(t-u)/3$

6.5. Other examples

The examples described above are but a sample of the many loop-level amplitudes that have representations that manifest the duality between color and kinematics. Among them are various examples of supersymmetric [4, 6–8, 11, 14, 16, 19, 20, 398] and nonsupersymmetric [10, 12, 15, 156, 348] gauge-theory amplitudes, form factors [9, 17, 18, 22, 23], string theory amplitudes, and their field theory limits [13, 371, 409–411]. Additionally, a systematic method to determine BCJ numerators for one-loop amplitudes which makes use of the global constraints on the loop-momentum dependence of the numerators imposed by the kinematic Jacobi identities was introduced in reference [11].

It has moreover been shown that the leading and subleading [5, 412] factorization theorems of gauge and gravity theories are consistent with the double-copy procedure to all orders in perturbation theory, thus providing some all-loop-levels evidence for this conjecture. In another interesting example, the duality has been applied to QCD scattering amplitudes, in order to find hidden relations between coefficients of loop integrals [413, 414].

The multitude of nontrivial examples suggests that the duality between color and kinematics does extend to loop amplitudes, even though no proof exists as yet. Finding a proof would likely provide a guide towards more systematic constructions of representations of amplitudes that manifest the duality.

Even when they are expected to exist, the construction of duality-satisfying amplitudes representations is not always straightforward. An alternative, discussed and illustrated on the two-loop four-point all-plus pure-YM amplitude in [415], is to relax the demand that the duality be manifest off shell and impose instead that it be manifest only on a spanning set of generalized unitarity cut. The double-copy construction then yields an expression which coincides with the corresponding supergravity amplitude on a spanning set of cuts; the two must therefore be the same.

It has proven difficult to find representations of the five-loop four-point $\mathcal{N}=4$ SYM amplitude for which the duality between color and kinematics is manifest. In particular, the expected power-counting constraints suggested by supersymmetry appear to not be compatible with duality and D-dimensional unitarity cuts. To find the corresponding $\mathcal{N}=8$ supergravity amplitude the generalized double-copy construction provides an efficient approach, as it uses gauge-theory amplitudes' representations that should exhibit the duality but do not manifest it; we review it in the next Section. The success of the generalized double copy, using the five-loop amplitudes' representations constructed in [38, 218, 416], strongly suggests that it should be possible to manifest the duality for the five-loop four-point $\mathcal{N}=4$ SYM amplitude. Presumably, this will require integrands that relax some the simplifying assumptions, such as locality or manifest relabeling symmetry of the diagrams. Using string theory to define global diagram labels in the field-theory limit, there has been some very interesting progress on finding integrands that manifest CK duality [417].

7. Generalized double copy

Whenever gauge-theory amplitudes are available in a form that manifests the duality between color and kinematics, the BCJ double-copy construction provides the most efficient means for obtaining the corresponding gravity integrands. However, in some cases, such as the five-loop four-point amplitude of $\mathcal{N}=8$ supergravity, it has proven difficult to find such representations. In other cases, such as the all-plus two-loop five-gluon amplitude in pure-YM theory, the BCJ form of the amplitude has a superficial power-count much worse than that of Feynman diagrams [15] and thus an analysis of UV properties of its double copy is cumbersome at best. It can therefore be advantageous to have a double-copy method for converting generic representations of gauge-theory amplitude to gravity ones, without first constructing BCJ representations for them. Such a procedure has been developed in [416] and applied in [38, 218] to construct the five-loop four-point integrand of $\mathcal{N}=8$ supergravity and to extract its UV properties after integration[49].

If we start with a generic representation of a gauge-theory amplitude where BCJ duality is not manifest and apply the double-copy substitution rule (44), in general, we do not obtain a correct gravity amplitude. Nevertheless, this 'naive double copy' can be systematically corrected to give the desired amplitude. As we summarize below, the correction terms have a regular pattern reminiscent of the KLT tree-level amplitudes relations [86], allowing us to obtain the most complicated corrections directly from gauge theory.

7.1. Generalities

To start the generalized double-copy construction we first need to reorganize slightly the two (possibly distinct) gauge-theory amplitudes that comprise the two sides of the double copy. Starting with any local representations of the amplitudes, which may include four- or higher-point contact terms, we reorganize them into a format that has only three-point vertices and the maximum number of propagators. If a given term has fewer propagators we multiply and divide by the propagators needed to form diagrams with only cubic vertices that correspond to the color factor of the given term. Once the gauge-theory amplitudes are written in this format the next step is to apply the double-copy substitution (44) to these amplitudes, despite neither gauge theory manifesting the BCJ duality between color and kinematics. As already mentioned, this so-constructed naive double-copy expression is, in general, not a correct (super)gravity amplitude. Nonetheless, it is a good starting point for obtaining the full gravity amplitude as, by construction, it reproduces the maximal and next-to-maximal cuts of the desired (super)gravity amplitude. (See appendix C for a description of the method of maximal cuts.) In maximal cuts, where all propagators are cut, the amplitude is reduced to a sum of

[49] Another possible method, proposed in [415] and illustrated on the two-loop four-point pure-YM amplitude in D dimensions, is to demand that the duality between color and kinematics holds only on unitarity cuts. This provides a straightforward construction of the generalized unitarity cuts of the double-copy theory, which need to be subsequently assembled into the complete gravity amplitude.

products of gauge-theory three-point tree amplitudes. Because on-shell gravity three-vertices are products of gauge-theory ones, maximal cuts trivially satisfy the double-copy property for any representation of the single copy amplitudes. The next to maximal cuts, where one of the propagators are left uncut, also automatically give the correct gravity expressions because, if present, the duality between color and kinematics is automatic for on-shell four-point tree amplitudes [1].

Beyond the next-to-maximal cuts, the naive double copy will generally *not* give correct unitarity cuts, and nontrivial corrections are necessary. These required corrections can be organized into contact terms via the method of maximal cuts described in appendix C. However, for complicated problems, such as $\mathcal{N}=8$ supergravity [218] at five-loops it becomes cumbersome to use the method of maximal cuts to obtain the missing terms.

Instead, it turns out that it is possible to construct general formulae that relate the necessary cut-correction terms to the violations of the kinematic Jacobi relations (41) in the gauge-theory amplitudes. The derivation of such formulae relies only on the existence of duality-satisfying representations for all tree-level amplitudes.

Indeed, the existence of BCJ representations at tree level implies that such representations should also exist for all cuts of gauge-theory amplitudes that decompose the loop integrand into products of tree amplitudes. This further implies that the corresponding generalized unitarity cuts of the gravity amplitude can be expressed in double-copy form,

$$\mathcal{C}_{\mathrm{GR}} = \sum_{i_1,\ldots,i_q} \frac{n^{\mathrm{BCJ}}_{i_1,i_2,\ldots i_q}\tilde{n}^{\mathrm{BCJ}}_{i_1,i_2,\ldots i_q}}{D^{(1)}_{i_1}\ldots D^{(q)}_{i_q}}\,, \tag{340}$$

where the n^{BCJ} and $\tilde{n}^{\mathrm{BCJ}}$ are the BCJ numerators associated with each of the two single-copy parent theories. In this expression the cut conditions are understood as being imposed on the numerators. Each sum runs over the diagrams of each tree amplitude composing the generalized cut and $D^{(m)}_{i_m}$ are the products of the uncut propagators associated to each diagram of mth tree amplitude. This notation is illustrated in figure 29 for an N^{2}MC at three loops. In this figure, each of the two four-point blobs is expanded into three diagrams, giving a total of nine diagrams. For example, the combination of indices $i_1 = 1$ and $i_2 = 1$ refers to the three-loop diagram obtained by taking the first diagram from each blob and connecting it to the three-point vertices; the result, in the ordering of diagrams chosen for each of the two four-point amplitude, is the first cubic diagram on the first line of figure 29. The denominators in equation (340) correspond to the thick (colored) lines in the diagrams.

The BCJ numerators in equation (340) are related [2, 41] to those of an arbitrary representation by a generalized gauge transformation which shifts the numerators subject to the constraint that the amplitude is unchanged; the shift parameters follow the same labeling scheme as the numerators themselves,

$$n_{i_1,i_2,\ldots i_q} = n^{\mathrm{BCJ}}_{i_1,i_2,\ldots i_q} + \Delta_{i_1,i_2,\ldots i_q}\,. \tag{341}$$

The shifts $\Delta_{i_1,i_2,\ldots i_q}$ are constrained to leave the corresponding cuts of the gauge-theory amplitude unchanged. Using such transformations we can reorganize a gravity cut in terms of cuts of a naive double copy and an additional contribution,

$$\mathcal{C}_{\mathrm{GR}} = \sum_{i_1,\ldots,i_q} \frac{n_{i_1,i_2,\ldots i_q}\tilde{n}_{i_1,i_2,\ldots i_q}}{D^{(1)}_{i_1}\ldots D^{(q)}_{i_q}} + \mathcal{E}_{\mathrm{GR}}(\Delta)\,, \tag{342}$$

where the cut conditions are imposed on the numerators. Rather than expressing the correction $\mathcal{E}_{\mathrm{GR}}$ in terms of the generalized-gauge-shift parameters, it is useful to re-express the correction

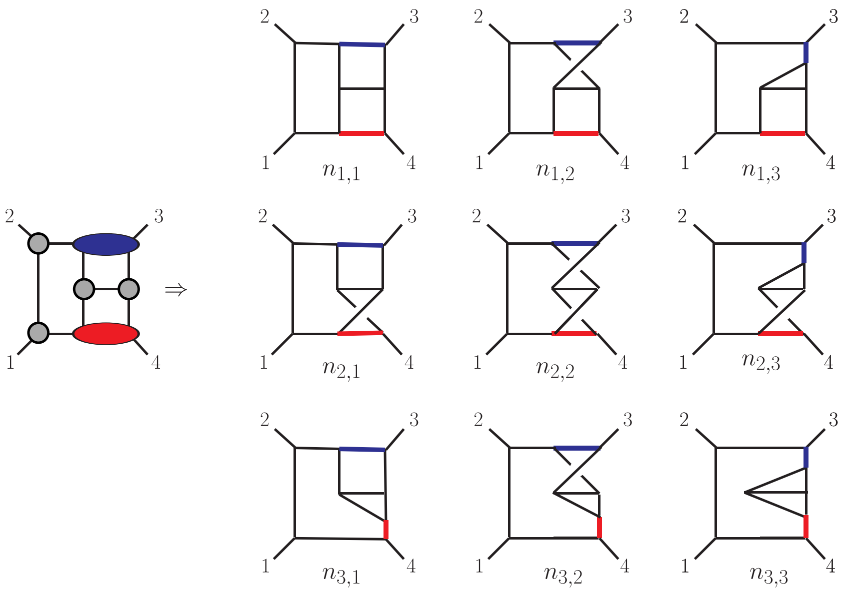


Figure 29. An example illustrating the notation in equation (340). Expanding each of the two four-point blob gives a total of nine diagrams. The $n_{i,j}$ correspond to labels used in the generalized unitarity cut. The shaded thick (blue and red) lines are the propagators around which BCJ discrepancy functions are defined.

terms as bilinears in the violations of the kinematic Jacobi relations (41) by the generic gauge-theory amplitude numerators. These violations are referred to as BCJ discrepancy functions.

As an example, the generalized unitarity cut in figure 29 is composed of two four-point tree amplitudes and the rest are three-point amplitudes. For any cut of this structure, two four-point trees connected to any number of three-point trees, the correction has a simple expression,

$$\mathcal{E}_{\mathrm{GR}}^{4\times 4} = -\frac{1}{d_1^{(1,1)} d_1^{(2,1)}} \left(J_{\bullet 1,1} \tilde{J}_{1,\bullet 2} + J_{1,\bullet 2} \tilde{J}_{\bullet 1,1} \right) , \tag{343}$$

where $d_i^{(b,p)}$ is the pth propagator of the ith diagram inside the bth amplitude factor[50] and

$$J_{\bullet 1,i_2} \equiv \sum_{i_1=1}^{3} n_{i_1 i_2} , \qquad J_{i_1,\bullet 2} \equiv \sum_{i_2=1}^{3} n_{i_1 i_2} , \qquad \tilde{J}_{\bullet 1,i_2} \equiv \sum_{i_1=1}^{3} \tilde{n}_{i_1 i_2} , \qquad \tilde{J}_{i_1,\bullet 2} \equiv \sum_{i_2=1}^{3} \tilde{n}_{i_1 i_2} , \tag{344}$$

are BCJ discrepancy functions[51]. Our notation is to label the type of cut by $m_1 \times m_2 \times \cdots m_k$ where each m_i specifies the number of legs on each tree amplitude with $m_i \geqslant 4$ composing the cut. These discrepancy functions vanish whenever the numerators involved satisfy the BCJ relations, even if the representation as a whole does not satisfy them. Such expressions are

[50] We will sometimes omit the second argument, p, when an amplitude factor has a single propagator.

[51] We will sometimes denote the BCJ discrepancy function with either $\bullet$ in the position i or by $\{i, 1\}$ when the ith amplitude factor has a single propagator (i.e. it is a four-point amplitude).

Table 16. A non-BCJ form of the three-loop four-point $\mathcal{N}=4$ SYM diagram numerators from [408]. We define $\tau_{ij}=2p_i\cdot p_j$, $s=(p_1+p_2)^2$, $t=(p_2+p_3)^2$ and $u=(p_1+p_3)^2$.

Diagram	$\mathcal{N}=4$ SYM numerators.
(a)–(d)	s^2
(e)–(g)	$s(p_5^2+\tau_{45})$
(h)	$s(\tau_{26}+\tau_{36})-t(\tau_{17}+\tau_{27})+st$
(i)	$s(p_5^2+\tau_{45})-t(p_5^2+\tau_{56}+p_6^2)-(s-t)p_6^2/3$

not unique and can be rearranged using various relations between discrepancy functions [218, 416, 418–420]. For example, a more symmetric version, equivalent to equation (343), is

$$\mathcal{E}^{4\times4}_{\mathrm{GR}}=-\frac{1}{9}\sum_{i_1,i_2=1}^{3}\frac{1}{d^{(1,1)}_{i_1}d^{(2,1)}_{i_2}}\left(J_{\bullet1,i_2}\tilde{J}_{i_1,\bullet2}+J_{i_1,\bullet2}\tilde{J}_{\bullet1,i_2}\right). \tag{345}$$

Similarly, a cut with a single five-point tree amplitude and the rest three-point tree amplitudes is given by

$$\mathcal{C}^{5}_{\mathrm{GR}}=\sum_{i=1}^{15}\frac{n_i\tilde{n}_i}{d^{(1,1)}_{i}d^{(1,2)}_{i}}+\mathcal{E}^{5}_{\mathrm{GR}}\qquad\text{with}\qquad\mathcal{E}^{5}_{\mathrm{GR}}=-\frac{1}{6}\sum_{i=1}^{15}\frac{J_{\{i,1\}}\tilde{J}_{\{i,2\}}+J_{\{i,2\}}\tilde{J}_{\{i,1\}}}{d^{(1,1)}_{i}d^{(1,2)}_{i}}, \tag{346}$$

where $J_{\{i,1\}}$ and $J_{\{i,2\}}$ are BCJ discrepancy functions associated with the first and second propagator of the ith diagram. (See [218] for further details.)

As the cut level k increases, the formulae relating the amplitudes' cuts with the cuts of the naive double copy become more intricate, but the basic building blocks remain the BCJ discrepancy functions. Formulas like (345), (346) and their generalizations can enormously streamline the computation of the contact term corrections and are especially helpful at five loops at the N^{2}MC and N^{3}MC level, where calculating the contact terms via the maximal-cut method can be rather involved. Beyond this level, the contact terms become much simpler due to a restricted dependence on loop momenta and are better dealt with using the method of maximal cuts and KLT relations [86], as described in [218].

7.2. Three-loop example

To illustrate the discussion above, we now present a relatively simple though nontrivial construction of the three-loop four-point amplitude of $\mathcal{N}=8$ supergravity, which was studied in several other different approaches [2, 6, 407, 408]. As described in section 6 the most efficient way to construct it is to first obtain a BCJ representation of corresponding $\mathcal{N}=4$ SYM amplitude and then apply the double-copy construction. Instead, we construct it here through the generalized double copy, from a non-BCJ form of the $\mathcal{N}=4$ SYM amplitude of [408] whose numerators are included in table 16 with the momentum labeling in figures 27(a)–(i), corresponding to the one of [2]. An overall factor of stA_4^{tree} is not included in table 16.

Following the generalized double-copy construction, the $\mathcal{N}=8$ supergravity numerators of diagrams (a)–(i) are squares of the corresponding $\mathcal{N}=4$ SYM ones:

$$N^{\mathcal{N}=8}_{(x)}=n^2_{(x)}, \tag{347}$$

where $x\in\{a,\ldots,i\}$. This defines the naive double copy. This is not the complete supergravity amplitude given that the gauge-theory numerators do not satisfy the BCJ relations (41),

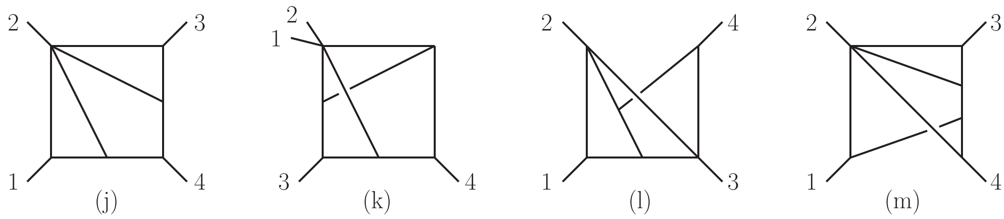


Figure 30. Nonvanishing contact terms appearing in the generalized double copy construction of the three-loop four-point amplitude of $\mathcal{N}=8$ supergravity.

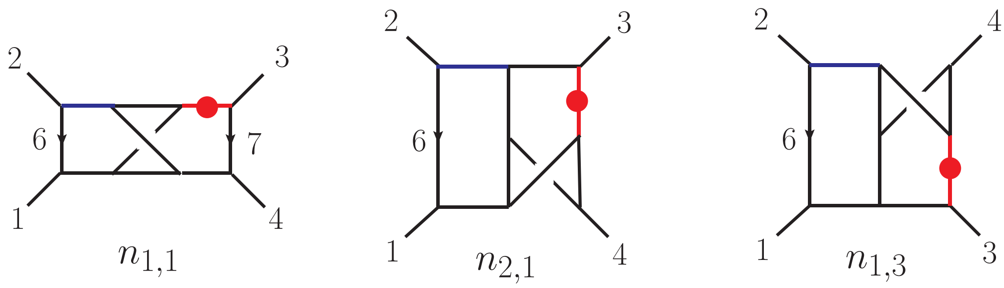


Figure 31. The three diagrams whose kinematic numerators contribute to $J_{\{1,1\},1}$. The thick shaded (red) line marks the off-shell legs participating in the dual Jacobi relation. The shaded (red) dot indicates the off-shell leg of the second amplitude factor.

as can be confirmed by checking its generalized unitarity cuts. To complete the supergravity amplitude we need to find the missing contact terms.

Given that the N^1MC-level contact terms are automatically accounted for in the naive double copy, contact terms first appear at the N^2MC level. There are a total of 62 possible independent such contact diagram, corresponding to diagrams obtained by starting from the first nine diagrams in figure 27 and collapsing all pairs of propagators. Of these, all but the four diagrams (j)–(m) in figure 30 vanish.

As an example, consider the contact diagram in 30(l), composed of two four-point vertices. We obtain it from equation (343). First, we identify the nine cubic diagrams that contribute to it (some are vanishing) and pick one whose numerator we label as $n_{1,1}$; we choose diagram (c) in figure 27. The two J-functions are calculated by relabeling the appropriate numerators to the labels of figure 31. For example, $J_{\{u_1,1\},1}$ is obtained from the $\mathcal{N}=4$ SYM numerators of the three diagrams shown in figure 31,

$$n_{1,1}=s^2, \qquad n_{2,1}=s\left(t+\tau_{26}+\tau_{36}\right), \qquad n_{3,1}=s\left(u-\tau_{36}\right), \tag{348}$$

corresponding to relabeling of diagrams (c) and (g) in figure 27. Summing and applying momentum conservation gives $J_{\{1,1\},1}=s\tau_{26}$. Similarly, $J_{1,\{1,1\}}=s\tau_{37}$. With these labels, the two off-shell inverse propagators are τ_{26} and τ_{37}, so that from equation (343) the $\mathcal{N}=8$ supergravity contact term numerator for diagram (l) is

$$N_{(\mathrm{l})}^{\mathcal{N}=8}=-2\frac{J_{\{1,1\},1}J_{1,\{1,1\}}}{\tau_{26}\tau_{37}}=-2s^2\,. \tag{349}$$

The other three independent contact terms corresponding to diagrams (j), (k) and (m), can similarly be obtained from equation (346), with the result

$$N_{(\mathrm{j})}^{\mathcal{N}=8} = -\tfrac{1}{9}\left(s-t\right)^2 , \qquad N_{(\mathrm{k})}^{\mathcal{N}=8} = N_{(\mathrm{m})}^{\mathcal{N}=8} = -2s^2. \tag{350}$$

All other nonvanishing contact terms are relabelings of these.

7.3. Towards general formulae

This generalized double-copy procedure has been systematically used to obtain the five-loop four-point integrand of $\mathcal{N}=8$ supergravity [218], which was then used to analyze the UV properties of this theory at five loops [38]. In this case, it was sufficient to work out formulae for the extra corrections up to the N^{3}MCs, because beyond this the missing contact terms are simple enough to straightforwardly obtain by numerical analysis.

As discussed before, equations (343) and (346) can be used for all N^{2}MCs in any double-copy theory. These are sufficient to determine the three-loop four-point amplitude in $\mathcal{N}=8$ supergravity, because of its low power count. Beyond this order the corresponding formulae for $\mathcal{E}$ depend on the detailed labeling of the corresponding cut. We include here $\mathcal{E}_{\mathrm{GR}}^{4\times4\times4}$ and $\mathcal{E}_{\mathrm{GR}}^{5\times4}$ and comment on $\mathcal{E}_{\mathrm{GR}}^{6}$ given as an ancillary file in [218].

The additional terms that promote a cut composed of three four-point amplitude factors of the naive double copy to the cut of the corresponding double-copy theory [218, 416] are obtained by following the steps detailed in section 7.1. It is convenient to organize then into the contribution of single- and double-discrepancy functions:

$$\mathcal{E}_{\mathrm{GR}}^{4\times4\times4} = T_1 + T_2\,. \tag{351}$$

They are

$$\begin{aligned} T_1 &= -\sum_{i_3=1}^{3}\frac{J_{\bullet1,1,i_3}\tilde{J}_{1,\bullet2,i_3}}{d_1^{(1)}d_1^{(2)}d_{i_3}^{(3)}} - \sum_{i_2=1}^{3}\frac{J_{\bullet1,i_2,1}\tilde{J}_{1,i_2,\bullet3}}{d_1^{(1)}d_{i_2}^{(2)}d_1^{(3)}} - \sum_{i_1=1}^{3}\frac{J_{i_1,\bullet2,1}\tilde{J}_{i_1,1,\bullet3}}{d_{i_1}^{(1)}d_1^{(2)}d_1^{(3)}} + \left\{J\leftrightarrow\tilde{J}\right\} ,\\ T_2 &= \frac{J_{\bullet1,1,1}\tilde{J}_{1,\bullet2,\bullet3}}{d_1^{(1)}d_1^{(2)}d_1^{(3)}} + \frac{J_{1,\bullet2,1}\tilde{J}_{\bullet1,1,\bullet3}}{d_1^{(1)}d_1^{(2)}d_1^{(3)}} + \frac{J_{1,1,\bullet3}\tilde{J}_{\bullet1,\bullet2,1}}{d_1^{(1)}d_1^{(2)}d_1^{(3)}} + \left\{J\leftrightarrow\tilde{J}\right\} , \end{aligned} \tag{352}$$

where e.g. $\tilde{J}_{1,\bullet2,\bullet3}$ is defined as

$$\tilde{J}_{i_1,\bullet2,\bullet3} = \sum_{i_2=1}^{3}\sum_{i_3=1}^{3}\tilde{n}_{i_1,i_2,i_3}\,, \tag{353}$$

with n_{i_1,i_2,i_3} being the numerators of the cut of the naive double copy. As mentioned earlier, we dropped the second upper label in $d_i^{(b,p)}$ defined below equation (343) because four-point diagrams have only a single propagator, so $b=1$ for all terms in equation (352).

To simplify T_2 we used the relations

$$\begin{aligned} &\frac{J_{1,\bullet2,\bullet3}}{d_1^{(1)}} = \frac{J_{2,\bullet2,\bullet3}}{d_2^{(1)}} = \frac{J_{3,\bullet2,\bullet3}}{d_3^{(1)}}\,, \quad \frac{J_{\bullet1,1,\bullet3}}{d_1^{(2)}} = \frac{J_{\bullet1,2,\bullet3}}{d_2^{(2)}} = \frac{J_{\bullet1,3,\bullet3}}{d_3^{(2)}}\,,\\ &\frac{J_{\bullet1,\bullet2,1}}{d_1^{(3)}} = \frac{J_{\bullet1,\bullet2,2}}{d_2^{(3)}} = \frac{J_{\bullet1,\bullet2,3}}{d_3^{(3)}}\,, \end{aligned} \tag{354}$$

which identify various double-discrepancy functions.

The additional terms that promote a cut composed of one five-point and one four-point amplitude factors of the naive double copy to the cut of the corresponding double-copy theory can be organized as

$$\mathcal{E}_{\rm GR}^{5\times4} = \sum_{i=1}^{15}\sum_{j=1}^{3}\frac{1}{d_i^{(1,1)}d_i^{(1,2)}d_j^{(2,1)}}\left[-\frac{1}{6}J_{\{i,1\},j}\tilde{J}_{\{i,2\},j} - \left(-\frac{1}{3}\right)\times\frac{1}{6}\left(J_{\{i,1\},j}\tilde{J}_{\{i,2\},\bullet 2} + J_{\{i,2\},j}\tilde{J}_{\{i,1\},\bullet 2}\right)\right.$$
$$\left. -\frac{1}{5}J_{\{i,1\},j}\tilde{J}_{i,\bullet 2} - \frac{1}{5}J_{\{i,2\},j}\tilde{J}_{i,\bullet 2} + \frac{1}{30}\sum_{k\in\mathcal{J}_i}\sigma_{k,i}J_{\{k,1\},j}\tilde{J}_{i,\bullet 2} + \frac{1}{30}\sum_{k\in\mathcal{J}_i}\sigma_{k,i}J_{\{k,2\},j}\tilde{J}_{i,\bullet 2}\right] + \{J\leftrightarrow\tilde{J}\}\,, \tag{355}$$

where $\mathcal{J}_i$ is the set of five diagrams connected to diagram i through Jacobi relations on the two propagators, including diagram i which appears once, and $\sigma_{k,i}$ are the signs with which their color factors enter in the color Jacobi relations, with the normalization that $\sigma_{i,i} = 1$[52]. While equation (354) is quite different from the corresponding $\mathcal{E}_{\rm GR}^{5\times4}$ in [218], the two expressions are in fact equivalent, as can be shown by reducing to a basis of BCJ discrepancy functions, or by directly evaluating the additional terms for a choice of representation of the five-point amplitude. The essential advantage of equation (354) is that it does not make reference either to a specific ordering of the diagrams of the five-point amplitude or to a specific choice of order of propagators for each diagram. These features may be the key to extending equation (354) to cuts with higher-point tree-level amplitude factors.

Similarly to $\mathcal{E}_{\rm GR}^{4\times4}$ and $\mathcal{E}_{\rm GR}^{5}$, both $\mathcal{E}_{\rm GR}^{4\times4\times4}$ and $\mathcal{E}_{\rm GR}^{5\times4}$ are not local. To extract their corresponding contact terms it is necessary to subtract the contribution of the 4×4- and 5-contact terms which contribute to the $4\times4\times4$ and 5×4 cuts. The strategy discussed in the previous section applies here as well, so we will not repeat it.

Expressions for the extra terms that promote cuts with higher-point tree-level factors of the naive double copy to the corresponding cuts of the double-copy theory can be obtained following the discussion in section 7.1. For example, the additional terms for a cut with a single six-point factor are included in the ancillary file `ExtraJ_6pt.m` of [218]. Unlike $\mathcal{E}_{\rm GR}^{4\times4}$, $\mathcal{E}_{\rm GR}^{5}$, $\mathcal{E}_{\rm GR}^{4\times4\times4}$ and $\mathcal{E}_{\rm GR}^{5\times4}$ above however, $\mathcal{E}_{\rm GR}^{6}$ is presented in terms of a basis of independent discrepancy functions, obtained by solving the constraints they obey due to their definition in terms of the cut kinematic numerators of the single-copy parent theories. The expression is also not manifestly organized in terms of the kinematic denominators of the 105 diagrams of the six-point tree-level diagram. While, for these reasons, the available $\mathcal{E}_{\rm GR}^{6}$ is not manifestly crossing symmetric, it is sufficient for greatly simplifying the analytic structure of N^{3}MC with a single six-point tree amplitude, compared to the direct construction of such cuts via e.g. the KLT relations.

A feature of the nonsymmetric correction terms $\mathcal{E}_{\rm GR}$ expressed in terms of the some basis of BCJ discrepancy functions is that, when evaluated on a cut, they may lead to terms that behave as $0/0$. These are harmless when the 0 in the numerator is manifest, since it corresponds to an absent diagram. Sometimes, however, the 0 in the numerator is not manifest and arises due to a cancellation between distinct terms, that can leave behind a nontrivial finite piece. When this

[52] That is, the color factors of the corresponding diagrams obey the relation

$$c_i + \sum_{k\in\mathcal{J}_i}\sigma_{k,i}c_k = 0\,,$$

which is just the sum of the two Jacobi relations on the two propagators of diagram i.

occurs, the simplest strategy is to take advantage of the asymmetry in the formula, to relabel it to avoid such problematic cases.

Generalized double-copy formulae such as those reviewed here, give the cuts of any double-copy theory in terms of generic representations of the amplitudes of the single-copy parent theories. It is therefore an interesting problem to find similar general formulae for more complicated—perhaps all—cuts at any loop order. We can argue based on the gauge invariance of the single-copy theories that the correction terms must be linear in the BCJ discrepancy functions of each of the single-copy theories [218]. That is,

$$\mathcal{E} = \sum_{i,j} M_{ij} J_i \tilde{J}_j \,, \tag{356}$$

for some appropriate matrix M_{ij} whose entries are rational functions of the kinematic invariants of the cut. This structure is compatible with the fact that the corrections should all vanish if the duality between color and kinematics were manifest in *either* one of the two single copies [41]. A further heuristic argument for the general form (356) of the correction terms $\mathcal{E}$ relies on an understanding of the structure of the terms that need to be added to cuts of the naive double copy in order to restore the linearized diffeomorphism invariance expected of the cuts of amplitudes of a gravitational theory. As we saw in sections 1 and 2, a gauge transformation of tree-level amplitudes—and thus also of the cuts of a loop amplitude—is given by a sum of terms each of which is proportional to some linear combination of color Jacobi relations. Consequently, a linearized diffeomorphism transformation of the naive double copy yields a sum of terms each containing a BCJ discrepancy function from either one of the two single copies. To restore diffeomorphism invariance these terms must be cancelled by the transformation of further terms that are added to the cuts of the naive double copy. Assuming that the structure of these terms is the same for all double copy theories, they must be of the form (356). See [218] for more details.

While the generalized double-copy method has already been successful for the highly nontrivial case of $\mathcal{N} = 8$ supergravity at five loops [218, 416], its development is only at the beginning. Having a general tool for converting gauge-theory amplitudes in any representation to gravity ones is clearly useful and important. A good starting point would be to derive general formulae for tree-level amplitudes [218, 418–420] in terms of a naive double copy, plus corrections in terms of the BCJ discrepancy functions. At present such formulae are known only through six points. If an elegant solution to the tree-level problem can be found, it should be immediately applicable to finding a general solution to the loop-level one. One obvious application would be towards a definitive resolution of the UV behavior of extended supergravity theories. This would require calculations beyond those that have already been carried out (see e.g. [38, 292]), and would likely need a version of the generalized double copy to be practical. $\mathcal{N} = 5$ supergravity at five loops is an especially interesting case for future study, given that at four loops it exhibits an enhanced cancellation of UV divergences [292]. It is important to know whether this continues at higher loops.

8. Classical double copy

As we have seen at length, the duality between color and kinematics and the double-copy construction are essential tools in the construction of gauge and (super)gravity scattering amplitudes at higher-loop orders and/or at higher multiplicity. In close analogy with tree-level scattering amplitudes, the perturbative construction of solutions of the classical equations of motion of a field theory (perhaps in the presence of sources) also exhibits an expansion in tree-level diagrams. One may consequently expect that, with an appropriate definition, some version of double-copy construction may lead to a construction of solutions of Einstein's equations (perhaps also in the presence of other fields) in terms of solutions of YM equations of motion (perhaps also in the presence of other fields). If one could turn the double copy into a systematic tool for analyzing classical solutions one could hope for new advances analogous to the ones that have occurred for scattering amplitudes.

As we shall discuss below, such a relation between classical solutions is not without subtleties and comes with quantifiable differences from the case of flat-space scattering amplitudes. Flat-space scattering amplitudes carry an inherent simplicity in that they are completely independent of gauge and field variable choices. However, in contrast to scattering amplitudes, generic classical solutions change nontrivially under gauge transformations and, moreover, they are sensitive to the nonlinear terms in the gauge transformations. Thus, to relate gauge and gravity solutions it is necessary to make correlated gauge choices in the two theories; the principles for making such choices are unclear. Related to this, the form of the equations of motion depends strongly on the choice of field variables. Thus, any naive extension of the scattering-amplitudes' double copy of fields can be completely obscured by nonlinear coupling-dependent terms that depend on some *a priori* chosen form of the equations of motion.

As yet, no coherent set of rules for the construction of double copies for generic classical solutions in gravity theories has been formulated, though a variety of nontrivial tantalizing examples have been found. (See e.g. [50–77].) Ideally, any such rules should smoothly generalize those of scattering amplitudes and reduce to them in the appropriate limits. The classes of examples that have been constructed and analyzed emphasize both the similarities and the differences between classical solutions and scattering amplitudes, and expose the subtleties that need to be addressed in order to formulate a general framework. Their existence, however, suggests that it may be possible to find generic solutions of a gravity theory in terms of solutions of the two gauge theories that give its scattering amplitudes. The most obvious application of these ideas are towards improving calculations of as well as calculations in post-Newtonian expansion of gravitational interaction potentials as well as calculations potentially relevant to gravitational-wave detection. These type of calculations can be phrased in terms of scattering amplitudes [78, 79, 421–426] and therefore are likely to lead to useful new results, such as the computation of the third post-Minkowskian contribution to the conservative two-body potential [80, 82].

In this section we describe the known constructions of gravity classical solutions in terms of gauge-theory solutions, commonly referred to as 'classical double copies'. We outline their relation and similarities with the double copy of scattering amplitudes and summarize the examples that have been discussed in this framework. We start with a description of perturbative solutions in gravity before turning to complete double copies.

8.1. Perturbative classical solutions vs. tree-level amplitudes

There is a close relation between solutions of classical equations of motion of some field theory and the Green's functions of that theory. The classical field generated by an arbitrary source is the generating functional for the tree-level connected Green's functions. Given a field theory of some field ϕ with Lagrangian $\mathcal{L}$, a solution of the equation of motion with general sources,

$$\frac{\delta\mathcal{L}}{\delta\phi} = \zeta\,, \tag{357}$$

is given in terms of the generating functional of connected tree-level Green's functions by [427]

$$\phi[x,\zeta] = \frac{\delta W[\zeta]^{\text{tree}}}{\delta\zeta}\,, \tag{358}$$

and moreover

$$W[\zeta] = \int d^D x\, \left(\mathcal{L}[\phi[x,\zeta]] - \zeta\phi[x,\zeta]\right). \tag{359}$$

The relation between Green's functions and scattering amplitudes given by the LSZ reduction implies in turn that, by amputating the sources, W becomes the generating functional of tree-level S-matrix elements. This may be realized by taking the source to be the quadratic operator acting on an on-shell wave solution of the free equation of motion. The solution (358) with such sources is the generating function of Berends-Giele currents—i.e. Green's functions of fundamental fields with exactly one leg off shell[53]; it therefore may also be interpreted as a solution of the Berends-Giele off-shell recursion relation [430]. This idea was used in [431, 432] to construct an implicit representation (referred to as the 'perturbiner') of gluon scattering amplitudes in four-dimensional YM theory and the gravitational dressing of certain classes of such amplitudes. Tree-level amplitudes of higher-dimensional and supersymmetric YM theories have been constructed using this method in [114, 433] and in certain effective field theories and deformations of YM theories in [172, 434]. It was also was used in [43] to construct the kinematic algebra dual to the color algebra in self-dual YM theory. Solutions for the supersymmetric versions of Berends-Giele current that manifest CK duality were given in [435].

Thus, equations (357) and (358) allow us to construct perturbative approximations of solutions with the appropriate source in terms of the scattering amplitudes of the theory. Moreover, should it be possible to resum the scattering amplitudes into a generating functional, equation (358) provides an exact solution of the equation of motion with the appropriate sources. Depending on the chosen sources, the construction can be carried out either in momentum space (if the sources are momentum eigenstates) or in position space.

Introducing sources in gauge and gravity theories can be confusing for at least two reasons. First, fixed sources coupling to vector fields or with the graviton may break gauge invariance. A resolution of this would-be problem is the gauge-fixing that is necessary for any (tree-level) computation, which already breaks gauge invariance. One then adds sources in the gauge-fixed theory, in which the question of gauge invariance should not arise. Second, related, nonabelian vector fields and gravitons self-interact and consequently they can self-source. Examples are all solutions of vacuum Einstein's equations as well as solutions of classical YM equations such as the instanton. For a stable configuration the matter stress tensor should be covariantly constant with respect to the metric that it sources; thus, it has some knowledge of the solution. This implies that the perturbative construction of such solutions requires a judicious choice of source which may itself receive corrections order by order in perturbation theory. Examples were discussed in e.g. [436] and [437] for the Schwarzschild and Reissner-Nordström black holes, respectively.

Unlike scattering amplitudes, solutions of the classical field Equations can be changed by (1) field redefinitions (2) coordinate changes and (3) gauge transformations (if gauge

[53] Green's functions with two legs off shell have been constructed in gauge theories coupled to fundamental matter in [428, 429].

symmetries are present)[54]. [55] As yet, Lagrangians that manifest CK duality are known to only a few perturbative orders [41, 42, 150, 151]. It is natural to expect that, if one had such a complete Lagrangian, classical solutions constructed through a classical double copy would solve its equations of motion. It is natural to expect that nontrivial field redefinitions and coordinate transformations are necessary to map such a solution to the field variables of a more standard Lagrangian. In fact, the perturbative Lagrangians manifesting the double-copy properties of gravity require this as well as elimination of auxiliary fields.

To illustrate the perturbative construction of solutions of supergravity equations of motion we outline here the derivation of the first terms [437] of the Reissner-Nordström solution—a charged black hole of (super)gravity coupled with a vector field A_μ of field strength $F_{\mu\nu}$. The vanishing-charge limit leads to the corresponding (first) term(s) in the Schwarzschild solution, discussed in [436]. The relevant action is

$$\begin{aligned} S &= S_G + S_{EM} + S_{\text{gauge fixing}} + S_\zeta\,, \\ \mathcal{L}_G &= \frac{1}{\kappa^2}\sqrt{-g}g^{\mu\nu}R_{\mu\nu}\,, \qquad\qquad \mathcal{L}_{EM} = \frac{1}{16\pi}\sqrt{-g}g^{\mu\rho}g^{\nu\sigma}F_{\mu\nu}F_{\rho\sigma}\,, \\ \mathcal{L}_\zeta &= \frac{1}{2}g^{\mu\nu}\left(\zeta^M_{\mu\nu}+\zeta^{EM}_{\mu\nu}\right) + A_\mu\zeta^\mu \equiv \frac{1}{2}\sqrt{-g}g^{\mu\nu}\left(T^M_{\mu\nu}+T^{EM}_{\mu\nu}\right) + \sqrt{-g}A_\mu j^\mu\,, \\ \mathcal{L}_{\text{gauge fixing}} &= -\frac{1}{2\pi}\left(\partial^\mu A_\mu\right)^2 + \frac{1}{2}\left(\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\right)\right)^2. \end{aligned} \tag{360}$$

To construct a perturbative solution around Minkowski space the metric is assumed of the form[56]

$$g^{\mu\nu} = \eta^{\mu\nu} + \kappa h^{\mu\nu}\,. \tag{361}$$

There are several sources that lead to the desired solution. One may choose, for example, the stress tensor of a charged point particle. Alternatively, one may choose an extended source—a sphere of radius ϵ of uniform mass density ρ and uniform charge density σ. The general form of the stress tensor is

$$T_\nu{}^\mu = (\rho+p)\,u_\nu u^\mu + p\delta_\nu{}^\mu, \quad g_{\mu\nu}u^\mu u^\nu = 1\,, \tag{362}$$

where ρ is the mass density function and p the (potentially phenomenological) pressure. In the case of a 'ball of dust' with uniform mass and charge densities, the components of the source turn out to be (after choosing $u=(1,0,0,0)$ and imposing covariant constancy of the stress tensor) [437]

$$\begin{aligned} \zeta^M_{00} &= \rho\theta(\epsilon-r) = \frac{3m}{4\pi\,\epsilon^3}\theta(\epsilon-r)\,, \qquad \zeta^M_{ij} = p^{(0)}\eta_{ij} = \frac{3Q^2}{8\pi\,\epsilon^6}\left(r^2-\epsilon^2\right)\delta_{ij}\theta(\epsilon-r)\,, \\ \zeta_\mu &= \sigma\delta^0_\mu\theta(\epsilon-r) = \frac{3Q}{4\pi\,\epsilon^3}\theta(\epsilon-r)\,\delta^0_\mu\,, \end{aligned} \tag{363}$$

[54] Symmetries of the equations of motion which are not symmetries of the action, such as parts of the U-duality symmetry of four-dimensional supergravity theories, may be used to generate inequivalent solutions from known ones. See e.g. [438] for a review.

[55] The same choices also affect Feynman rules; however, when Feynman rules are combined into a scattering amplitude there is no dependence upon these choices, although solutions of the classical equations of motion (and also Green's functions) depend on them.

[56] Note that this choice is different form the one typically used for perturbative S-matrix calculations and in later subsections, but it is useful here as it avoids nonlinear terms involving the metric fluctuation and the sources.

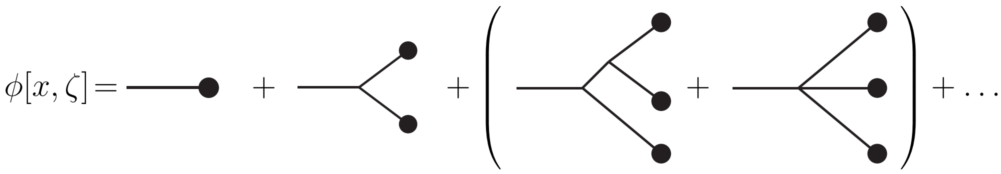


Figure 32. The first few terms in the expansion of a classical solution in terms of sources. Each heavy dot represents a source ζ. The free end is at position x. The weight of each vertex is not specified and may contain derivatives acting on the propagators connecting it to other vertices, sources or the point x.

where m and Q are the total mass and charge, respectively. The pressure $p^{(0)}$ is chosen such that this configuration of mass and charge densities is static under Newtonian gravitational attraction and Coulomb repulsion. As the metric receives κ^n corrections, so will the pressure function (hence the upper label '(0)' in ζ_{ij}^M above).

The first correction to the flat space metric and the electromagnetic field due to the sources (363), given by the first two diagrams in figure 32, is

$$\begin{aligned}
\langle A^\mu(x)\rangle_\zeta &= \int d^d y\, \Delta^{\mu\nu}(x-y)\,\zeta_\nu(y) + \dots\,, \\
\kappa\langle h^{\mu\nu}(x)\rangle_\zeta &= \frac{\kappa^2}{2}\int d^d y\, \Delta^{\mu\nu,\rho\sigma}(x-y)\,\zeta^M_{\rho\sigma}(y) \\
&\quad + \kappa^2 \int d^d y\, d^d x_1 d^d x_2\, \Delta^{\mu\nu,\rho\sigma}(x-y)\,\gamma_{\rho\sigma,\eta\tau}(\partial_y)\,\langle A^\eta(y)\rangle\langle A^\tau(y)\rangle\,. \qquad (364)
\end{aligned}$$

Here $\Delta^{\mu\nu,\rho\sigma}$ is the graviton propagator in the chosen de-Donder gauge (cf $\mathcal{L}_{\text{gauge fixing}}$), $\Delta^{\mu\nu}$ is the photon propagator in Lorentz gauge and $\gamma_{\rho\sigma,\eta\tau}(\partial_y)$ describes the graviton-photon three-point interaction. We note that, due to the κ dependence in expansion of the metric (361), the trilinear graviton-photon vertex contributes before the three-graviton vertex.

The extended nature of the source implies that the vector potential is different for $r < \epsilon$ and $r > \epsilon$. Denoting by tilde the Fourier-transform of the source,

$$\begin{aligned}
\langle A^\mu(y)\rangle &= \delta_0^\mu \int d^3 p\, \frac{e^{ip\cdot x}}{-p^2}\tilde{\zeta}_0(p) = \delta_0^\mu\left(\frac{Q}{r}\theta(r-\epsilon) + \left(\frac{3Q}{2\epsilon} - \frac{Qr^2}{2\epsilon^3}\right)\theta(\epsilon-r)\right) \\
&\equiv \delta_0^\mu U\,, \qquad (365)
\end{aligned}$$

which is just the Coulomb potential of the assumed charge distribution. Defining similarly the Newtonian potential of the given mass distribution,

$$W \equiv \int d^3 p\, \frac{e^{ip\cdot x}}{-p^2}\tilde{\zeta}^M_{00}(p) = \frac{\rho}{4\pi r}\theta(r-\epsilon) + \left(\frac{3\rho}{8\pi\epsilon} - \frac{\rho r^2}{8\pi\epsilon^3}\right)\theta(\epsilon-r)\,, \qquad (366)$$

and the action of the inverse Laplace operator on a time-independent function $F(x)$ as

$$\frac{1}{\nabla^2}F(x) \equiv \frac{1}{4\pi}\int d^3 y\, \frac{F(y)}{|x-y|}\,, \qquad (367)$$

the components of the metric fluctuations around flat Minkowski space are

$$
\begin{aligned}
\kappa\langle h^{00}\rangle_\zeta &= 8\pi G\left(W+3\frac{1}{\nabla^2}p^{(0)}-\frac{\eta_{kl}}{4\pi}\frac{1}{\nabla^2}\partial^k U\partial^l U\right),\\
\kappa\langle h^{ij}\rangle_\zeta &= 8\pi G\left(W-\frac{1}{\nabla^2}p^{(0)}-\frac{\eta_{kl}}{4\pi}\frac{1}{\nabla^2}\partial^k U\partial^l U\right)\delta^{ij}-4G\frac{1}{\nabla^2}\partial^i U\partial^j U,\\
\kappa\langle h^{i0}\rangle_\zeta &= 0.
\end{aligned}
\tag{368}
$$

Evaluating the integrals and defining the physical mass

$$
M = m+\frac{3}{5}\frac{Q^2}{\epsilon},
\tag{369}
$$

it follows [437] that for $r>\epsilon$ the metric components are

$$
\begin{aligned}
g^{00} &= 1+\frac{2MG}{r}-\frac{Q^2G}{r^2}+\mathcal{O}\left(G^2\right),\\
g^{ij} &= -\left(1-\frac{2MG}{r}\right)\delta^{ij}+\frac{Q^2G}{r^4}x^i x^j+\mathcal{O}\left(G^2\right),\\
g^{i0} &= 0.
\end{aligned}
\tag{370}
$$

This matches the Reissner–Nordström solution in Cartesian coordinates and de Donder gauge [437][57]:

$$
\begin{aligned}
ds^2 = {}& \frac{r^2+Q^2G-M^2G^2}{(r+MG)^2}dt^2-\left(1+\frac{MG}{r}\right)^2\left(dx^i\right)^2\\
&+\frac{\left(Q^2G-M^2G^2\right)(r+MG)^2}{r^4\left(r^2+Q^2G-M^2G^2\right)}\left(x_i\,dx^i\right)^2.
\end{aligned}
\tag{371}
$$

We note that the $Q\to 0$ limit yields the Schwarzschild solution [436] as well as that the size of the mass and charge distribution do not affect the exterior solution, in agreement with Birkhoff's theorem, which states that any spherically symmetric solution of the vacuum field equations must be static and asymptotically flat. The size of the distribution enters however the definition (369) of the physical mass M for nonvanishing electric charge. We also note that equation (369) is a reflection of the field backreaction on sources. In fact, the redefinition (369) is necessary for the solution to have a smooth limit to a point source. The relation between the physical mass M and the 'free mass' m receives further corrections as higher orders are included.

8.2. Perturbative spacetimes and the double copy

The double-copy formulation of classical gravity calculations has the potential to streamline calculations such as those outlined in the previous subsection by exploiting the close relation between the tree expansion in figure 32 and that of tree-level S-matrix elements. Given that we do not as yet have a general framework for applying the double copy to perturbative solutions, detailed analyses of specific examples, as we do below, help identify the correct physical extension of the amplitudes double-copy rules to this setting. Before we proceed to summarize the various options and illustrate their application to this problem, we begin with several

[57] While this is different from the standard form of the Reissner-Nordström solution, it can be mapped to it by a coordinate transformation and field redefinition.

comments which connect it to some of the calculations above and alert the reader to points that will arise.

As noted in section 2, the double-copy spectrum naturally contains a dilaton and a two-index antisymmetric tensor (or equivalently a pseudo-scalar in four dimensions). As for tree-level scattering amplitudes where these unwanted states can be projected out at tree level by a suitable choice of asymptotic states, solutions of Einstein's Equation may be found by choosing gauge-theory sources such that their double copy does not source the dilaton and/or the anti-symmetric tensor [57, 58]. Choosing gauge-theory sources that are then used in the double copy appears to bypass the need for a judicious choice a matter stress tensor as source for the gravity solution; however, prescribed properties of supergravity solutions and their corresponding sources undoubtedly translate into properties of gauge-theory sources. At the time of this writing, a complete dictionary has not yet been formulated.

CK duality as defined for scattering amplitudes in section 2, requires that external lines are on the free mass shell. Thus, in the tree expansion in figure 32 the duality can be expected to hold only up to terms that vanish if the sources obeyed free-field equations of motion. The discussion in section 2 then implies that such a feature leads to breaking of linearized gauge (diffeomorphism) invariance in the double-copy theory due to the presence of sources. This may be interpreted as the double-copy realization of the fact that gravity sources break diffeomorphism invariance. For the same reason, gravity field equations can be satisfied by a double-copy field configuration only up to terms proportional to the free equations of motion of the sources. Thus, for a comparison with a direct solution of supergravity equations of motion, such terms must be eliminated by field, coordinate and source redefinitions. This mirrors the backreaction of gravitational field on its source, illustrated in the previous subsection. It is not *a priori* obvious that gauge-theory classical solutions which differ by gauge transformations lead through the double copy to gravity solutions that differ by field redefinitions and coordinate transformations.

Perturbative spacetimes and their relation to perturbative solutions of the YM equations of motion were discussed in [58]. Below we outline their construction. As in the calculation of scattering amplitudes, we begin with the YM action (see equation (5)), whose equations of motion in the presence of sources are

$$\partial^{\mu} F^{a}{}_{\mu\nu} + g f^{abc} A^{b\mu} F^{c}{}_{\mu\nu} = \zeta^{a}_{\mu}, \tag{372}$$

where g is the coupling constant and the field-strength tensor $F^{a}{}_{\mu\nu}$ is

$$F^{a}{}_{\mu\nu} = \partial_{\mu} A^{a}_{\nu} - \partial_{\nu} A^{a}_{\mu} + g f^{abc} A^{b}_{\mu} A^{c}_{\nu}. \tag{373}$$

It is straightforward to include matter fields. Apart from their equations of motion, inclusion of the matter fields also gives specific expressions for the sources ζ. We will not discuss this possibility any further, choosing ζ to be non-dynamical and focusing on the gauge sector. The goal, following [58], is to solve perturbatively equation (372),

$$A^{a}_{\mu} = A^{(0)a}_{\mu} + g A^{(1)a}_{\mu} + g^{2} A^{(2)a}_{\mu} + \cdots, \tag{374}$$

and construct from it a solution of the double-copy theory.

The action for Einstein gravity coupled with a dilaton and an antisymmetric tensor, which is the double copy of two pure D-dimensional gauge theories, is given in equation (81). For the construction of a perturbative solution of its equations of motion, the fields are expanded as

$$h_{\mu\nu} = h^{(0)}_{\mu\nu} + \frac{\kappa}{2} h^{(1)}_{\mu\nu} + \left(\frac{\kappa}{2}\right)^2 h^{(2)}_{\mu\nu} + \cdots ,$$
$$B_{\mu\nu} = B^{(0)}_{\mu\nu} + \frac{\kappa}{2} B^{(1)}_{\mu\nu} + \left(\frac{\kappa}{2}\right)^2 B^{(2)}_{\mu\nu} + \cdots ,$$
$$\phi_{\mu\nu} = \phi^{(0)}_{\mu\nu} + \frac{\kappa}{2} \phi^{(1)}_{\mu\nu} + \left(\frac{\kappa}{2}\right)^2 \phi^{(2)}_{\mu\nu} + \cdots . \tag{375}$$

We can combine these different fields into a single field H. Since the asymptotic values of these fields are all obtained by projection from the tensor product of the two asymptotic gauge fields, it is convenient to have the field H which has this property at every order in κ. That is, in its expansion in κ,

$$H_{\mu\nu} = H^{(0)}_{\mu\nu} + \frac{\kappa}{2} H^{(1)}_{\mu\nu} + \left(\frac{\kappa}{2}\right)^2 H^{(2)}_{\mu\nu} + \cdots , \tag{376}$$

$H^{(n)}$ is the double copy of the nth order term in the expansion of the gauge-theory field. There are no cross terms between different orders in the vector field expansion (374). This is a consequence of the fact that different orders are given by different tree configurations in figure 32 and thus do not mix in the double copy.

On shell, at the linearized level and in the appropriate gauges[58] it is possible [58] to formulate the equations of motion in terms of a linear combination of the three fields:

$$H^{(0)}_{\mu\nu} = h^{(0)}_{\mu\nu} + B^{(0)}_{\mu\nu} + P^q_{\mu\nu} \phi^{(0)} . \tag{377}$$

In the absence of sources they are

$$\partial^\rho \partial_\rho H^{(0)}_{\mu\nu} = 0 . \tag{378}$$

A source modifies the right-hand side appropriately and must have the transversality and trace properties of $H^{(0)}_{\mu\nu}$. The field (376) has been referred to in [58] as the 'fat graviton', in contrast with the 'skinny graviton', $h_{\mu\nu}$. In equation (377) $P^q_{\mu\nu}$ is a projector, which depends on a fixed null vector q, defining the physical dilaton. In position space it is

$$P^q_{\mu\nu} = \frac{1}{D-2} \left(\eta_{\mu\nu} - \frac{q_\mu \partial_\nu + q_\nu \partial_\mu}{q \cdot \partial} \right) . \tag{379}$$

Conversely, the three physical fields can be extracted from $H^{(0)}$ by projection:

$$\phi^{(0)} = \eta^{\mu\nu} H^{(0)}_{\mu\nu} , \qquad B^{(0)}_{\mu\nu} = \frac{1}{2} \left(H^{(0)}_{\mu\nu} - H^{(0)}_{\nu\mu} \right) ,$$
$$h^{(0)}_{\mu\nu} = \frac{1}{2} \left(H^{(0)}_{\mu\nu} + H^{(0)}_{\nu\mu} \right) - P^q_{\mu\nu} H^{(0)\rho}{}_\rho . \tag{380}$$

8.2.1. Linearized solution. Following [58], to solve the YM equation (372) we choose the Lorenz gauge, $\partial^\mu A^a_\mu = 0$, and to leading order in the coupling the equation becomes

$$\partial^2 A^{(0)a}_\mu = \zeta^a_\mu . \tag{381}$$

Consistency with the gauge condition requires that ζ be transverse. To start instead with a scattering state it suffices to replace $\zeta^a_\mu \to \varepsilon_\mu c^a \partial^\nu \partial_\nu \exp(ip \cdot x)$ where c^a is some color wave

[58] The de Donder gauge for the graviton and the Lorentz gauge for the tensor field.

function and take the limit $p^2 \to 0$ at the end of calculations. To cover both options simultaneously one may denote the solution to this equation by $A_\mu^{(0)a}$ while not using its specific form, though in specific examples it may be necessary to be more specific.

Wave solutions of the YM free-field equations,

$$A_\mu^{(0)a} = \sum_j c_j^a \varepsilon_\mu^j(p)\, e^{ip\cdot x}, \tag{382}$$

with little-group indices j, $p\cdot\varepsilon^j = 0 = q\cdot\varepsilon^j$ and c_j^a color wave functions, can be straightforwardly double-copied to wave solutions of the free field equation of the action (81). In the absence of sources this is just a reorganization of the usual double copy of scattering states. The gravity solution,

$$H_{\mu\nu}^{(0)}(x) = h_{ij}\varepsilon_\mu^i(p)\,\varepsilon_\nu^j(p)\, e^{ip\cdot x}, \tag{383}$$

can be decomposed into the graviton, B-field and dilaton using (380). The constant factor h_{ij} is arbitrary and can be chosen such that the gravity and gauge-theory asymptotic waves have the same normalization. It can also be used to project out the B-field and dilaton and obtain a linearized solution of Einstein's equations, e.g.

$$c_j^a = c^a a_j\,, \qquad h_{ij} = a_i a_j\,, \qquad a\cdot a = 0\,, \tag{384}$$

where a_i are 'kinematic gauge-theory wave functions'.

In general, whether or not the B-field and dilaton can be turned off depends on the gauge-theory sources and on their relation to gravity sources. The relevant solutions of the (381), in position and momentum space, for an arbitrary source is

$$A_\mu^{(0)a}(x) \propto \int d^D y \frac{\zeta_\mu^a(y)}{|x-y|^{D-2}}\,, \quad \mathcal{F}\left[A_\mu^{(0)a}\right](p) = \frac{\mathcal{F}\left[\zeta_\mu^a\right](p)}{p^2}\,, \tag{385}$$

where $\mathcal{F}$ is the Fourier-transform operator. The rules for constructing the corresponding linearized gravity solution and sources are yet to be completely clarified. Here we attempt to formalize several possibilities, while leaving others for future development.

In identifying suitable relations between gauge and gravity sources it is important that the result can be interpreted as the linearized stress tensor of some field theory and thus that it conforms with energy conditions expected of such a stress tensor [53]. Not every possible construction has this property; indeed, it was shown in [53] that, while the source for Kerr–Schild solutions (whose linearized approximation is exact and will be discussed in some detail in section 8.3) can be obtained by specifying the charge distribution sourcing the corresponding gauge-theory solutions and imposing $\nabla_\mu T^\mu{}_\nu = 0$, they do not obey simultaneously the weak and strong energy conditions. While discussions of energy conditions have appeared in the literature (see below for references), a thorough analysis is currently absent and we will refrain from attempting one here. We emphasize that the classical double copy is best *defined* so that it yields a solution of (378); moreover, its sources should be constructed out of the gauge-theory sources such that they do not have any unphysical features. In general, it is necessary to verify whether the resulting source obey reasonable energy conditions before attempting to promote it from a non-dynamical source to a dynamical one, realized in terms of the fundamental fields of a quantum theory. These requirements may be used to identify some of the rules of the construction.

We begin with a source of the type

$$\zeta_\mu^a = c^a \zeta_\mu\,, \tag{386}$$

with constant color factor c^a and transverse ζ_μ. Even though c^a need not have any particular algebraic properties, it is natural to take at face value the fact that the solution (385) is given by the first Feynman diagram in figure 32 and apply the usual double-copy rules: $c^a \to \tilde{\zeta}_\mu$. Even though nonlinear corrections to a YM solution with this source vanish because $f^a{}_{bc}c^b c^c = 0$, nonlinear corrections to its gravity counterpart may be present; we shall see this explicitly in section 8.2.2. Similarly to scattering-amplitudes double copy, it is very important to *not* discard color factors that vanish due to the summation over the color indices. (Examples where this is crucial are found in [6, 69].)

Gauge-theory sources may exhibit a less transparent separation of color and kinematics, e.g.

$$\zeta(x)^a_\mu = \sum_i c^a_i \zeta(x)^i_\mu \,, \tag{387}$$

with several distinct independent color factors c^a_i and transverse (position-dependent) ζ^i_μ. Similarly to the case of asymptotic scattering states, we may still apply the (color factor) $\to$ (kinematic factor) replacement in momentum space with the same twist as in that case (and in the case of a wave solution) of allowing for a constant relative rotation of sources. Formally

$$\sum_i c^a_i \mathcal{F}\left[\zeta^i_\mu\right](p) \longrightarrow \sum_{i,j} h_{ij}\mathcal{F}\left[\zeta^i_\mu\right](p)\,\mathcal{F}\left[\tilde{\zeta}^i_\mu\right](p) \,, \tag{388}$$

where $\mathcal{F}$ is the Fourier-transform operator. In general it may be possible to allow h_{ij} to be a function of momentum; Lorentz invariance demands that it should be a function of p^2 and thus it can only lead to shifts of $H^{(0)}_{\mu\nu}$ by local functions. From this perspective, $h_{ij} \to h(p^2)_{ij}$ should be equivalent to field and/or coordinate redefinition in the gravity theory.

In both this case and in the simpler previous case (which may be obtained by taking the indices i and j to take a single value), the resulting space linearized solution is

$$H^{(0)}_{\mu\nu}(p) = \frac{\sum_{ij} h\left(p^2\right)_{ij}\mathcal{F}\left[\zeta^i_\mu\right](p)\,\mathcal{F}\left[\tilde{\zeta}^j_\nu\right](p)}{p^2} \,. \tag{389}$$

Comparing this the general solution of equation (378) with a source, we identify the numerator as the Fourier-transform of that source. Transforming back to position space, it follows that the gravity source is given by the convolution of the two YM sources with a kernel defined by the matrix h_{ij}:

$$\zeta_{\mu\nu}(x) = \int d^D y\, d^D z\, \mathcal{F}[h_{ij}](|x-y-z|)\, \zeta^i_\mu(y)\, \tilde{\zeta}^j_\mu(z) \,. \tag{390}$$

Such a relation between gauge and gravity sources was discussed in [54] and is reminiscent of the off-shell definition of the linearized fat graviton in equation (138).

Gauge transformations, whose linearized form is $A^a_\mu \to A^a_\mu + \partial_\mu \chi^a$, can map a solution such as (385) into one that has less straightforward identification of a momentum space 'kinematic numerator'. To explore this possibility let us assume that $A^a_\mu(x)$ has the general form[59]

$$A^{(0)a}_\mu(x) = \frac{1}{x^2}\sum_i c^a_i n^i_\mu(x) \,, \tag{391}$$

[59] Time-independent vector potentials, of the form $A^a_\mu(\vec{x}) = \sum_i c^a_i \mu^i(\vec{x})/|\vec{x}|$, can be treated similarly. The apparent difference in the engineering dimension between the expression of $A^a_\mu(\vec{x})$ here and that in equation (391) stems from the difference in the dimension of the measure of the three-dimensional and four-dimensional (inverse) Fourier-transform operator.

where $n(x)$ may contain terms that either eliminate the overall factor or introduce stronger singularities. The Fourier transform of this vector potential can be defined formally as

$$\mathcal{F}\left[A^{(0)a}_{\mu}\right](p) = \sum_{i} c^a_i \, \hat{n}^i_{\mu}(i\partial/\partial p)\, \mathcal{F}\left[\frac{1}{x^2}\right](p)\,, \tag{392}$$

where the operators $\hat{n}^i_{\mu}(i\partial/\partial p)$ are obtained from $n^i_{\mu}(x)$ by the formal replacement $x^{\mu} \mapsto i\partial/\partial p_{\mu}$. This operation is to be understood in the sense of distributions, i.e. the Fourier transform is taken in the presence of a test function that falls off sufficiently fast so integration by parts does not yield any boundary terms. Interpreting the operators $n^i_{\mu}(i\partial/\partial p)$ as the kinematic numerators, the linearized double copy may be defined as[60]

$$\mathcal{F}\left[H^{(0)}_{\mu\nu}\right](p) = \sum_{ij} h_{ij}\, \hat{n}^i_{\mu}(i\partial/\partial p)\, \hat{\tilde{n}}^j_{\nu}(i\partial/\partial p)\, \mathcal{F}\left[\frac{1}{x^2}\right](p)\,. \tag{393}$$

The commutation properties of the operators $\hat{n}$ and $\hat{\tilde{n}}$ together with the properties of h_{ij} determine whether or not this double copy yields a purely gravitational solution or the solution also contains nontrivial dilaton and/or anti-symmetric tensor. Fourier-transforming back to position space for a constant matrix h_{ij} suggests a (linearized) gravitational source (in de Donder gauge)

$$\zeta_{\mu\nu}(x) = \sum_{ij} h_{ij} n^i(x)\, \tilde{n}^j(x)\,. \tag{394}$$

See [54] for a further discussion on the relation of gauge and gravity sources in a time-dependent setting and [63] for examples where symmetries help identify the appropriate sources. The above construction is related to the position-space replacement rules of [51, 63].

A non-dynamical source can also be interpreted in the spirit of a (spontaneous) breaking of the gauge group and thus apply the corresponding double-copy rules discussed in section 5 together with the fact that the linearized solution (385) is given by the first Feynman diagram in figure 32. That is, the source is decomposed in irreducible representations of the unbroken (global part of the) gauge group and the double copy amounts to constructing gauge-invariant bilinears. This interpretation should also be subject to the consistency conditions discussed in section 5 regarding the spectrum of the double-copy theory.

Ultimately gravitational sources should be dynamical (we shall review this in section 8.4); as a step in this direction while eschewing the full dynamics of matter fields one may demand, as was done in [57], that the gauge-theory source obeys covariant current conservation,

$$D^{\mu}\zeta^a_{\mu} = 0\,. \tag{395}$$

Imposing it anticipates that ζ^a_{μ} can be realized in terms of some other fields, in a gauge invariant Lagrangian without settling on a specific realization.

The previous discussion and examples above refer to cases in which the sources of at least one of the two gauge theories are smooth functions, perhaps with compact support. If both momentum-space sources contain singular distributions their product requires a careful definition, especially if their product is ill-defined, such as a product of Dirac δ-functions. A physical perspective together with the expectation that there exists a Lagrangian that manifests the double-copy properties of equation (81) suggests a natural prescription. Because the momenta of the two gauge theories are identified through the double copy, it is natural that constraints

[60] This construction may *in principle* be generalized to $\hat{n} = \hat{n}(p, i\partial/\partial p)$. We leave this to the readers who read this footnote.

on it be imposed only once. Thus, if overlapping constraints are imposed by the gauge-theory sources, they should be included only once in the double copy of the source. It is perhaps interesting that this prescription yields identical (linearized) gravity solutions from distinct (linearized) gauge-theory solutions—e.g. two point-like sources present for all times *vs.* one point-like source present for all time and one instantaneous source.

To illustrate this, let us consider the field of a static point-like charge. The four-current is proportional to $u = (1,0,0,0)$ and the vector potential is [58]

$$A_\mu^{(0)a}(x) = gc^a u_\mu \frac{1}{4\pi r}, \quad A_\mu^{(0)a}(p) = gc^a u_\mu \frac{\delta^{(1)}(p^0)}{p^2}. \tag{396}$$

Consequently, $H^{(0)\mu\nu}$ is

$$H^{(0)\mu\nu}(p) = \frac{\kappa}{2} M u^\mu u^\nu \frac{\delta^{(1)}(p^0)}{p^2}, \tag{397}$$

which can be easily Fourier-transformed to position space. In writing this expression we made certain identifications between the gauge coupling and constants in the gravity theory. Since $H^{(0)\mu\nu}$ is symmetric, $b^{\mu\nu} = 0$; it is not traceless, so there is a nontrivial dilaton

$$\phi = H^{(0)\mu}{}_\mu = +\frac{\kappa}{2}\frac{M}{4\pi r}. \tag{398}$$

Using equation (380) and the projector (379), the correction to the metric is

$$h^{\mu\nu} = \frac{\kappa}{2}\frac{M}{4\pi r}\left(u^\mu u^\nu + \frac{1}{2}(\eta^{\mu\nu} - q^\mu l^\nu - q^\nu l^\mu)\right), \quad \text{with} \quad l = \frac{1}{r+z}(0,x,y,r+z). \tag{399}$$

Running a similar construction in the opposite direction, shock-wave solutions of Einstein's equations which are also solutions of linearized Einstein's equations were shown in [50] to be related, through a double-copy procedure, to certain wave solutions of YM theory. The relevant gravitational source $\zeta_{\mu\nu}$ is identified such that the scattering of some particle off a high-energy graviton is equivalent to all orders in perturbation theory to the scattering off $\zeta_{\mu\nu}$; the gravitational shock wave, given by the Aichelberg-Sexl [439], is the solution of (linearized) Einstein's equation with this source. The corresponding gauge-theory source ζ^a_μ was similarly constructed, i.e. such that the scattering of some particle off a high energy gluon is equivalent to all orders in perturbation theory to the scattering off ζ^a_μ. The source turned out to be of the type (386) and the gauge-theory shock wave is the solution of (linearized) YM equations with this source. The two waves are related by the usual color→kinematics replacement. By construction, scattering off the gravitational wave can also be obtained through this replacement from scattering off the gauge-theory wave, to all orders in perturbation theory.

8.2.2. Nonlinear corrections. With a linearized solution in hand, nonlinear corrections can be computed directly, by evaluating increasingly higher orders in the tree expansion in figure 32. The goal however it to explore the realization of nonlinear corrections to the gravity solutions as a double copy of the nonlinear corrections to the YM solutions. We will review this here, loosely following [58]. As we shall see, this comparison will emphasize the importance of the choice of fields, a feature that will be further discussed for complete solutions.

Nonlinear corrections to a linearized solution are expressed, through the tree expansion in figure 32, as convolutions of the linearized solution with kernels given by Feynman vertices. Since however, the gravity source depends on the metric it sources, one may either include explicitly such modifications (as $\mathcal{O}(\kappa^{n\geqslant 2})$ corrections to the source) or ignore them and obtain a solution for a choice of fields such that source changes are absent. These two perspectives have an analog in the two YM theories, where sources may either be corrected order by order in perturbation around the trivial solution such that they are e.g. covariantly constant, $D\cdot\zeta=0$, or they are fixed, respectively.

The first nonlinear correction to some solution $A_\mu^{(0)c}$ follows easily in terms of the standard three-point vertex. To utilize the same rules as for amplitudes double copy, it is convenient to present it in momentum space:

$$
\begin{aligned}
A^{(1)a\mu}(-p_1) = {} & \frac{i}{2p_1^2} f^{abc} \int \frac{d^D p_2}{(2\pi)^D} \frac{d^D p_3}{(2\pi)^D} (2\pi)^D \delta(p_1+p_2+p_3) \\
& \times \left[(p_1-p_2)^\gamma \eta^{\mu\beta} + (p_2-p_3)^\mu \eta^{\beta\gamma} + (p_3-p_1)^\beta \eta^{\gamma\mu} \right] A_\beta^{(0)b}(p_2) A_\gamma^{(0)c}(p_3)\,.
\end{aligned}
\tag{400}
$$

The factor in the square parenthesis is the usual kinematic part of the off-shell three-gluon vertex and has the same antisymmetry properties as the color factor. While this expression may be simplified somewhat by making use of the transversality of $A^{(0)a\mu}$, we will choose not to do so.

Taking two configurations like (400) and replacing the color factors of one with the kinematics of the other while leaving the propagators untouched (which, apart from using the same double-copy rules for amplitudes also includes the application of the results of the previous subsection $A_\mu^{(0)a}(p)\tilde{A}_\nu^{(0)b}(p) \to H_{\mu\nu}^{(0)}(p)$) leads to

$$
\begin{aligned}
H^{(1)\mu\mu'}(-p_1) = {} & \frac{1}{4p_1^2} \int \frac{d^D p_2}{(2\pi)^D} \frac{d^D p_3}{(2\pi)^D} (2\pi)^D \delta(p_1+p_2+p_3) \\
& \times \left[(p_1-p_2)^\gamma \eta^{\mu\beta} + (p_2-p_3)^\mu \eta^{\beta\gamma} + (p_3-p_1)^\beta \eta^{\gamma\mu} \right] \\
& \times \left[(p_1-p_2)^{\gamma'} \eta^{\mu'\beta'} + (p_2-p_3)^{\mu'} \eta^{\beta'\gamma'} + (p_3-p_1)^{\beta'} \eta^{\gamma'\mu'} \right] H_{\beta\beta'}^{(0)}(p_2) H_{\gamma\gamma'}^{(0)}(p_3)\,.
\end{aligned}
\tag{401}
$$

This expression has the same structure as the first nonlinear correction to the solutions of the equations of motion of the action (81) except that the trilinear interaction of gravitons, B fields and dilatons was replaced by the factorized integrand kernel above. This factorization is the same as that of the three-point amplitudes from equation (81). It can be seen explicitly by starting from the complete three-point vertices and using transversality *and* the on-shell condition for the external states. While $H_{\gamma\gamma'}^{(0)}$ is transverse by construction, it obeys, in general, a free-field equation with a source. Thus, $H^{(1)\mu\mu'}(-p_1)$ given above represents the first correction to a gravity solution for the choice of a fluctuations such that the trilinear vertex is free of terms that vanish on the free mass shell. This vertex is related to the one following from the expansion of the Lagrangian by a field redefinition.

There exists further freedom in the relation between $H^{(1)\mu\mu'}(-p_1)$ and the fluctuations of the metric, B field and dilaton. At the linearized level the later are given by the decomposition (377). For higher-order corrections however this decomposition may be modified. As discussed in the beginning of this section, the amplitudes double copy guarantees only that the asymptotic states—or linearized solutions—double copy. At higher orders in κ there may exists further terms in the relation between gauge-theory and gravity fields which are projected out when the LSZ reduction is applied to a Green's function. At the first nonlinear order this is

$$
H_{\mu\nu}^{(1)} = h_{\mu\nu}^{(1)} + B_{\mu\nu}^{(1)} + P_{\mu\nu}^q \phi^{(1)} + \mathcal{T}_{\mu\nu}^{(1)}\left(h^{(0)}, b^{(0)}, \phi^{(0)}\right), \tag{402}
$$

and at arbitrary order

$$H^{(n)}_{\mu\nu} = h^{(n)}_{\mu\nu} + B^{(n)}_{\mu\nu} + P^{q}_{\mu\nu}\phi^{(n)} + \mathcal{T}^{(n)}_{\mu\nu}\left(h^{(m)}, b^{(m)}, \phi^{(m)}, m < n\right). \tag{403}$$

Such terms may be interpreted as field redefinitions connecting the initial choice of gravity fields (375) to the ones 'chosen' by the double copy. They also capture various choices that can be made during the calculation, such as gauge choices and—highlighted by their appearance in the first nonlinear correction—use of the free/lower order equations of motion in the definition of vertices. Terms of this type may be eliminated by nontrivial choices of the kernel in equation (389). These 'transformation functions' [58] may be determined by comparing the perturbative solution of the equations of motion of the action (81) with the result of the double copy. The main physical information they contain is that they provide the connection between the fields natural from a double-copy perspective and the natural fluctuations in the gravity Lagrangian. In the special case of the self-dual theory, it is known how to choose a parametrization of the metric perturbation such that the double copy is manifest [43]. For these field variables $\mathcal{T}_{\mu\nu} = 0$ to all orders in the tree diagram expansion of self-dual spacetimes.

An example illustrating this discussion and dramatically emphasizing the relevance of the choice of field variables was given in [58] using the linearized gravity solution in equation (397) and its gauge-theory counterpart in equation (396). This example also emphasizes the importance of *not* dropping terms whose color factors vanish *after* summation over color indices. The first nonlinear correction $H^{(1)}$ to equation (397) was obtained in [58]; it is

$$H^{(1)}_{\mu\nu}(x) = -\left(\frac{\kappa}{2}\right)^2 \frac{M^2}{4\,(4\pi\,r)^2}\hat{r}_\mu \hat{r}_\nu\,, \tag{404}$$

where $\hat{r}_\mu = (0, \mathbf{x}/r)$. It turns out that a nontrivial transformation function is necessary to turn $H = \eta + \kappa H^{(0)} + \kappa^2 H^{(1)}$ into a solution of the equations of motion to $\mathcal{O}(\kappa^2)$ in the variables (375). It is given by [58]

$$\begin{aligned}\mathcal{T}^{(1)\mu\nu}(-p_1) = \int \frac{d^D p_2}{(2\pi)^D}\frac{d^D p_3}{(2\pi)^D}\frac{(2\pi)^D\delta(p_1+p_2+p_3)}{4p_1^2}&\{H^{(0)}_{2\,\alpha\beta}H^{(0)\alpha\beta}_{3}p_1^\mu p_1^\nu + 8p_2^\alpha H^{(0)}_{3\,\alpha\beta}H^{(0)\beta(\mu}_{2}p_1^{\nu)}\\ &+ 8p_2\cdot p_3\, H^{(0)\mu\alpha}_{2}H^{(0)\nu}_{3\ \ \alpha} - 2\eta^{\mu\nu}p_2\cdot p_3\, H^{(0)}_{2\,\alpha\beta}H^{(0)\alpha\beta}_{3} + 4\eta^{\mu\nu}p_2^\alpha H^{(0)}_{3\,\alpha\beta}H^{(0)\beta\gamma}_{2}p_{3\gamma}\\ &+ P^{\mu\nu}_{q}[2(D-6)p_2\cdot p_3\, H^{(0)}_{2\,\alpha\beta}H^{(0)\alpha\beta}_{3} - 4(D-2)p_2^\alpha H^{(0)}_{3\,\alpha\beta}H^{(0)\beta\gamma}_{2}p_{3\gamma}]\}\,,\end{aligned} \tag{405}$$

where we used the shorthand notation

$$H^{(0)}_{i\,\mu\nu} \equiv H^{(0)}_{\mu\nu}(p_i)\,, \quad \text{and} \quad p^{(\mu}q^{\nu)} \equiv \frac{1}{2}(p^\mu q^\nu + p^\nu q^\mu)\,. \tag{406}$$

This first transformation function $\mathcal{T}^{(1)\mu\nu}$ holds for all cases that have symmetric and transverse $H^{(0)}_{\mu\nu}$ and $h^{(0)}_{\mu\nu}$.

Since equation (396) is an exact solution of the YM equations of motion, one may wonder whether it is possible that it has some other, physically equivalent form which can be double-copied to an exact solution of dilaton-axion-gravity in some field variables. To this end, it is necessary that the first correction to this equivalent form of equation (396) vanishes before summation over color indices. We shall see in section 8.3 that this is indeed possible.

Proceeding to higher orders is in principle straightforward, but quite tedious in practice. The new features compared to the discussion above relates to the need of a representation of the corrections to the YM equations which manifest CK duality up to terms that are projected out by the LSZ reduction. Since the only difference between the asymptotic states of scattering amplitudes and $A^{a(0)}_\mu$ is that the latter obey an on-shell condition with sources, CK duality can be satisfied only up to such terms. Similarly to scattering amplitudes, a generic perturbative classical solution is related to one that exhibits the duality (in this restricted sense) by generalized gauge transformations. As in that case, such transformations are not always easy to find. As in that case, a Lagrangian whose Feynman rules lead to manifestly CK-dual representation or the use of the generalized double-copy construction can alleviate this issue.

To quintic order in fields, the Lagrangian in [41] provides the requisite Feynman rules to obtain the gauge-theory perturbative classical solution in a form that can be double copied directly. This was exploited in [58], where the second nonlinear correction was discussed. As explained there, the quartic YM vertex does not contribute to a symmetric double copy and the second term in the perturbative solution of YM equations is given entirely in terms of the three-point vertex:

$$
\begin{aligned}
A^{(2)a\mu}(-p_1) = {} & \frac{i}{p_1^2} f^{abc} \int \frac{d^D p_2}{(2\pi)^D} \frac{d^D p_3}{(2\pi)^D} (2\pi)^D \delta(p_1 + p_2 + p_3) \\
& \times \left[(p_1 - p_2)^\gamma \eta^{\mu\beta} + (p_2 - p_3)^\mu \eta^{\beta\gamma} + (p_3 - p_1)^\beta \eta^{\gamma\mu} \right] A^{(0)b}_\beta(p_2) A^{(1)c}_\gamma(p_3) \,.
\end{aligned} \tag{407}
$$

It leads to the second correction $H^{(2)}$ in the gravitational solution

$$
\begin{aligned}
H^{(2)\mu\mu'}(-p_1) = {} & \frac{1}{2p_1^2} \int \frac{d^D p_2}{(2\pi)^D} \frac{d^D p_3}{(2\pi)^D} (2\pi)^D \delta(p_1 + p_2 + p_3) \\
& \times \left[(p_1 - p_2)^\gamma \eta^{\mu\beta} + (p_2 - p_3)^\mu \eta^{\beta\gamma} + (p_3 - p_1)^\beta \eta^{\gamma\mu} \right] \\
& \times \left[(p_1 - p_2)^{\gamma'} \eta^{\mu'\beta'} + (p_2 - p_3)^{\mu'} \eta^{\beta'\gamma'} + (p_3 - p_1)^{\beta'} \eta^{\gamma'\mu'} \right] H^{(0)}_{\beta\beta'}(p_2) H^{(1)}_{\gamma\gamma'}(p_3) \,.
\end{aligned} \tag{408}
$$

The graviton, antisymmetric tensor and dilaton components can be easily extracted using the projectors; to connect this general expression to a solution with specific sources in specific coordinates $\mathcal{T}^{(2)}$ must be computed as well. We refer to [58] for details.

Exercise 8.1. Explore the possibility of using a quasi-classical solution obtained by folding scattering amplitudes in BCJ representation against external sources to construct solutions for the gravity field equations. This is equivalent to removing terms proportional to the free-field equations from Green's functions and using the result to construct an ansatz for a classical solution. The resulting double-copy field configuration should be correct—for some choice of field variables—up to terms that are proportional to the free field equations, i.e. up to field redefinitions.

Steps towards the double copy of nonlinear classical solutions beyond second order were taken in [172, 434] for the special case of perturbiners or Berends-Giele currents. Starting from the perturbiners of certain effective field theories [434] and F^3 and F^4-deformed YM theory, perturbiners of the corresponding gravity theories were constructed using the KLT relations. Because only one leg of the Berends-Giele current is off shell, the relation between the objects thus constructed and the 'true' gravitational perturbiner is simpler than in the most general case: it consists only of a gauge transformation and involves no field redefinition.

The need for a Lagrangian yielding CK-satisfying Feynman rules or, more generally, of Green's functions manifesting CK duality on all of their internal lines may be circumvented through the generalized double-copy construction discussed in section 7. Generalizing slightly to Green's functions, the starting point is any general perturbative expressions for the gauge-theory solutions expressed in terms of cubic diagrams; quartic vertices, if present, are resolved in the usual way. Because of lack of manifest CK duality, their double copy does not yield solutions of the equations of (81) up to field redefinitions. The formulae discussed in section 7 provide the correction terms. As in the examples discussed earlier in this section, transformation functions are probably necessary to relate the result of the generalized double copy to a solution in some chosen coordinates. It remains an open problem to have an *a priori* understanding of the choice of fields in the gravitational theory that set all transformation functions to zero.

8.3. Complete solutions; Kerr–Schild coordinates

In the discussion of perturbative construction of gravity solutions in section 8.2 we encountered, following [58], linearized solutions which are exact solutions of YM equations—such as that in equation (396)—which double copy to linearized solutions of gravity which receive higher-order corrections. While, as emphasized there, this can be understood as a consequence of the special properties of the color factors of the YM solution, it is important to understand whether there exists a choice of field variables for which these contributions to not arise at all and consequently the transformation functions vanish identically to all orders in classical perturbation theory. The general expectation is that if a gauge-theory solution does not receive corrections beyond nth order in perturbation theory, then its corresponding gravity solution will also be exact beyond that order.

As pointed out in [51], following [440], a particular ansatz for the metric linearizes the source-free Einstein's equations and thus can potentially give these metrics as double copies of solutions of YM Equations which do not receive nonlinear corrections. They are known as Kerr–Schild metrics; the ansatz is given in terms of a scalar function ϕ (which is *not* the dilaton) and a vector k which is null and geodesic with respect to the background metric $\bar{g}_{\mu\nu}$:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa h_{\mu\nu} \equiv \bar{g}_{\mu\nu} + \kappa\phi k_\mu k_\nu\,, \qquad \bar{g}_{\mu\nu} k^\mu k^\nu = 0\,, \qquad \left(k\cdot\bar{\nabla}\right) k_\mu = 0\,. \tag{409}$$

The background (or fiducial) metric $\bar{g}_{\mu\nu}$ is also used to raise and lower indices on the metric fluctuation h and $\bar{\nabla}^\mu$ is the corresponding background-covariant derivative. One component of k can be set to unity, thus absorbing its dynamics in ϕ. The Kerr–Schild form is special in that the metric perturbation—or the graviton—explicitly decomposes into a direct product of the vector k_μ with itself. The remarkable property of this ansatz is that it linearizes the Ricci tensor and reduces Einstein's equations to a single nontrivial relation between the function ϕ and the source. The components of the Ricci tensor are

$$R^\mu{}_\nu = \bar{R}^\mu{}_\nu + \kappa\left[-h^\mu{}_\rho\bar{R}^\rho{}_\nu + \frac{1}{2}\bar{\nabla}_\rho\left(\bar{\nabla}_\nu h^{\mu\rho} + \bar{\nabla}^\mu h^\rho{}_\nu - \bar{\nabla}^\rho h^\mu{}_\nu\right)\right], \tag{410}$$

where $\bar{R}^\mu{}_\nu$ is the Ricci tensor associated with the background metric $\bar{g}_{\mu\nu}$. We emphasize that the linear dependence on the metric fluctuation h in equation (409) holds only for the index positions in equation (410).

A simple choice of background metric is $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$ (with a mostly-minus signature), used at length in this context in [51]. For this choice the background-covariant derivatives become regular derivatives. Further choosing $k^0 = 1$, the components of the Ricci tensor are

$$\begin{aligned}
R^0{}_0 &= \frac{1}{2}\partial^i\partial_i\phi\,,\\
R^i{}_0 &= -\frac{1}{2}\partial_j\left[\partial^i\left(\phi k^j\right) - \partial^j\left(\phi k^i\right)\right],\\
R^i{}_j &= \frac{1}{2}\partial_l\left[\partial^i\left(\phi k^l k_j\right) + \partial_j\left(\phi k^l k^i\right) - \partial^l\left(\phi k^i k_j\right)\right],\\
R &= \partial_i\partial_j\left(\phi k^i k^j\right).
\end{aligned} \tag{411}$$

All Latin indices run over the space-like directions. Thus, the scalar function ϕ is determined by a Poisson-type equation.

A generalization of the Kerr–Schild ansatz in equation (409) is the double-Kerr–Schild ansatz [441], which is given in terms of two scalar functions and two null, geodesic and mutually orthogonal vectors:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa h_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa \left(\phi k_\mu k_\nu + \psi l_\mu l_\nu\right) , \\ k^2 = l^2 = k \cdot l = 0 , \qquad \left(k \cdot \bar{\nabla}\right) k_\mu = 0 , \qquad \left(l \cdot \bar{\nabla}\right) l_\mu = 0 , \tag{412}$$

where as before $\bar{g}_{\mu\nu}$ is a background metric and is used in all index contractions and background-covariant derivatives $\bar{\nabla}$. For these field variables the Ricci tensor is

$$R^\mu{}_\nu = \bar{R}^\mu{}_\nu + \kappa \left[-h^\mu_\rho \bar{R}^\rho{}_\nu + \frac{1}{2} \bar{\nabla}_\rho \left(\bar{\nabla}_\nu h^{\mu\rho} + \bar{\nabla}^\mu h^\rho_\nu - \bar{\nabla}^\rho h^\mu{}_\nu \right) \right] + R^\mu{}_{\nu,\mathrm{non-lin.}} , \tag{413}$$

$$R^\mu{}_{\nu,\mathrm{non-lin.}} = -\frac{\kappa^2}{2} [\frac{1}{2} \bar{\nabla}^\mu h(k)^\rho{}_\delta \bar{\nabla}_\nu h(l)^\delta{}_\rho + h(l)^{\mu\delta} \bar{\nabla}_\rho \bar{\nabla}_\nu h(k)^\rho{}_\delta \\ + \bar{\nabla}_\rho (h(l)^{\rho\delta} \bar{\nabla}_\delta h(k)^\mu{}_\nu + 2h(l)^\rho{}_\delta \bar{\nabla}_{(\nu} h(k)^{\mu)}{}_\delta - 2h(l)^{\mu\delta} \bar{\nabla}^{[\rho} h(k)^{\delta]}{}_\nu)] + (k \leftrightarrow l), \tag{414}$$

where

$$h(k)_{\mu\nu} = \phi k_\mu k_\nu , \qquad h(l)_{\mu\nu} = \psi l_\mu l_\nu . \tag{415}$$

The linearity of Einstein's equations in Kerr–Schild variables implies that any single Kerr–Schild metric can also be thought of as a double Kerr–Schild metric.

In higher dimensions further generalizations are possible, involving up to $D-2$ null, geodesic and mutually orthogonal vectors with the same properties as k and l. Additionally, it was argued in [441] that in the so-called Plebansky coordinates, the nonlinear part of the Ricci tensor, $R^\mu_{\nu,\mathrm{non-lin.}}$, vanishes identically and solutions of the linearized Einstein's equations are also exact solutions.

8.3.1. Kerr–Schild exact solutions. In this section we shall review the double-copy interpretation of the Schwarzschild solution, emphasizing its realization vis-à-vis the discussion in the previous section. We will then summarize and comment on generalizations of this approach to other spacetimes.

The Kerr–Schild form of the Schwarzschild solution is (see e.g. [442])

$$g_{\mu\nu} = \eta_{\mu\nu} + \frac{\kappa^2}{8\pi} \frac{M}{r} k_\mu k_\nu , \tag{416}$$

where M is positive and the null four-vector k is chosen such that the line element is rotationally invariant:

$$k^\mu = \left(1, x^i/r\right) , \qquad r^2 = \sum_{i=1}^{3} x^i x_i . \tag{417}$$

The double-copy form of this solution was discussed at length in [51]. With the definition of the linearized double copy for singular sources discussed in section 8.2.1 and up to identification of parameters, it can be seen that the departure of the metric (416) from Minkowski space is given by the double copy of

$$A^a_\mu = \frac{g c^a k_\mu}{4\pi r} . \tag{418}$$

It can be straightforwardly verified that A^a_μ satisfies Maxwell's equations with a point-like source at the origin. Equation (418), however, is not the standard potential of such a source. Rather, as shown in [51], it is related to the standard potential of a point-like charge (418) by a gauge transformation with parameter Λ^a:

$$\begin{aligned}
&A^a_\mu = A'^a_\mu + \partial_\mu \Lambda^a\,,\\
&A'^a_\mu = \frac{g c^a u_\mu}{4\pi r}\,, \qquad u_\mu = (1,0,0,0)\,, \qquad \Lambda^a = \frac{g c^a}{8\pi}\log r^2\,,\\
&j^a_\mu = -g c^a u_\mu \delta^{(3)}(\mathbf{x})\,.
\end{aligned} \tag{419}$$

It is interesting to contrast the two gauge-equivalent vector potentials A^a_μ and A'^a_μ. Both are proportional to a single color vector c^a and because of this they both are formally exact solutions of the nonlinear YM equations. For the latter one, A', the vanishing color factors of the corrections are multiplied by nontrivial kinematic dependence and thus, as discussed in section 8.2.2, there are nonlinear corrections that should be included which are crucial for transforming its double copy into a solution of the full Einstein's equations. For the former, A, one can check that the vanishing color factors come together with vanishing kinematic dependence. Therefore, the corrections to the double copy of two A vectors also vanish and thus the linearized double copy does not receive nonlinear corrections. This underscores the importance of the gauge choice for the gauge-theory solutions that participate in the classical double copy.

To recover the Schwarzschild solution the parameters of the two theories are replaced as

$$\frac{\kappa}{2} \leftrightarrow g\,, \qquad M \leftrightarrow |c|\,. \tag{420}$$

We note that the norm of the color vector c^a, which may be identified as the charge of the source under the sole Cartan generator of the gauge group that is nontrivial, corresponds to the mass of the Schwarzschild black hole. This seems to suggest a relation between the uniqueness of the Coulomb-like solution and Birkhoff's theorem.

It is interesting and important to note that, despite the gauge-theory solutions being sourced by the same charge distributions and being gauge-equivalent, their classical double copies as defined here are inequivalent. Indeed, while the solution constructed in this section has only a nontrivial metric, the one constructed perturbatively in section 8.2.2 stating from equation (396) also has a nontrivial dilaton which cannot be removed while preserving a nontrivial metric. With the current understanding of the classical double copy, the fact that gauge-equivalent gauge-field configurations lead to inequivalent gravitational field configurations appears to be an unavoidable feature. At this juncture, it seems best to start with valid gravitational solutions and work backwards to gauge theory.

Considerations similar to the ones outlined above have been used to give a double-copy interpretation to the Kerr back hole, black brane solutions, shock-wave and plane-wave solutions [51] and to the (anti) de Sitter spaces in [52]. In the latter cases the cosmological constant is related to the charge density of a uniform charge distribution. Gravity solutions with additional matter fields turned on, such as the Taub-NUT space, have a double Kerr–Schild form and, as argued in [52], have a double-copy interpretation (in the same sense as discussed above) in terms of a dyon solution whose electric and magnetic charges are related to the mass and the NUT charge.

Kerr–Schild solutions with time-dependent sources, describing accelerating black holes, have been discussed in [54] where a relation was constructed between the electromagnetic radiation of an accelerating charge and the gravitational radiation of an accelerating point mass and thus represents an effective description of the complete vacuum solution. The contraction of the corresponding sources with gluon and graviton polarization vector/tensor gives the amplitude for the Bremsstrahlung process. Other gravitational wave solutions, including vacuum solutions, which are of the double-Kerr–Schild type, were discussed in [74].

All YM solutions that appeared in the constructions reviewed here are also solutions of Maxwell's equations. Using the fact, discussed in section 5, that YM theory can be interpreted as a double copy of itself with a theory of a bi-adjoint scalar field, more complicated solutions can be constructed by taking the double copy of e.g. a Maxwell solution with a solution of the bi-adjoint scalar theory. This observation was explored in [52, 62], while solutions of bi-adjoint scalar theory were constructed in [55, 73] and [61]. The details pertaining to the relation between the sources of various solutions remain to be fully worked out. This perspective also makes contact with the off-shell Lagrangian double copy of [56, 255].

Following [53], the gravitational stress tensor of certain double-copy Kerr–Schild solutions was expressed linearly in terms of the current sourcing the gauge-theory solution. With this relation, in most cases they are not stress-energy tensor of a perfect fluid and contains shear stresses and, moreover, they do not obey the weak-energy condition. It is possible that other choices of coordinates and field variables display double-copy behavior that simultaneously map YM solutions to gravitational ones and satisfy the energy conditions.

Further generalizations, involving a nontrivial fiducial metric $\bar{g}$ in the Kerr–Schild ansatz, were discussed in [52, 62, 63]. As discussed in [52, 62], if the fiducial metric is of Kerr–Schild type, then every such solution can also be interpreted as a (multiple) Kerr–Schild metric with Minkowski space as fiducial metric (referred to as Type A constructions in [62]). Among the examples discussed are the de Sitter and anti de Sitter generalizations of the Schwarzschild black hole.

Solutions with a non-Kerr–Schild background metric have been discussed in [62] (referred to there as Type B constructions) and in [63]. They are realized in terms of solutions of gauge theory on a space with the fiducial metric. The classical scale invariance of YM theories implies that, for a fiducial metric is conformally Minkowski, the gauge theory is effectively in flat space (up to a curvature-dependent scalar mass term). Examples of this type were discussed in [62]. Apart from black holes in asymptotically maximally-symmetric spaces which are also treated in this framework, [63] also gives double-copy interpretations to black strings, black branes, and various types of gravitational waves. The corresponding localized sources for the YM and scalar theories, for both stationary and time-dependent examples, are also identified and examples are given in terms of Kerr–Schild vectors k.

While a coherent picture for the classical double copy of exact gauge-theory solutions to exact (matter-coupled) gravity solutions is still to be formulated, the examples discussed in the literature and summarized here give hope that such a relation may be generically possible.

8.3.2. Good and bad coordinates: charged black holes from higher dimensions. To further illustrate the importance of the choice of field variables for the interpretation of the result of classical double-copy constructions as exact solutions of Einstein's equations (perhaps coupled to additional matter) let us briefly discuss the charged black hole solution in the presence of an additional scalar field. (See also [62] for a discussion of charged black holes). The equations of motion are standard[61]

[61] The corresponding action is in the string frame, and may be mapped to the Einstein frame by a rescaling of the five-dimensional metric.

$$
\begin{aligned}
R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R &= \frac{1}{2}\phi^2 \left(g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} - \frac{1}{4} g_{\mu\nu} F \cdot F \right) + \frac{1}{\phi} \left(\nabla_\mu \nabla_\nu \phi - g_{\mu\nu} \nabla \cdot \nabla \phi \right), \\
\nabla \cdot \nabla \phi &= \frac{1}{4} \phi^3 F \cdot F, \\
\nabla^\alpha F_{\alpha\mu} &= -3 \frac{\nabla^\alpha \phi}{\phi} F_{\alpha\mu},
\end{aligned} \tag{421}
$$

and can be obtained by Kaluza–Klein reduction from Einstein's equations in five dimensions through the usual ansatz

$$
g_5 = \begin{pmatrix} g_4 + \phi^2 A \otimes A & \phi^2 A \\ \phi^2 A & \phi^2 \end{pmatrix}. \tag{422}
$$

As we shall see, in these field variables the charged black hole solution does not a clear classical double-copy interpretation; we will identify the field variables in which it does, paralleling the smooth relation of double copy between theories related by dimensional reduction.

The four-dimensional charged black hole can be obtained via Kaluza–Klein reduction from a five-dimensional black string. In Kerr–Schild form, it is

$$
g = \eta + \varphi \hat{k} \otimes \hat{k}, \qquad \eta^{\mu\nu} \hat{k}_\mu \hat{k}_\nu = 0, \qquad \varphi = \frac{M}{r_3}, \tag{423}
$$

where r_3 is the radial coordinate in the three coordinates transverse to the string. A suitable solution of the constraints constants defining $\hat{k}$ is that it is a boost of the vector $(1, \hat{r}_3, 0)$ where $\hat{r}_3$ is the unit vector in three dimensions orthogonal to the string:

$$
\hat{k} = (\gamma, \hat{r}_3, \beta\gamma) \equiv (k, \beta\gamma), \qquad \gamma^2 = \frac{1}{1-\beta^2}. \tag{424}
$$

Using the reduction ansatz in equation (422), the four-dimensional fields are:

$$
\begin{aligned}
g_4 &= \eta + \frac{\varphi}{1 + \beta^2\gamma^2\varphi} k \otimes k, \qquad k = (\gamma, \hat{r}_3), \qquad k^2 = 1 - \gamma^2 = -\frac{\beta^2}{1-\beta^2}, \\
\phi &= \sqrt{1 + \beta^2\gamma^2\varphi}, \qquad A = \frac{\beta\gamma\varphi}{1 + \beta^2\gamma^2\varphi} k.
\end{aligned} \tag{425}
$$

It is not difficult to check that this field configuration is a solution of equation (421).

It is also not difficult to see that this field configuration departs from the Kerr–Schild ansatz in that the vector k defining the departure of the metric from Minkowski space is time-like rather than null. Moreover, the dependence on φ suggests that all fields are given by a nontrivial resummation of tree diagrams.

Another choice of field variables,

$$
g_5 = \begin{pmatrix} \tilde{g}_4 & \tilde{A} \\ \tilde{A} & 1 + \tilde{\phi} \end{pmatrix}, \tag{426}
$$

which is closely related to the dimensional reduction of asymptotic states of scattering amplitudes, is more suitable for a classical double-copy interpretation. Indeed, the four-dimensional fields are

$$
\begin{aligned}
\tilde{g}_4 &= \eta + \varphi k \otimes k, \qquad k = (\gamma, \hat{r}_3), \qquad k^2 = 1 - \gamma^2 = -\frac{\beta^2}{1-\beta^2}, \\
\tilde{\phi} &= \beta^2\gamma^2\varphi, \qquad \tilde{A} = \varphi k,
\end{aligned} \tag{427}
$$

which are related in the sense described in section 8.2.1 to the following solution of gauge theory coupled to a scalar field:

$$A^a_{\mathrm{YM}} = \varphi k c^a \,, \qquad \phi^a_{\mathrm{YM}} = \mathcal{N} \varphi \, c^a \,, \tag{428}$$

where c^a is some color vector.

We note that the field configuration (427) is not a solution of equation (421), but it is a solutions of the equations obtained from them through the field redefinition mapping the fields in equation (422) to those in equation (426). Moreover, while structurally similar to the Kerr–Schild ansatz, it is not of the same type because the vector k is time-like. The violation of the null condition is compensated by the contribution of the vector and scalar fields. While the relation between (428) and (427) is linear, the fact that k is not null allows in principle for a nonvanishing kinematic part in the nonlinear corrections to (428) and thus to potential nonlinear corrections to their double copy, cf section 8.2.2. The fact that (427) is an exact solution suggests absence of the nonlinear corrections to the scalar and vector fields in equation (428). These features may allow further generalization of the classical double-copy interpretation of solutions of Kerr–Schild type. An alternative construction of the charged dilatonic black hole solution discussed here, which uses the standard four-dimensional equations of motion (421) and a generalization of the double-Kerr–Schild ansatz (412) which also includes certain internal dimensions, was discussed in [443].

Exercise 8.2. Show that the first nonlinear correction to (427) vanishes by evaluating it in terms of the kinematic factors of the corrections to the gauge-theory solution (428).

8.4. Radiation

Earlier in this section we have reviewed and illustrated various possible definitions of the double copy of classical solutions gauge theories to solutions of Einstein's equations coupled perhaps with additional matter and summarized the existing results. One of the fundamental results of general relativity (and, in fact, of any gravity theory) is the emission of gravitational waves—classical gravitational radiation emitted in processes involving massive astrophysical bodies such as neutron stars or black holes, perhaps with macroscopic intrinsic angular momentum.

From the perspective of general relativity such calculations have a long history, with numerical and perturbative results in various approximation schemes, which we will not review here; see [129, 130, 444] for reviews. They have been stunningly confirmed through the direct experimental detection of gravitational waves by the LIGO and Virgo collaborations [127]. We expect that the double-copy approach to such calculations will lead to important technical simplifications and bring new insight into these problems.

The first nontrivial contribution to the radiation process involves five particles: the two incoming and outgoing massive bodies and the outgoing graviton. To evaluate this, it is necessary to fix a model for the massive bodies that can be included in the double-copy construction. In [57] they were represented in terms of gauge fields, effectively as the linearized solutions in equation (396). The classical double copy then (effectively) yields a gravity solution whose linearized form is (397) and thus the massive objects being scattered source both gravitons and dilatons. It moreover appears that the double-copy rules used in [57] assume that a Lagrangian that manifests CK duality is available. Indeed, the gauge-group generators are replaced with the kinematic dependence of the off-shell three-point vertex, thus assuming that the latter have the same algebraic properties as the former. While for a Lagrangian that manifests CK duality this replacement is, of course, equivalent to the usual rules in section 2, for a general

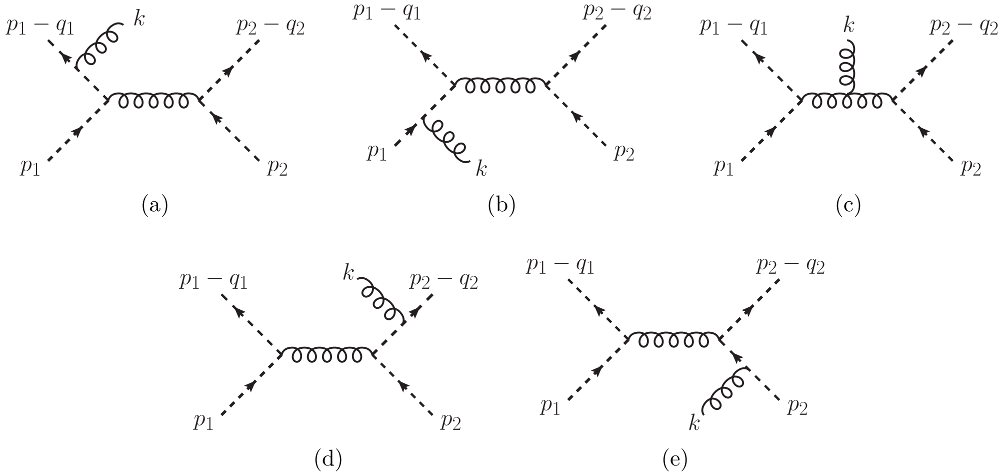


Figure 33. The five cubic diagrams for inelastic scalar scattering with gluon production in gauge theory. The legs carrying momenta p_1 and p_2 are incoming and the remaining ones are outgoing.

Lagrangian, however, further terms may be necessary. Nevertheless, the result reproduces direct calculations [57, 59] in dilaton-coupled gravity. Similar techniques have been used to obtain the corresponding results in Einstein-YM theory [445].

In a different approach, suggested in [424] based on earlier ideas of [446, 447], incoming and outgoing spinless massive bodies are represented as double copies of minimally-coupled massive scalar Φ. The Lagrangian is

$$\mathcal{L} = -\frac{1}{2}\mathrm{Tr}F^{\mu\nu}F_{\mu\nu} + \sum_i \left[(D_\mu\Phi_i)^\dagger (D^\mu\Phi_i) - m_i^2\,|\Phi|^2\right], \tag{429}$$

with the scalar field in some (complex) representation of the gauge group and D_μ the corresponding covariant derivative. Then, a certain classical limit is taken to ensure that, as for classical particles, masses are parametrically larger than their spatial momenta. In this approach one can choose the couplings of these particles such that, on the one hand, CK duality is present and on the other their double copy does not yield a dilaton source. Due to Birkhoff's theorem, this model is sufficient describe the gravitational wave emission far from the horizon of black holes, where the large (classical) masses ensures that the linearized emission is captured accurately. We outline the relevant calculation, following [424].

The five diagrams contributing to the scattering process $\Phi_i\Phi_j \to \Phi_i\Phi_j h_{\mu\nu}$ are shown in figure 33 and the corresponding amplitude is

$$\mathcal{A} = -i\left(\frac{n_\mathrm{a}c_\mathrm{a}}{D_\mathrm{a}} + \frac{n_\mathrm{b}c_\mathrm{b}}{D_\mathrm{b}} + \frac{n_\mathrm{c}c_\mathrm{c}}{D_\mathrm{c}} + \frac{n_\mathrm{d}c_\mathrm{d}}{D_\mathrm{d}} + \frac{n_\mathrm{e}c_\mathrm{e}}{D_\mathrm{e}}\right). \tag{430}$$

The denominators $D_\mathrm{a},\ldots,D_\mathrm{e}$ and the color factors $c_\mathrm{a},\ldots,c_\mathrm{e}$ are easily read from the diagrams in figure 33, taking into account that the scalar Φ_i is in some complex representation R_i with generators $T^a_{R_i}$. The kinematic numerators follow from the Feynman rules of the Lagrangian (429)[62]. They are:

[62] We note that, as discussed in previous sections, a quartic scalar term is not necessary for CK duality because the scalar fields are taken in a complex representation of the gauge group.

$$
\begin{aligned}
n_{\mathrm{a}} &= (2p_1+q_2)\cdot(2p_2-q_2)\ \varepsilon\cdot(2p_1+2q_2) - \left(2p_1\cdot q_2+q_2^2\right)\ \varepsilon\cdot(2p_2-q_2)\,, \\
n_{\mathrm{b}} &= (2p_1-k-q_1)\cdot(2p_2-q_2)\ 2\varepsilon\cdot p_1 + 2p_1\cdot k\,\varepsilon\cdot(2p_2-q_2)\,, \\
n_{\mathrm{c}} &= (2p_1-q_1)^{\mu}(2p_2-q_2)^{\rho}\left[(k+q_2)_{\mu}\eta_{\nu\rho} + (q_1-q_2)_{\nu}\eta_{\rho\mu} - (k+q_1)_{\rho}\eta_{\mu\nu}\right]\varepsilon^{\nu}\,, \\
n_{\mathrm{d}} &= (2p_1-q_1)\cdot(2p_2+q_1)\ \varepsilon\cdot(2p_2+2q_1) - \left(2p_2\cdot q_1+q_1^2\right)\ \varepsilon\cdot(2p_1-q_1)\,, \\
n_{\mathrm{e}} &= (2p_1-q_1)\cdot(2p_2-k-q_2)\ 2\varepsilon\cdot p_2 + 2p_2\cdot k\,\varepsilon\cdot(2p_1-q_1)\,,
\end{aligned}
\tag{431}
$$

where ε is the gluon polarization vector. The color identities that are important for the gauge invariance of $\mathcal{A}$ in equation (430) are

$$
c_{\mathrm{a}} - c_{\mathrm{b}} = c_{\mathrm{c}} \quad c_{\mathrm{d}} - c_{\mathrm{e}} = c_{\mathrm{c}}\,. \tag{432}
$$

It can be easily checked that the numerators (431) obey the corresponding kinematic relations.

The double-copy amplitude follows from the usual rules, see section 2:

$$
\mathcal{M} = -i\left(\frac{n_{\mathrm{a}}n_{\mathrm{a}}}{D_{\mathrm{a}}} + \frac{n_{\mathrm{b}}n_{\mathrm{b}}}{D_{\mathrm{b}}} + \frac{n_{\mathrm{c}}n_{\mathrm{c}}}{D_{\mathrm{c}}} + \frac{n_{\mathrm{d}}n_{\mathrm{d}}}{D_{\mathrm{d}}} + \frac{n_{\mathrm{e}}n_{\mathrm{e}}}{D_{\mathrm{e}}}\right). \tag{433}
$$

The tensor product of the two outgoing gluon polarization vectors can be projected onto a graviton state. For internal lines a more involved projection is necessary [188]. We shall return to it shortly.

To relate the amplitude just constructed to the classical scattering of massive bodies it is necessary to focus on the classical kinematic regime. There exists many ‘classical limits’ of a field theory and all of them involve the limit of vanishing Planck’s constant, which must therefore be restored (on dimensional grounds) in the field theory expressions. The limit we are interested in is also the one in which masses and other quantum numbers, such as external momenta and charges, are parametrically large compared to the momenta exchanged between particles. Thus, the classical limit is equivalent with a large mass expansion[63] [421]:

$$
m_i \to \frac{m_i}{\hbar}\,, \qquad g \to \frac{g}{\hbar}\,, \qquad \hbar \to 0\,, \qquad p_i^{\mu} \to m_i v_i^{\mu}\,, \qquad v_i^2 = 1\,. \tag{434}
$$

Because the coupling (charges) and masses are scaled simultaneously, this limit makes parts of tree-level and loop-level diagrams of the same order and consequently all such contributions enter nontrivially in this classical limit [421].

The limit (434) must be taken while enforcing the exact on-shell condition for all external particles. In particular

$$
(p_i-q_i)^2 = m_i^2 - 2m_i v_i\cdot q_i + q_i^2 = m_i^2 \Rightarrow 2m_i v_i\cdot q_i = q_i^2\,. \tag{435}
$$

Thus, if external momenta are parametrically larger than the exchanged ones, this equation can be satisfied only if

$$
v_i\cdot q_i \sim \mathcal{O}\left(m_i^{-1}\right)\,. \tag{436}
$$

This condition must be enforced when taking the classical limit of equation (433). Defining the variables

$$
\begin{aligned}
P_{12}^{\mu} &\equiv k\cdot v_1\, v_2^{\mu} - k\cdot v_2\, v_1^{\mu}, \\
Q_{12}^{\mu} &\equiv (q_1-q_2)^{\mu} - \frac{q_1^2}{k\cdot v_1}v_1^{\mu} + \frac{q_2^2}{k\cdot v_2}v_2^{\mu},
\end{aligned}
\tag{437}
$$

[63] Other formulations of the classical limit, leading to the same result but with a different physical reasoning, were discussed in [78, 80, 82].

this classical limit of the amplitude (433) is

$$\mathcal{M}_{\rm cl} = -16 i m_1^2 m_2^2\, \varepsilon_{\mu\nu} \left[4\frac{P_{12}^\mu P_{12}^\nu}{q_1^2 q_2^2} + 2\frac{v_1\cdot v_2}{q_1^2 q_2^2}\left(Q_{12}^\mu P_{12}^\nu + Q_{12}^\nu P_{12}^\mu\right) \right. \\ \left. + (v_1\cdot v_2)^2 \left(\frac{Q_{12}^\mu Q_{12}^\nu}{q_1^2 q_2^2} - \frac{P_{12}^\mu P_{12}^\nu}{(k\cdot v_1)^2 (k\cdot v_2)^2} \right) \right], \tag{438}$$

where $\varepsilon_{\mu\nu} \equiv \varepsilon_\mu \varepsilon_\nu$. As pointed out in [424], there is a close relation between this amplitude and the metric perturbation (i.e. radiation field) constructed in [57]. The metric perturbation is given by the Fourier transform to position space of $\mathcal{M}_{\rm cl}$ with respect to the incoming scalar momenta, subject to the constraints imposed by the on-shell conditions for all external momenta. We note that the mass of the particles enters only as an overall factor in the amplitude (438) and, consequently, in the associated metric perturbation. This property, implying that the features of the metric are essentially independent of a (spinless) source, may be interpreted physically as a reflection of Birkhoff's theorem.

To obtain the analogous results in Einstein's gravity theory it is necessary to project out the dilaton and antisymmetric tensor field from all diagrams. As reviewed in section 5 following [188], this can be done by introducing further 'ghost' fields whose couplings are adjusted such that they remove the (un)desired degrees of freedom[64]. For the case at hand the relevant Lagrangian is [424]

$$\mathcal{L} = -\frac{1}{2}{\rm Tr} F_{\mu\nu}F^{\mu\nu} + {\rm Tr} D^\mu \chi D_\mu \chi \\ + \sum_i \left[(D_\mu \Phi_i)^\dagger D^\mu \Phi_i - m_i^2 \Phi_i^\dagger \Phi_i - 2X m_i \Phi_i^\dagger \chi \Phi_i \right], \tag{439}$$

where D_μ is the gauge-covariant derivative, χ is the adjoint ghost, X is its coupling to be determined. The mass factors are included such that the ghost field has canonical dimension and X is dimensionless. While in the adjoint representation, the ghost field is allowed to double copy only with itself.

The unknown coupling X can be determined by comparing the $2 \to 2$ massive scalar scattering obtained through double copy from the Lagrangian (439) with the massive scalar scattering in scalar-coupled general relativity. The result is

$$X^4 = \frac{1}{D-2}, \tag{440}$$

where D is the spacetime dimension. The Lagrangian (439) can then be used to evaluate the additional Feynman diagrams that remove the dilaton and axion contribution to equation (438). The complete amplitude in scalar-coupled pure gravity is

$$\mathcal{M}_{\rm GR} = -16 i m_1^2 m_2^2\, \varepsilon_{\mu\nu} \left[4\frac{P_{12}^\mu P_{12}^\nu}{q_1^2 q_2^2} + 2\frac{v_1\cdot v_2}{q_1^2 q_2^2}\left(Q_{12}^\mu P_{12}^\nu + Q_{12}^\nu P_{12}^\mu\right) \right. \\ \left. + \left((v_1\cdot v_2)^2 - \frac{1}{D-2} \right) \left(\frac{Q_{12}^\mu Q_{12}^\nu}{q_1^2 q_2^2} - \frac{P_{12}^\mu P_{12}^\nu}{(k\cdot v_1)^2 (k\cdot v_2)^2} \right) \right], \tag{441}$$

[64] An alternative possibility is to carry out the double copy while keeping track of the helicity of internal fields and making sure that only graviton modes appear on all internal lines of diagrams [80, 82].

where the polarization tensor $\varepsilon_{\mu\nu}$ now includes only graviton degrees of freedom. As shown in [424], this reproduces the result of the (far more complicated) direct computation in general relativity coupled to point particles.

It is interesting to note that, repeating the calculation in [57, 59] such that the sources correspond to massive spinless bodies and removing the dilaton and axion contribution through a procedure similar to the one described above appears to lead [424] to a different result than (441). While the origin of the difference is not clear, it is possible that they are due to the unusual double-copy rules employed in [57, 59].

Exercise 8.3. Evaluate the amplitude for the graviton production in the scattering of massive charged spinless bodies YME theory and compare the result with that of [445].

8.5. Further comments

The close relation between Green's functions and scattering amplitudes of QFTs suggests that relations between scattering amplitudes of different theories may translate, in particular gauges and for special choices of field variables, into relations between classical solutions of the corresponding equations of motion. In this section we reviewed at length examples in which this expectation is realized and certain solutions of gauge theories can be used to construct, through a classical double copy, certain solutions of gravity theories. Important points—such as the relation between the gauge choice for the gauge-theory solution and the properties of the corresponding gravity solution, or the identification of the best choice of gravity field variables such that no transformation functions are present—remain to be fully understood and the complete rules of the classical double-copy construction to be spelled out. The examples we discussed, as well as the additional ones that may be found in the literature, show that such an approach can have useful applications to current problems in gravitational physics. Chief among them is precision predictions of gravitational waves; as we saw in section 8.4, the (classical) double-copy construction may help streamline the evaluation of the expected signal from the relevant astronomical events.

Further applications of the double copy to gravitational wave physics, which we did not discuss in detail, relate to the calculation of gravitational interaction potential in the post-Newtonian expansion. While standard methods, using the gravitation Lagrangian, are well-developed and results through fourth post-Newtonian order are available [448–451], double-copy calculations such as in [452–454] may bring a novel perspective to this problem. Indeed, advances based on the double copy and new developments [78] in the effective field approach [422, 455] resulted in a new state of the art result at the third order in Newton's constant [80].

A common feature of the classical solutions constructed to date through such methods is that, in the appropriate field variables, Einstein's equations become linear. This includes the case of the Kerr black hole which was shown in [456] to be related to a certain complex deformation of the Coulomb potential. It is also shown that the change in momentum in a scattering event, known as 'the impulse', can be described via a double copy of a point charge. Progress towards further understanding some of the rules of the classical double copy may follow from finding examples where nonlinear contributions are nonvanishing. Perhaps the easiest approach to exploring such cases is the analysis of a Kerr–Schild solution for another choice of field variables; an example would be to repeat the calculations in [436, 437] using modern approaches.

The construction of a gravitational Lagrangian whose fields are explicitly constructed in terms of those of the two single-copy gauge theories may also lead to new ways of relating the solutions of the two theories. The linearized Lagrangians of certain $\mathcal{N} = 2$ supergravities have been organized in this fashion in [56, 255].

Another (notoriously difficult) problem which may benefit from the existence of a systematic classical double copy is gravitational perturbation theory around a curved ground state such as the (anti) de Sitter space, or the Schwarzschild black hole. The double copy will likely relate it to gauge-theory perturbation theory around a nontrivial classical solution of YM equations of motion. Attempts in this direction have been discussed in [60] and [72] where, respectively, the three- and four-point amplitudes of gravitons around a gravitational plane wave were expressed in terms of three- and four-gluon amplitudes around a particular gauge-theory plane wave. While developing general methods for such calculations is an interesting problem in its own right, it is likely that their main applications will be to gauge/string duality.

The study of the classical double copy is in its infancy and many avenues remain to be explored; we expect that the resulting methods will yield important new progress in gravitational physics, especially on the problem of gravitational radiation.

9. Conclusions

The duality between color and kinematics and the double-copy construction offer a radically-different perspective on gravity theories compared to traditional Lagrangian or Hamiltonian approaches. For the well-studied case of scattering amplitudes, the duality provides powerful means for converting results in gauge theory to those of gravity. This has led to progress in studying the behavior of various gravitational theories at high perturbative orders, such as the UV behavior of extended supergravity at four and five loops [38, 291–293] and the third post-Minkowskian corrections to the classical Hamiltonian for compact binaries [80, 82]. At present there are no other means to evaluate such high orders.

Remarkably, the idea of CK duality and of the double-copy structure extends to theories with no obvious connection to gauge or gravity theories, as reviewed in section 5. The fact that the scattering amplitudes of theories whose Lagrangians seem to have little to do with each other contain the same kinematical objects is rather striking and points to new nontrivial constraints shared by consistent theories. Additionally, by now the duality and double copy have been established for a large number of examples of classical solutions [50–77].

There are several areas where further progress would be welcomed. For example, it is not at the moment clear how far the notion of CK duality and the double copy can be carried beyond scattering amplitudes. Many of the examples of classical solutions that display the duality make use of special properties, such as the existence of Kerr–Schild forms of the metric. It would be very important to find more general examples. Classical solutions are inherently more difficult to study because they depend on coordinate and gauge choices and, without the appropriate choices, the double-copy structure is obscured. This may be contrasted with scattering amplitudes, which are independent of the choice of gauge and, to a large extent, field variables, making it much easier to formulate double-copy relations. To avoid carrying out complicated case-by-case analyses, a key step is to find underlying principles for choosing gauges and field variables in both single- and double-copy theories that make it more straightforward to identify relations between off-shell quantities. It would also be interesting to see if the more invariant color-trace-based formulation of the duality [48, 157–161] might shed light on extensions of CK duality beyond scattering amplitudes.

A possible path to unraveling the principles for choosing gauges and field variables may be the study of correlation functions. The computation of correlation functions of gauge-invariant operators in gauge theories may be approached through generalized unitarity [457], which relates it to the construction of tree-level scattering amplitudes and form factors of the same gauge-invariant operators. In this respect, CK duality has been formulated and used for the form factors of certain operators in four- and five-loop calculations in $\mathcal{N}=4$ SYM theory [9, 17, 18, 22]. Therefore, it seems plausible to extend CK duality to the correlation functions of these operators. However, a puzzle arises if one considers the natural step of constructing the double copy of such correlation functions. Since correlation functions of gauge-invariant operators in gauge theories are gauge invariant, one may conclude following the discussion in sections 2 and 4 that the corresponding gravitational correlation functions are automatically diffeomorphism-invariant. It is well-known however that local diffeomorphism-invariant operators do not exist in gravitational theories. Since correlators of gauge-dependent operators depend on choices of field variables, it appears that the gravitational correlation functions obtained though the double copy should be understood as being given for a particular choice of field variables and perhaps also for particular choice of gauge. A further puzzle originates from contrasting the results of the double copy for conformal gauge theories that admit a string-theory dual to the results of the corresponding string theory in anti-de-Sitter space. On the one hand, gauge-theory correlation functions are given by string-theory correlators with prescribed boundary conditions in anti-de-Sitter space; on the other, the gauge-theory correlators can be used to construct correlators in a gravitational theory (*not* a string theory) in a Minkowski vacuum and with the same amount of supersymmetry as the AdS one. While technically difficult, it would clearly be interesting to understand the implications of such a relation.

Apart from formal developments such as the ones described above, perturbative calculations in curved spacetime are playing an increasingly-important role in the current development of our understanding of the Universe. Initial attempts to use the double-copy construction in this context, involving calculations in certain plane wave spacetimes, have been discussed in [60, 62, 72, 458, 459]. As in flat space, the double-copy construction may help by relating such calculations with simpler ones in gauge theory, especially for spacetimes which are themselves classical double copies. It is obviously nontrivial to extend the insights of flat-spacetime scattering to the many conceptual and technical challenges posed by cosmological correlators in de Sitter, yet there is already a developing program [460–465] leveraging the identification of S-matrix elements emerging as residues of well-defined singularities of such quantities.

While CK duality and the associated double copy have been crucial for uncovering the UV properties of various supergravities [33, 34, 36, 38, 218, 293] and for identifying a new set of nontrivial enhanced UV cancellations [292], to move forward it is essential to gain a thorough grasp on the structures or symmetries that are responsible for the appearance of the latter. Progress in this direction has been reported in [34, 40, 466], but much more remains to be done to have a satisfactory understanding. Presumably, the duality and double copy play a key role in these cancellations.

Another important topic is to expand the web of theories related by the duality and double-copy construction. As illustrated in figure 17 of section 5, theories that may appear to be unrelated are bound together by double-copy relations. In many of these cases, the connection is rather obscure from a Lagrangian perspective, e.g. that DBI theory has a relation to the special Galileon theory, by sharing the NLSM as a composite theory via the double copy. A crucial open question is whether it is possible to get a complete classification of all double-copy-constructible theories. An equally important question is whether all supergravity theories can be expressed in a double-copy format [121, 240]. In this review, we discuss a large number of examples, which are collected in tables 4 and 5. It is a surprising fact that the only known unitary UV completions of gravity and higher-dimensional YM, the closed and open superstring, require their constituent effective field theories to be compatible with the field-theory adjoint double-copy to all orders of α' at tree-level [169, 171, 310, 355].

Another basic research direction is to find the underlying algebra behind the duality between color and kinematics. A natural expectation is that the kinematic Jacobi identities are due to an infinite-dimensional Lie algebra [43, 45, 47, 49]. Indeed, for the case of self-dual field configurations, corresponding to amplitudes with identical helicity, the algebra has been identified as that of the area-preserving diffeomorphisms in one lightcone and one transverse direction [43]. However, extending this observation to general helicity or field configurations has proven to be challenging.

Constructing a Lagrangian that automatically generates Feynman rules that manifest the duality would greatly help with finding double-copy relations between classical solutions. However, at present, only perturbative order-by-order constructions of such Lagrangians are known [41, 42, 150, 151]. From the perspective of gravity theories, Lagrangians that display the required factorization of Lorentz indices [94] have been obtained to all orders [95, 467]. However, as yet, it is unclear which all-orders gauge-theory Lagrangians can reproduce them though double copy. One difficulty is that such Lagrangians would likely contain an infinite number of auxiliary fields to make them local.

To further streamline higher-loop computations would be particularly desirable. While there has been enormous progress in carrying out such computations to relatively high orders (see e.g. [6, 38, 293, 468] for four and five-loop calculations) we should always strive to go further. At high orders, it can be nontrivial to find representations of loop integrands that manifest CK duality [15, 415]. As described in section 7, these difficulties can be bypassed via a generalized double copy [218, 416] that can be used to convert any representation of gauge-theory amplitudes to corresponding gravity ones, relying only on the proven existence of the duality at tree level. Finding generalizations of this procedure for any number of loops or legs would be important.

Strengthening connections between the double copy and other advances in scattering amplitudes would also be advantageous. In particular, the amplituhedron [469] gives novel geometric descriptions of amplitudes. A detailed formulation has been given for the planar sector of $\mathcal{N} = 4$ SYM theory. Making contact with CK duality requires extending these results to the nonplanar sector. Evidence suggests that this may be possible [195, 470].

The double copy seems to hint at some kind of interpretation of gravitons as composed of spin-1 particles. Of course, these cannot be any kind of naive bound states, which are forbidden by the Witten-Weinberg theorem [471]. Still, the double copy strongly suggests that gravitons and gluons ultimately belong together, presumably along the lines realized by string theory. (See [472–474] for steps in this direction.) Understanding any fundamental physical implications of the way gauge and gravity theories are intertwined by the double copy is a key problem that deserves further attention.

The application of the double copy to gravitational-wave physics [127] is currently the subject of intense investigation, specifically regarding the post-Newtonian [475] and post-Minkowskian [476, 477] approaches to the inspiral phase of binary mergers (see the following reviews for details and [129–132]). A nontrivial application of CK duality to the study of gravitational radiation has been discussed with a worldline formulation in [57], where the duality has been established through next-to-leading order [69]. Related progress was also reported in [59, 64, 65, 445]. Other investigations related to gravitation-wave physics that directly draw from scattering-amplitudes methods can be found in [54, 79, 422–426, 478–481]. A systematic and scalable approach for obtaining high-order corrections to conservative two-body potentials in the post-Minkowskian framework was presented in [78]. This has been successfully used to find the third post-Minkowskian corrections [80, 82], starting from two-loop amplitudes obtained via the double copy. It is noteworthy that this is one order beyond previous calculations [423, 482, 483]. While their impact on improving templates for LIGO/Virgo is currently under study [81], these results should also offer new insights into the general structure of high-order two-body Hamiltonians.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

We thank T Adamo, L Borsten, E Bjerrum-Bohr, J Bourjaily, L Dixon, M Duff, A Edison, S Ferrara, M Günaydin, S He, E Herrmann, Y-T Huang, G Kälin, R Kallosh, D Kosower, A Luna, D. Lüst, C Mafra, G Mogull, R Monteiro, S Nagy, H Nicolai, A Ochirov, D O'Connell, J Parra-Martinez, L Rodina, O Schlotterer, C-H Shen, S Stieberger, F Teng, J Trnka, A Tseytlin, P Vanhove, I Vazquez-Holm, C White and S Zekioglu for helpful discussions, collaboration and comments during the course of writing this review. Z B is supported by the U.S. Department of Energy (DOE) under Grant No. DE-SC0009937. JJMC is grateful for the support of Northwestern University, CEA/CNRS-Saclay, and the European Research Council under ERC-STG-639 729, *Strategic Predictions for Quantum Field Theories*. R R is supported by the U.S. Department of Energy (DOE) under Grant No. DE-SC0013699. The research of M C and H J is supported by the Knut and Alice Wallenberg Foundation under Grants KAW 2013.0235 and KAW 2018.0116 - *From Scattering Amplitudes to Gravitational Waves*, the Ragnar Söderberg Foundation (Swedish Foundations' Starting Grant), and the Swedish Research Council under Grant 621-2014-5722. This work was performed in part at the Munich Institute for Astro- and Particle Physics (MIAPP) of the DFG cluster of excellence 'Origin and Structure of the Universe' and in part at the Aspen Center for Physics, which is supported by National Science Foundation Grant PHY-1607 611. In addition, Z B thanks Mani L Bhaumik for generous support over the years.

Appendix A. Notation and list of acronyms

In this appendix, we summarize our notation and conventions for the reader's convenience.

Throughout the review, we denote with $\mathcal{A}_n(1,\cdots,n)$ gauge-theory color-dressed amplitudes, while $A_n(1,\cdots,n)$ is used to indicate color-stripped partial amplitudes. In some sections, where theories containing different fields are discussed, it is convenient to adopt the notation

$$\mathcal{A}_n\left(1\Phi_1,\ldots,n\Phi_n\right)\,, \tag{442}$$

which makes explicit which field is associated to each external leg of the amplitude. Superamplitudes are denoted with the same symbol as the corresponding amplitudes, i.e. it should be clear from the context whether $A_n(1,\cdots,n)$ and $\mathcal{A}_n(1,\cdots,n)$ refer to an amplitude or to the corresponding superamplitude. Writing the S-matrix as $S=1+iT$, our amplitudes correspond to the iT term, i.e. they give the output of the Feynman-diagram calculation. Amplitude in a gravitational theory are denoted as $\mathcal{M}_n(1,\cdots,n)$. For notational simplicity, we set $\kappa=2$ in most formulas, where κ is the gravitational coupling constant.

Our conventions for the phase in the BCJ representation of gauge-theory and gravity amplitudes follows the one in [156] and departs from the original BCJ papers [1, 2]. With this choice, YM numerators are real when written in terms of polarization vectors.

Calculations presented in this review involve a spacetime metric of mostly-minus signature. Our spinor-helicity conventions are obtained from the ones of [91] by the minimal replacement

$$(\eta_{\mu\nu})_{\mathrm{E\&H}} \to -(\eta_{\mu\nu})_{\mathrm{our}}\ . \tag{443}$$

In particular, angle and square brackets are the same as in [91].

Gauge-group fundamental fields are represented with high indices and anti-fundamental fields are represented with low indices. For example, the generator for the fundamental representation is written as

$$(t^a)_i^{\ j} \equiv t^a_{i\bar{\jmath}}\ . \tag{444}$$

Generators are normalized as

$$\mathrm{Tr}\left(t^a t^b\right) = \frac{\delta^{ab}}{2}\ , \tag{445}$$

and obey commutation relations of the form

$$\left[t^a, t^b\right] = i f^{abc} t^c\ . \tag{446}$$

In amplitude calculations it is convenient to rescale the group-theory generators and structure constants as

$$T^a \equiv \sqrt{2} t^a\ , \quad \tilde{f}^{abc} \equiv i\sqrt{2} f^{abc}\ , \tag{447}$$

so that we have the identity

$$\mathrm{Tr}\left(T^a T^b\right) = \delta^{ab}\ . \tag{448}$$

In particular, color factors entering the formula (35) are written in terms of T^as and $\tilde{f}^{abc}$s, i.e. are written in terms of hermitian objects carrying a factor of $\sqrt{2}$ with respect to the Feynman-rule normalization. When we encounter fields in a matter (non-fundamental) representation $\mathcal{R}$, we denote the corresponding generators as $t^a_{\mathcal{R}}$. In some sections of this review, for example in section 5, we frequently use hatted indices for gauge-group indices of the gauge theories entering the double-copy construction to differentiate them from global indices.

Finally, we conclude this appendix with a list of acronyms commonly used throughout the review:

UV	Ultraviolet
IR	Infrared
KLT	Kawai–Lewellen–Tye (formula/relations)
CK duality	Color/kinematics duality
BCJ	Bern–Carrasco–Johansson (duality/relations)
YM	Yang–Mills
SYM	Super-Yang-Mills
NLSM	Nonlinear Sigma model
DDM	Del Duca–Dixon–Maltoni (amplitude representation)
POP	Partially-ordered permutations
BCFW	Britto–Cachazo–Feng–Witten (relations/recursion in field theory)
KK	Kleiss–Kuijf (amplitude relations)
QCD	Quantum chromodynamics
CPT	Charge-Parity-Time reversal (transformations)
MHV	Maximally-helicity-violating (amplitudes)
LSZ	Lehmann–Symanzik–Zimmermann (reduction)
1PI	One particle irreducible (effective action)
BMS	Bondi-Metzner-Sachs (transformations)
SUSY	Supersymmetry (tables and figures only)
CSG	Conformal supergravity
BLG	Bagger–Lambert–Gustavsson (theory)
ABJM	Aharony–Bergman–Jafferis–Maldacena (theory)
VEV	Vacuum expectation value
$\mathrm{YM_{DR}}$	Yang-M-ills-scalar theory from dimensional reduction
YME	Yang–Mills–Einstein (theory)
DBI	Dirac–Born–Infeld (theories)
CHY	Cachazo–He–Yuan (formalism, also known as scattering equations)
SG	Supergravity (tables and figures only)
MZVs	Multiple zeta values
QFT	Quantum field theory
LIGO	Laser interferometer gravitational-wave observatory

Appendix B. Spinor helicity and on-shell superspaces

In explicit expressions for amplitudes, such as those in section 6 or appendix C, it is very convenient to adopt a helicity (circular polarization) basis for the asymptotic states of gluons or gravitons. In this appendix, we summarize the spinor-helicity formalism [88, 89, 484–488], which offers a convenient Lorentz covariant formalism for describing helicity, leading to remarkably compact expressions for scattering amplitudes. The resulting states fit naturally into on-shell supermultiplets [489].

B.1. Basics of spinor helicity

The spinor-helicity formalism expresses the positive- and negative-helicity polarizations of gluons (vectors) in terms of massless Weyl spinors

$$\varepsilon^{+}_{\mu}(k;q) = \frac{\langle q|\sigma_\mu|k]}{\sqrt{2}\langle qk\rangle}\,, \qquad \varepsilon^{-}_{\mu}(k;q) = \frac{[q|\sigma_\mu|k\rangle}{\sqrt{2}[kq]}\,, \tag{449}$$

where q is an arbitrary null 'reference' momentum which can be chosen independently for each external state of amplitudes and, because of gauge invariance[65], drops out of the final expressions. We use the compact notation

$$\langle ij\rangle \equiv \frac{1}{2}\bar{u}(k_i)(1+\gamma_5)u(k_j)\,, \qquad [ij] \equiv \frac{1}{2}\bar{u}(k_i)(1-\gamma_5)u(k_j)\,,$$
$$\langle q|\sigma_\mu|k] \equiv \frac{1}{2}\bar{u}(q)\gamma^\mu(1-\gamma_5)u(k)\,, \qquad [q|\sigma_\mu|k\rangle \equiv \frac{1}{2}\bar{u}(q)\gamma^\mu(1+\gamma_5)u(k)\,, \tag{450}$$

with $u(k)$ following the standard textbook notation for solutions of the Dirac equation [100]. The spinors $|i\rangle$ and $|i]$ transform in the $(\mathbf{1},\mathbf{2})$ and $(\mathbf{2},\mathbf{1})$ representations of the four-dimensional Lorentz group, respectively. The spinor products (450) are antisymmetric in their arguments. An important identity is the Schouten identity:

$$\langle ij\rangle\langle kl\rangle = \langle il\rangle\langle kj\rangle + \langle ik\rangle\langle jl\rangle\,, \tag{451}$$

which is a consequence of the vanishing of all three-index antisymmetric tensor with each index taking two values. The spinor products (450) are related to the usual scalar products by

$$\langle ij\rangle[ji] = 2k_i\cdot k_j = s_{ij}\,, \tag{452}$$

where the k_i are null four momenta. The Fierz identity is in our conventions is

$$\langle i|\sigma^\mu|j]\,\langle k|\sigma_\mu|l] = 2\langle ik\rangle[lj]\,. \tag{453}$$

Helicity amplitudes (that is, amplitudes with polarization vectors or tensors in helicity notation) can be given a manifestly crossing symmetric representation. To this end it is necessary to assign all momenta to have the same orientation, either all outgoing or all incoming. When switching between the two different orientations the helicity label is reversed. This is, of course, natural: since the helicity measures the projection of the spin on the momentum, changing the orientation of the momentum reverses the helicity. Using spinor helicity we can obtain exceptionally compact expressions for gauge-theory scattering amplitudes [88].

The physical graviton polarization tensors in helicity notation are direct products of the gluon ones,

$$\varepsilon^+_{\mu\nu}(k;q) = \varepsilon^+_\nu(k;q)\,\varepsilon^+_\nu(k;q)\,, \quad \varepsilon^+_{\mu\nu}(k;q) = \varepsilon^+_\nu(k;q)\,\varepsilon^+_\nu(k;q)\,. \tag{454}$$

Their tracelessness , $\varepsilon^{+\mu}_\mu = 0$, follows from the Fierz identity (453) with the appropriate choice of spinors:

$$\langle q|\sigma_\mu|k]^2 = 0\,, \qquad [q|\sigma_\mu|k\rangle^2 = 0\,. \tag{455}$$

The relation (454) between graviton and gluon polarizations is the simplest manifestation of the double copy. Of course, the double copy holds for the full nonlinear theory, not just for polarization tensors.

Loop calculations require regularization; we will not discuss details of this issue here except to note that to maximize the benefits of the spinor-helicity formalism, which is intrinsically four-dimensional, it is necessary to choose a compatible version of dimensional regularization [391, 490].

[65] Linearized gauge transformations, $\varepsilon_\mu(p)\to\varepsilon_\mu(p)+f(p)p_\mu$, is realized as shifts of the spinors associated to the reference vector, $|q\rangle \to |q\rangle + f(p)|p\rangle$, *etc*.

Exercise B.1. Starting with equations (10)–(13) apply equation (449) to obtain four-gluon amplitudes for the various helicity configurations. How clean can you make these expressions? See, for example, [88].

B.1.1. Massive spinor helicity. As some of the theories described in this review involve massive fields, we will briefly outline how to adapt the spinor-helicity formalism to this case. A first possibility is to write a massive momentum k in terms of two massless momenta [491]:

$$k = k_{\perp} + \frac{m^2}{2k\cdot q}q\,, \tag{456}$$

where $k_{\perp}$ is massless and we have also introduced a massless reference momentum q. Polarizations for massive vectors are then written as

$$\varepsilon_{+}^{\mu}(k;q) = \frac{\langle q|\sigma^{\mu}|k_{\perp}]}{\sqrt{2}\langle q\,k_{\perp}\rangle}\,,\quad \varepsilon_{-}^{\mu}(k;q) = \frac{[q|\sigma^{\mu}|k_{\perp}\rangle}{\sqrt{2}[k_{\perp}\,q]}\,,\quad \varepsilon_{0}^{\mu}(k;q) = \frac{1}{m}\left(k_{\perp}^{\mu} - \frac{q^{\mu}}{2k\cdot q}\right)\,, \tag{457}$$

where the first two physical polarizations reproduce (449) in the massless limit and $\varepsilon_0^{\mu}(k;q)$ gives the longitudinal polarization. While this formalism allows us to find relatively compact expressions for amplitudes with massive fields, the reference momentum q does not drop out from the final expressions. This is to be expected: for a massive particle helicity is not a Lorentz-invariant quantity, so the decomposition (457) depends on the frame.

A more elegant approach involves a doublet of spinors $\lambda_{\alpha}^{a}, \tilde{\lambda}_{\dot{\beta}}^{a}$ which transform covariantly under the $SO(3) \cong SU(2)$ little group appropriate for describing massive particles in four dimensions. Massive momenta are then written as [492]:

$$k_{\alpha\dot{\beta}} = k_{\mu}\sigma^{\mu}_{\alpha\dot{\beta}} = \epsilon_{ab}|k^a\rangle_{\alpha}[k^b|_{\dot{\beta}} = \epsilon_{ab}\lambda_{\alpha}^{a}\tilde{\lambda}_{\dot{\beta}}^{b}\,, \tag{458}$$

where a, b are little group $SU(2)$ indices and $\alpha, \dot{\beta}$ are four-dimensional Weyl spinor indices. Massive vector polarizations are written n terms of the spinors $\lambda_{\alpha}^{a}, \tilde{\lambda}_{\dot{\beta}}^{a}$ as

$$\varepsilon_{\mu}^{ab}(k) = \frac{\langle k^{(a}|\sigma_{\mu}|k^{b)}]}{\sqrt{2}m}\,, \tag{459}$$

where the little group indices are symmetrized. This formalism can be straightforwardly extended to construct polarization tensors for higher-spin massive fields [492] and presents close analogies with massless spinor-helicity in six dimensions [493].

B.2. On-shell superamplitudes

For supersymmetric amplitudes, on-shell superspace provides a convenient organization of amplitudes according to their physical helicity states which also tracks the relationships between the different component amplitudes. The power of such an on-shell superspace follows from the fact that, for generic momentum configurations, scattering amplitudes are insensitive to the nonlinear parts of (super)symmetry transformations (see section 4 for more details.). This greatly simplifies the evaluation of state sums in both the on-shell recursion [162] and generalized unitarity [163, 164, 166] by allowing that all physical states be treated simultaneously.

To illustrate the ideas we use $\mathcal{N} = 4$ SYM theory [489] as an example. Similar constructions exist in cases with less than maximal supersymmetry [494] as well as supergravity theories [495]. These superspaces are obtained by extending the usual momentum space

Table 17. The states of $\mathcal{N}=8$ supergravity organized via the double copy. The $SU(8)$ representations are decomposed in representations of the $SU(4)\times SU(4)$ subgroup which is manifest in the construction. For gauge theory $(g^+,\lambda^+,\phi,\lambda^-,g^-)$ carry helicity $(+1,\frac{1}{2},0,-\frac{1}{2},-1)$, while the helicity of the supergravity states are the sum of gauge-theory helicities. The $\pm$ decorating the entries represent the sign of the helicity of the corresponding state. The double-copy states can be reorganized into the standard $\mathcal{N}=8$ multiplet containing 256 physical states, cf equation (461).

	${g_R}^+$	${f_R}^+_{\tilde{I}}$	$\phi_{R\tilde{I}\tilde{J}}$	${f_R}^-_{\tilde{I}\tilde{J}\tilde{K}}$	${g_R}^-_{\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
${g_L}^+$	h^+	$\psi^+_{\tilde{I}}$	$A^+_{\tilde{I}\tilde{J}}$	$\chi^+_{\tilde{I}\tilde{J}\tilde{K}}$	$\phi_{\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
${f_L}^+_I$	ψ^+_I	$A^+_{I\tilde{I}}$	$\chi^+_{I\tilde{I}\tilde{J}}$	$\phi_{I\tilde{I}\tilde{J}\tilde{K}}$	$\chi^-_{I\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
ϕ_{LIJ}	A^+_{IJ}	$\chi^+_{IJ\tilde{I}}$	$\phi_{IJ\tilde{I}\tilde{J}}$	$\chi^-_{IJ\tilde{I}\tilde{J}\tilde{K}}$	$A^-_{IJ\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
${f_L}^-_{IJKL}$	χ^+_{IJKL}	$\phi_{IJKL\tilde{I}}$	$\chi^-_{IJKL\tilde{I}\tilde{J}}$	$A^-_{IJKL\tilde{I}\tilde{J}\tilde{K}}$	$\psi^-_{IJKL\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$
${g_L}^-_{IJKL}$	ϕ_{IJKL}	$\chi^-_{IJKL\tilde{I}}$	$A^-_{IJKL\tilde{I}\tilde{J}}$	$\psi^-_{IJKL\tilde{I}\tilde{J}\tilde{K}}$	$h^-_{IJKL\tilde{I}\tilde{J}\tilde{K}\tilde{L}}$

(parametrized in terms of spinor variables) with unconstrained Grassmann variables, η^I with $I=1,\ldots,\mathcal{N}$, which transform in the fundamental representation of the R-symmetry group and carry unit little group weight. The bosonic spinor variables carry kinematic information, while the Grassmann variables carry information on the helicity and R-symmetry representation of the external states. On-shell superfields—i.e. fields defined on these superspaces—have a finite expansion in the fermionic variables, with each coefficient being a component fields of definite helicity and R-symmetry representation. $\mathcal{N}=4$ SYM has a simple structure because all states can be assembled into a single CPT-self-conjugate on-shell superfield:

$$\Phi(\eta)=g^++\eta^If^+_I+\frac{1}{2}\eta^I\eta^J\phi_{IJ}+\frac{1}{3!}\eta^I\eta^J\eta^Kf^-_{IJK}+\frac{1}{4!}\eta^I\eta^J\eta^K\eta^Lg^-_{IJKL}, \tag{460}$$

where g^+ is the positive helicity gluon, f^+_I four positive helicity Majorana fermions, ϕ_{IJ} six real scalars, $f^{I-}\equiv\frac{1}{3!}\epsilon^{IJKL}f^-_{JKL}$ four negative helicity Majorana fermions and g^- is the negative helicity gluons, for a total of $8+8$ physical states (not including color degrees of freedom). The case of $\mathcal{N}=8$ supergravity is similar, with a four-dimensional CPT-self-conjugate on-shell superfield containing fields up to helicity ±2:

$$\begin{aligned}\Phi(\eta)=&\,h^++\eta^I\psi^+_I+\frac{1}{2}\eta^I\eta^JA_{IJ}+\frac{1}{3!}\eta^I\eta^J\eta^K\chi^+_{IJK}+\frac{1}{4!}\eta^I\eta^J\eta^K\eta^L\phi_{IJKL}\\&+\frac{1}{5!}\eta^I\eta^J\eta^K\eta^L\eta^M\chi^-_{IJKLM}+\frac{1}{6!}\eta^I\eta^J\eta^K\eta^L\eta^M\eta^NA^-_{IJKLMN}\\&+\frac{1}{7!}\eta^I\eta^J\eta^K\eta^L\eta^M\eta^N\eta^O\psi^-_{IJKLMNO}+\frac{1}{8!}\eta^I\eta^J\eta^K\eta^L\eta^M\eta^N\eta^O\eta^Ph^-_{IJKLMNOP}.\end{aligned} \tag{461}$$

Each supergravity state is a double copy of the gauge-theory states[66]. The 256 physical states of $\mathcal{N}=8$ supergravity correspond to the 16×16 direct product of states of two $\mathcal{N}=4$ SYM theories, as shown in table 17.

Supersymmetry transformations act on on-shell superfields (i.e. single-particle supersymmetry transformations) as

$$Q^{\dot\alpha}_I=\tilde\lambda^{\dot\alpha}\frac{\partial}{\partial\eta^I}\quad,\quad Q^I_{\dot\alpha}=\eta^I\frac{\partial}{\partial\tilde\lambda^{\dot\alpha}}\quad,\quad Q^{\alpha I}=\lambda^\alpha\eta^I\quad,\quad Q_{\alpha I}=\frac{\partial^2}{\partial\lambda^\alpha\partial\eta^I}; \tag{462}$$

[66] As we discussed in section 5, supergravity states can more generally be understood as being in one-to-one correspondence with gauge-invariant bilinears of gauge-theory states.

the (linearized) supersymmetry transformations of component fields are extracted by acting on superfields with these generators and reading off the coefficient of the desired combination of Grassmann variables η. Multi-particle supersymmetry generators are obtained by summing the single-particle ones over all the particles; for example, $Q_I^{\dot\alpha}$ acts on a product of n distinct fields as

$$Q_I^{\dot\alpha} = \sum_{i=1}^{n} Q_{iI}^{\dot\alpha} = \sum_{i=1}^{n} \tilde\lambda_i^{\dot\alpha} \frac{\partial}{\partial \eta_i^I}\,. \tag{463}$$

For supersymmetry algebras with less-than-maximal supersymmetry not all multiplets are CPT-self-conjugate; in such cases the fields of the theory form (perhaps several) CPT-conjugate pairs.

Scattering amplitudes in supersymmetric field theories can be combined into superamplitudes, defined as polynomials in Grassmann variables such that the coefficient of each monomial is a component amplitude whose helicity configuration is dictated by the η factors that multiply it and the structure of the superfields of the theory. The details depend on the amount of supersymmetry; as above, we illustrate these ideas for $\mathcal{N}=4$ SYM theory. See [91, 494] for less supersymmetric cases. On-shell supersymmetry Ward identities, relating component amplitudes with different external field configurations, can be derived by demanding that superamplitudes are annihilated by the multi-particle supersymmetry generators. A detailed descriptions may be found in [91, 495–497]. The unconstrained nature of the Grassmann variables makes it straightforward to translate summations of on-shell states needed in unitarity cuts or on-shell recursion into Grassmann integrations, which take care of the state bookkeeping. See [498, 499] for details. This procedure ensures that all generalized cuts are manifestly supersymmetric.

The minimum number of Grassmann variables in superamplitudes enforces the conservation of the polynomial supercharge $Q^{\alpha\, I}$ in equation (462), sometimes referred to as the 'supermomentum':

$$\delta^{(8)}(Q) \equiv \delta^{(8)}\left(\sum_{j=1}^{n} \lambda_j^{\alpha} \eta_j^I\right) = \prod_{I=1}^{4} \sum_{i<j}^{n} \langle ij\rangle \eta_i^I \eta_j^I\,. \tag{464}$$

The superamplitude with this minimal number of Grassmann variables are referred to as maximally-helicity-violating (MHV) superamplitude and the corresponding component amplitudes are referred to in a similar manner. The name reflects the fact that these amplitudes, with all incoming particles, exhibit the maximum imbalance between positive and negative helicities[67]. The n-point maximally-helicity-violating (MHV) tree amplitudes of $\mathcal{N}=4$ SYM theory are [489]

$$\mathcal{A}_n^{\text{MHV}}(1,2,\cdots,n) = \frac{i}{\prod_{j=1}^{n}\langle j\,(j+1)\rangle}\,\delta^{(8)}\left(\sum_{j=1}^{n} \lambda_j^{\alpha}\eta_j^I\right), \tag{465}$$

where leg $n+1$ is to be identified with leg 1. The coefficient of the supermomentum-conserving δ-function is cyclically symmetric and can also be identified as the ratio

$$A_n^{\text{tree}}\left(1^-,2^-,3^+,\cdots,n^+\right)/\langle 12\rangle^4\,. \tag{466}$$

[67] At loop level, in non-supersymmetric theories the imbalance can be even larger, as the all-plus and single-minus gluon amplitudes and their conjugates are nonvanishing [264].

General n-point $\mathcal{N}=4$ SYM superamplitudes can be written as

$$\mathcal{A}_n = \mathcal{A}_n^{\mathrm{MHV}}\left(\mathcal{P}_n^{(0)} + \mathcal{P}_n^{(1)} + \cdots + \mathcal{P}_n^{(n-4)}\right), \tag{467}$$

where $\mathcal{P}_n^{(k)}$ is a polynomial of degree $4k$ in Grassmann variables and $\mathcal{P}_n^{(0)} = 1$. Through $k_{\mathrm{mid}} = [(n-4)/2]$, the component amplitudes captured by these terms have a smaller difference between the number of external states with positive and negative helicity and are referred to as (next-to)^k-MHV amplitudes (or, generically, non-MHV amplitudes). The component amplitudes with $k > k_{\mathrm{mid}}$ can be obtained from those with $k < k_{\mathrm{mid}}$ by conjugation[68]. We note that, while all component amplitudes of an MHV superamplitude are related to each other by supersymmetry transformations, several non-MHV component amplitudes are needed to generate entire superamplitude. The polynomials $\mathcal{P}_n^{(k\geqslant 1)}$ are expressed in terms of the R-invariants [501, 502] which manifest the dual superconformal invariance of tree-level amplitudes of $\mathcal{N}=4$ SYM theory. This symmetry led to the derivation [502] of an explicit form of all tree-level amplitudes of this theory and, consequently, also of pure Yang-Mills theory, as superpartners do not contribute at tree level.

The MHV superamplitudes of $\mathcal{N}=8$ supergravity have a form similar to the one of $\mathcal{N}=4$ SYM theory and are given by

$$\mathcal{M}_n^{\mathrm{MHV}} = \frac{M_n\left(1^-,2^-,3^+,\cdots,n^+\right)}{\langle 12\rangle^8}\,\delta^{(16)}\left(\sum_{j=1}^{n}\lambda_j^{\alpha}\eta_j^I\right), \tag{468}$$

where $M_n(1^-,2^-,3^+,\cdots,n^+)$ is a tree-level MHV pure-graviton amplitude. The simplicity of this result follows from the fact that, for MHV superamplitudes, the entire superspace content is contained in an overall supermomentum-conserving δ-function. The supergravity one is the double-copy of the gauge theory one and it evaluates to

$$\delta^{(16)}\left(\sum_{j=1}^{n}\lambda_j^{\alpha}\eta_j^a\right) = \prod_{I=1}^{8}\sum_{i<j}^{n}\langle ij\rangle\eta_i^I\eta_j^I\,. \tag{469}$$

For more general amplitudes the results are more complicated, but follow directly by applying the KLT or BCJ double copy to gauge-theory superamplitudes to obtain superamplitudes in supergravity; the manifestly-supersymmetric KLT relations were discussed in [503]. Since each gauge-theory amplitude exhibits a factor of the supermomentum-conserving δ-function, which is symmetric under permutation of the external lines, the supergravity amplitude inherits the complete supersymmetry of both gauge-theory factors. Thus, the supergravity R-symmetry group is $SU(N_L+N_R)$—see section 4 for more details. Moreover, as we discuss in section 4, double-copy supergravity exhibits an emergent $U(1)$ symmetry, which is part of its U-duality group. Among its implications is the vanishing of the double-copy of (super)amplitudes in different N^{k}MHV sectors.

As discussed in [499], we can obtain superamplitudes in theories with fewer supersymmetries by appropriately grouping R-symmetry indices. This can be realized by truncating the supermultiplets (460) and (461) such that a certain subset of the η-variables always appear together; the corresponding component amplitudes are obtained from those of the maximally-supersymmetric theory by restricting them to the combinations of Grassmann variables that

[68] Conjugation of superamplitudes exchanges η with their conjugates and thus also changes the type of on-shell superspace. The transformation to the original superspace is given by the fermionic Fourier transform of all conjugate η variables.. For a discussion of various superspaces see [500].

are allowed to appear for each external state. The simplest example is that the tree amplitudes of pure non-supersymmetric Yang-Mills theory are just the pure-gluon amplitudes of $\mathcal{N}=4$ SYM theory, only for these amplitudes the η variables appear in the combination $\eta^1\eta^2\eta^3\eta^4$; all other states contain a subset of their corresponding Grassmann variables and thus decouple from these amplitudes. In fact, using appropriate projections, one can even obtain QCD tree-level amplitudes with quarks from $\mathcal{N}=4$ tree amplitudes [504], leading to compact forms of QCD amplitudes[69].

As an example, the MHV tree amplitudes for external gauge supermultiplets in $\mathcal{N}$-extended SYM theory are given by [499]

$$\mathcal{A}_n^{\mathrm{MHV}}(1,2,\ldots,n) = \frac{\prod_{I=1}^{\mathcal{N}} \delta^{(2)}\left(Q^I\right)}{\prod_{j=1}^{n}\langle j\ \ (j+1)\rangle} \left(\sum_{i<j}^{n} \langle ij\rangle^{4-\mathcal{N}} \prod_{I=\mathcal{N}+1}^{4} \eta_i^I \eta_j^I\right), \tag{470}$$

with $\mathcal{N}$ counting the number of supersymmetries, $Q^I = \sum_{i=1}^{n} \lambda_i \eta_i^I$, and $n \geqslant 3$.

Using supersymmetric versions of MHV amplitudes [505] and on-shell recursion [162], general tree superamplitudes in gauge and gravity theories have been systematically constructed (see e.g. [222, 502, 504, 506]). As we briefly review in appendix C, they can be used as input building blocks to construct the integrands of loop superamplitudes. The cases of $\mathcal{N}<4$ superamplitudes, have been analyzed in [494] in some detail.

Exercise B.2. Using the supersymmetry generators (462) and the on-shell superfield (460), construct the linearized supersymmetry transformations of the component fields of $\mathcal{N}=4$ SYM theory. Derive the corresponding relations between the component MHV amplitudes.

Appendix C. Generalized unitarity

In this appendix we give a brief summary of the modern generalized unitary method [163–166, 194, 217] used in multiloop calculations, focusing on their applications in double-copy constructions. This provides some of the necessary background for our review of the generalized double-copy construction in section 7. We will present several examples illustrating the basic ideas and refer the reader to other reviews for further details [90, 196, 507, 508].

The generalized-unitary method systematically builds complete loop-level integrands using as input only on-shell tree-level amplitudes. A central feature is that simplifications and features of the latter are directly imported into the former. In particular, with this method we can use tree-level double-copy relations to construct gravity loop integrands. We also briefly review a variant of generalized unitarity, known as the maximal-cut method [217], which meshes well with the generalized double copy [416] discussed in section 7. A reorganization of the generalized-unitarity method that has various advantages is found in [509].

Traditionally, unitarity of the scattering matrix is implemented at the integrated level via dispersion relations [510]. For our purposes, however, it is much more useful to use it at the integrand level. We introduce the concept of a generalized cut that reduces an integrand to a sum of products of tree amplitudes $A_{(j)}^{\mathrm{tree}}$,

$$\sum_{\mathrm{states}} A_{(1)}^{\mathrm{tree}} A_{(2)}^{\mathrm{tree}} A_{(3)}^{\mathrm{tree}} \cdots A_{(m)}^{\mathrm{tree}}. \tag{471}$$

[69] The necessary change in the color factor reflecting the change in gauge-group representation can easily be accounted for in tree-level amplitudes though a multiplicative factor.

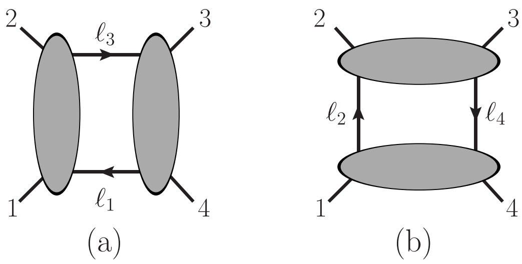


Figure 34. The (a) s and (b) t channel two-particle cuts of a one-loop four-point amplitude. The exposed lines are all on-shell and the blobs represent tree amplitudes.

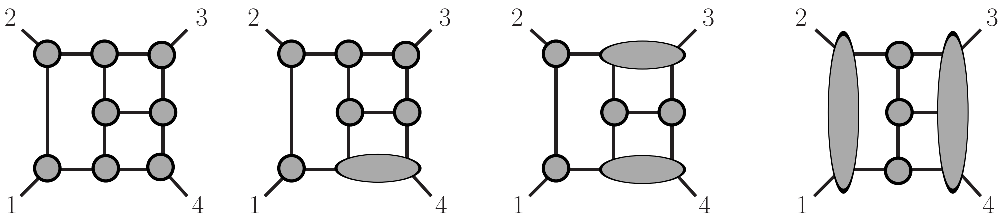


Figure 35. Examples of generalized cuts for a three-loop four-point amplitude. The exposed lines are all on-shell and the blobs represent tree amplitudes.

Each cut propagator is replaced with a delta function enforcing on-shell constraint for the corresponding momentum. The sum runs over all intermediate physical states that can contribute given the external states of the amplitude being constructed. Some generalized cuts of the one-loop four-point amplitude are shown in figure 34 and of the three-loop four-point amplitude in figure 35. In these Figures the exposed lines are all on-shell delta functions and the blobs represent on-shell tree amplitudes.

Loop integrands are determined by spanning set of generalized cuts, i.e. a set of cuts which receive contributions from all the terms that could possibly be generated by the Feynman graphs of the theory. Loop integrands are constructed by finding a single function whose cuts match all the products of tree amplitudes, summed over states corresponding to such a spanning set. Regardless of which set of cuts one uses to construct an integrand, one must always verify it on a minimal spanning set (i.e. a spanning set that contains the minimal number of cuts). To illustrate these ideas in practice we turn to a few simple unitarity cuts.

C.1. One-loop example of unitarity cuts

To illustrate the generalized unitarity method and how it meshes with double-copy ideas consider the two-particle cuts of a one-loop four-point color-ordered gauge-theory amplitude. In these amplitudes the color factors are stripped away and the external legs follow a cyclic ordering [88, 89]. The two-particle cuts of a one-loop four-point amplitude are obtained by putting two intermediate lines on shell, as illustrated in figure 34. For example, the s-channel cut in figure 34(a) is given by

$$C_s^{\text{gauge}} = \sum_{\text{states}} A_4^{\text{tree}}\left(-\ell_1, 1, 2, \ell_3\right) A_4^{\text{tree}}\left(-\ell_3, 3, 4, \ell_1\right), \tag{472}$$

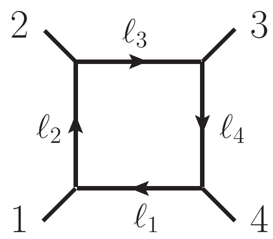


Figure 36. The one-loop box integral and loop momentum labels used in the cut construction.

where the sum runs over all physical states in the theory. The cuts are evaluated using momenta that place all cut-line momenta on shell, i.e. $\ell_i^2 = 0$ if the theory is massless.

A particularly simple example is the color-ordered one-loop four-point amplitude in $\mathcal{N} = 4$ SYM theory, which is useful for illustrating how gauge-theory unitarity cuts can be converted to gravity ones. For this theory, after summing over all physical states that cross the two-particle cut, the result takes a remarkably compact form [399],

$$C_s^{\text{gauge}} == -istA_4^{\text{tree}}(1,2,3,4)\frac{1}{(\ell_1 - p_1)^2}\frac{1}{(\ell_3 - p_3)^2}\,. \tag{473}$$

All momenta are on shell. This expression is valid for any external states of the theory; the cut depends on them only through the overall factor $A_4^{\text{tree}}(1,2,3,4)$. The t-channel cut in figure 34(b) is obtained by relabeling equation (473).

The most straightforward way to verify these equations is by using four-dimensional helicity states and on-shell superspace, but they hold in $D \leqslant 10$ dimensions as well (where maximal supersymmetric Yang-Mills theory is defined). Details may be found in [194].

Putting back the cut propagators and loop integration we find a function with the correct s-channel cut,

$$i\,stA_4^{\text{tree}}(1,2,3,4)\,I_4(s,t)\Big|_{s\text{-}cut}\,, \tag{474}$$

where $I_4(s,t)$ is the scalar box integral shown in figure 36 and given in equations (286) and (287). By the cut operation in the s channel in equation (474) we mean to remove the integration and to replace the two propagators $1/\ell^2$ and $1/(\ell - p_1 - p_2)^2$ with on-shell conditions, recovering equation (473) after identifying $\ell = \ell_1$ and applying momentum conservation. Similar considerations show that the t channel cut can be written as

$$i\,stA_4^{\text{tree}}(1,2,3,4)\,I_4(s,t)\Big|_{t\text{-}cut}\,. \tag{475}$$

Once unitarity cuts are written in this way, as the cuts of a single function, it is easy to see that the one-loop four-point amplitude, with no cut conditions, is obtained simply by removing the cut conditions,

$$A_{\mathcal{N}=4}^{1-\text{loop}}\left(1^-,2^-,3^+,4^+\right) = i\,stA_4^{\text{tree}}(1,2,3,4)\,I_4(s,t)\,, \tag{476}$$

matching the result in equation (282).

These basic ideas generalize to any massless gauge theory and underpin many theoretical studies, including those for collider physics (see e.g. [165, 511, 512]).

C.2. Converting gauge-theory unitarity cuts to gravity ones

As discussed in section 2, the BCJ forms of loop-level gauge-theory integrands can be directly converted to gravity ones. Because the unitarity-based construction uses tree amplitudes as input, we can also straightforwardly apply the KLT relations (79) to convert gravity unitarity cuts to sums of products of gauge-theory cuts. The BCJ form of the gauge-theory tree amplitudes (50) may also be used to apply the double copy to convert gauge-theory cuts to the corresponding gravity ones. The KLT form is especially useful when working with compact helicity amplitudes, while the BCJ form is helpful in D dimensions (i.e. when using dimensional regularization) with formal polarization vectors.

Consider first the two-particle cut of a one-loop four-point amplitude show in figure 34(a) in a gravity theory. Using the KLT form of the double copy, it is given by

$$\begin{aligned} C_{\mathrm{GR}} &= \sum_{\substack{\text{gauge}\\\text{states}}} M_4^{\text{tree}}\left(-\ell_1,1,2,\ell_3\right) \times M_4^{\text{tree}}\left(-\ell_3,3,4,\ell_1\right) \\ &= -s^2\left(\sum_{\substack{\text{gauge}\\\text{states}}} A_4^{\text{tree}}\left(-\ell_1,1,2,\ell_3\right) \times A_4^{\text{tree}}\left(-\ell_3,3,4,\ell_1\right)\right) \\ &\quad\times\left(\sum_{\substack{\text{gauge}\\\text{states}}} A_4^{\text{tree}}\left(\ell_3,1,2,-\ell_1\right) \times A_4^{\text{tree}}\left(\ell_1,3,4,-\ell_3\right)\right), \end{aligned} \tag{477}$$

where we applied the KLT relation (31) to rewrite each gravity tree amplitude in terms of a product of two gauge-theory amplitudes. In this expression we have assumed that the gravity theory of interest arises as a simple double copy, as it does for $\mathcal{N}=8$ supergravity. This allows us to decompose each state in the gravity theory into a 'left' and a 'right' gauge-theory state. For the case of $\mathcal{N}=8$ supergravity, summing over the states in the $\mathcal{N}=8$ multiplet is equivalent to summing independently over the left and right $\mathcal{N}=4$ SYM gauge-theory multiplets. For theories which are not simple double copies, such as pure gravity, one must remove unwanted states (i.e. dilaton and antisymmetric tensor) by inserting explicit physical-state projectors into the cuts. These projectors have been used effectively at two-loops to study ultraviolet properties of various theories, including pure gravity [338] as well as for computing the third post-Minkowskian correction to the conservative two-body Hamiltonian [80, 82]. In some cases, it is sufficient to evaluate the generalized unitarity cuts in four dimensions, where we can simplify the input gauge-theory amplitudes enormously by using helicity states. In this case, a simple way to control which particles circulate in the loops, is by correlating the state sum of the two gauge theories, For example, if we want only gravitons to cross the cuts, then for each term we should have identical helicity for the corresponding gluons in the state sum [80, 82]. A similar procedure works well for supergravity theories which are obtained as orbifolds of e.g. $\mathcal{N}=8$ supergravity where the orbifold action cannot be decomposed into independent actions on the left and right $\mathcal{N}=4$ SYM gauge theories [30].

The one-loop four-point amplitude in $\mathcal{N}=8$ supergravity is an instructive illustration of how we can recycle gauge-theory unitarity cuts into gravity ones. For this case, equation (477) immediately collapses because, up to relabeling, each gauge-theory state sum is the $\mathcal{N}=4$ SYM state sum in equation (473). Inserting the simplified $\mathcal{N}=4$ SYM cut (473) into (477) results in an equivalent relation for $\mathcal{N}=8$ supergravity,

$$
\begin{aligned}
C_{\mathrm{GR}} &= s^2 (st)^2 \left[A_4^{\mathrm{tree}}(1,2,3,4)\right]^2 \frac{1}{(\ell_1-p_1)^2 (\ell_3-p_3)^2 (\ell_3+p_1)^2 (\ell_1+p_3)^2} \\
&= s^2 \left[stA_4^{\mathrm{tree}}(1,2,3,4)\right]^2 \frac{1}{(\ell_1-p_1)^2 (\ell_3-p_3)^2 (\ell_1-p_2)^2 (\ell_3-p_4)^2} \\
&= i\, stuM_4^{\mathrm{tree}}(1,2,3,4) \left[\frac{1}{(\ell_1-p_1)^2} + \frac{1}{(\ell_1-p_2)^2}\right] \left[\frac{1}{(\ell_3-p_3)^2} + \frac{1}{(\ell_3-p_4)^2}\right]. \qquad (478)
\end{aligned}
$$

To obtain this we partial fractioned the product of propagators and used the KLT relations (30) and the BCJ amplitude relations (28). As for gauge-theory cuts, the ℓ_1 and ℓ_3 are on shell. The t- and u-channel formulae are obtained by relabeling the external legs in equation (478).

Following the same strategy as for the reconstruction of the one-loop four-point $\mathcal{N}=4$ SYM amplitude, it is then straightforward to obtain the $\mathcal{N}=8$ one-loop four-point amplitude,

$$
\mathcal{M}_4^{1-\mathrm{loop}}(1,2,3,4) = -i\left(\frac{\kappa}{2}\right)^4 stuM_4^{\mathrm{tree}}(1,2,3,4)\left(I_4(s,t)+I_4(s,u)+I_4(t,u)\right). \qquad (479)
$$

Here $I_4(s,t)$ is the box integrals defined in equation (286), while $I_4(s,u)$ and $I_4(t,u)$ are obtained by appropriate relabeling of external legs. This agrees with the form obtained using the BCJ double copy in section 6 and agrees with the result first obtained by Brink, Green, Schwarz [392] in the field-theory limit of superstring theory.

One can also use the BCJ double copy (45) for the component tree amplitudes of the gravity unitarity cuts. This is especially efficient when working in D dimensions (e.g. when using dimensional regularization), because compact helicity-based expressions for tree-level amplitudes are no longer available and, consequently, the result of the KLT relations will be cumbersome to use. In D dimensions, the BCJ form is a more natural form because it preserves the diagram structure when converting from gauge theory to gravity. For example, the two-particle cut (a) in figure 34, can be evaluated using the form of the double copy in equation (18),

$$
C_{\mathrm{GR}}^{(\mathrm{a})} = \sum_{\mathrm{pols.}} M_4^{\mathrm{tree}}(-\ell_1,1,2,\ell_3)\times M_4^{\mathrm{tree}}(-\ell_3,3,4,\ell_1)\,, \qquad (480)
$$

where the graviton tree amplitudes in the cut are obtained from the double copy form in equation (18) by simple relabelings. The sum over polarizations gives the physical-state projector. For gravitons in D dimensions the projector is

$$
P^{\mu\nu\rho\sigma}(p,q) = \sum_{\mathrm{pols.}} \varepsilon^{\mu\nu}(-p)\,\varepsilon^{\rho\sigma}(p) = \frac{1}{2}\left(P^{\mu\rho}P^{\nu\sigma}+P^{\mu\sigma}P^{\nu\rho}\right) - \frac{1}{D_s-2}P^{\mu\nu}P^{\rho\sigma}, \qquad (481)
$$

where $P^{\mu\nu}$ is the gluon physical-state projector

$$
P^{\mu\nu}(p,q) = \sum_{\mathrm{pols.}} \varepsilon^{\mu}(-p)\,\varepsilon^{\nu}(p) = \eta^{\mu\nu} - \frac{q^\mu p^\nu + p^\mu q^\nu}{q\cdot p}, \qquad (482)
$$

with momentum p and a null reference momentum q. In some cases, terms that vanish on-shell can be added to tree amplitudes so that the dependence on the reference momentum disappears because of the on-shell Ward identity for the gauge symmetry [80, 82, 513].

C.3. Method of maximal cuts

A refinement of the unitarity method [163, 164], which is especially helpful at higher loop orders, is the method of maximal cuts [217]. This method is not only a basic tool for checking and building double-copy gravity integrands, but it also plays a central role in the generalized-double-copy construction described in section 7. This construction was central to a computation determining the ultraviolet properties of $\mathcal{N}=8$ supergravity at five loops [38, 218, 416].

In the method of maximal cuts, the unitarity cuts are clustered in levels according to the number k of internal propagators allowed to remain off shell,

$$\mathcal{C}^{\mathrm{N}^k\mathrm{MC}} = \sum_{\text{states}} \mathcal{A}^{\text{tree}}_{m(1)} \cdots \mathcal{A}^{\text{tree}}_{m(p)}\,, \tag{483}$$

where $\mathcal{A}^{\text{tree}}_{m(i)}$ are tree-level $m(i)$-multiplicity amplitudes corresponding to the blobs, illustrated for various cuts of a three-loop four-point amplitude illustrated in figure 35. The level k is related to the multiplicity of the various factors by

$$k = \sum_{i=1}^{p} \left(m(i) - 3\right)\ . \tag{484}$$

The cuts (483) can be applied to either gauge or gravity amplitudes. As illustrated in the first diagram in figure 35, at the maximal cut (MC) level the maximum number of propagators are replaced by on-shell conditions and all tree amplitudes appearing in equation (483) are three-point amplitudes. At the next-to-maximal-cut (NMC) level, illustrated in the second cut of figure 35, a single propagator is placed off shell and so forth.

With this organization of generalized cuts, the integrands for L-loop amplitudes are obtained by first establishing an integrand whose maximal cuts are correct, then adding to it terms so that NMCs are all correct and systematically proceeding through the next^k maximal cuts (N^{k}MCs), until no further contributions are found. Where this process completes is dictated by the power counting of the theory and by choices made at each level. For example, if minimal power counting is assigned to each contribution, for $\mathcal{N}=4$ SYM four-point amplitudes, cuts through NMCs, N^{2}MCs and N^{3}MCs are sufficient at three [2], four [6] and five loops [514], respectively.

Most calculations (see e.g. [6, 9, 15, 17, 292, 293, 407, 408]) find it convenient to organize the integrands in terms of diagrams with purely cubic vertices, such as the three-loop ones illustrated in figure 27. Representations with only cubic diagrams have certain advantages: they are useful for establishing minimal power counting in each diagram, and the number of diagrams used to describe the result proliferate minimally with the loop order and multiplicity. A disadvantage is that ansätze are required for imposing various properties on each diagram, including the desired power counting, symmetry, and the multiple unitarity cuts to which a given diagram contributes. As the loop order increases, it becomes cumbersome to solve the requisite system of equations that imposes these constraints. We can avoid this in the generalized double-copy construction if we instead assign any new information obtained in a N^{k}MC to a contact diagram, as discussed in section 7 and illustrated in figure 37. This is necessarily local because the nonlocal contributions are accounted for at previous levels.

C.3.1. Sewing superamplitudes. We now briefly comment on the use of the on-shell superspace described in appendix B for the evaluation of the sums over the states crossing a unitarity cut. It turns out [498, 499, 515] that it can be conveniently expressed as an integration over the

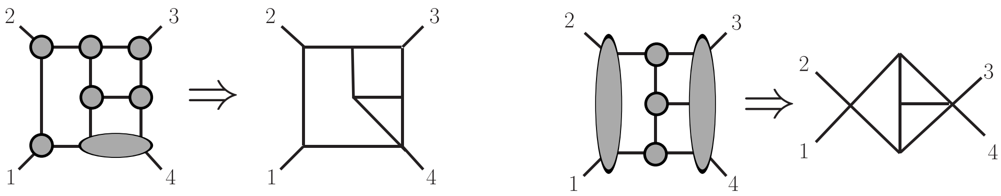


Figure 37. New contribution found via the method of maximal cuts can be assigned directly to contact terms. In these diagrams all the exposed lines are on-shell, and propagators within a blob remain off shell.

Grassmann parameters of the cut legs. The generalized supercut in $\mathcal{N}$-extended SYM theory is then given by[70]

$$\mathcal{C} = \int \left[\prod_{i=1}^{k} d^{\mathcal{N}}\eta_i\right] \mathcal{A}^{\text{tree}}_{(1)}\mathcal{A}^{\text{tree}}_{(2)}\mathcal{A}^{\text{tree}}_{(3)}\cdots\mathcal{A}^{\text{tree}}_{(m)}\,, \tag{485}$$

where $\mathcal{A}^{\text{tree}}_{(j)}$ are the tree-level superamplitudes connected by k on-shell cut legs. For each cut leg, the integral over η selects all possible states on that leg and sums up their contribution to the cut. These supercuts are functions on the on-shell superspace. For four and higher points the tree amplitudes $\mathcal{A}^{\text{tree}}_{(j)}$ are always proportional to a supermomentum delta function. Using the simple identity $\delta(A)\delta(B) = \delta(A+B)\delta(B)$, this implies that all such cuts are proportional to an overall supermomentum δ-function [498]. It turns out that such supercuts are sufficient for determining massless superamplitudes. This then implies that the four-dimensional cuts of any loop amplitude with four or more external legs must be proportional to an overall supermomentum conservation δ-function. Barring supersymmetry anomalies, this will also be the case for the corresponding superamplitudes.

Fermionic integration provides one of the several different methods for the evaluation of supersums in unitarity cuts [222, 498, 499, 515]. There are two main approaches for organizing the integration over the η parameters. In the first way, the fermionic δ-functions can be used to localize the integration, so that the evaluation of the supersum amounts to solving a system of linear equations [498, 499]. In a second complementary approach, 'index diagrams' are used to track the various contributions to the sum over states [499]. This approach leads to a simple algorithm for reading off the contribution of the entire supermultiplet from the purely gluonic ones and for reducing the number of supersymmetries. It was used in the construction of the complete four-loop four-point amplitude of $\mathcal{N}=4$ SYM theory [3].

The overall supermomentum-conserving δ-function has consequences on the ultraviolet properties of the theory akin to those of off-shell superspaces. In particular, in a theory with $\mathcal{N}$-extended supersymmetry, it implies that at least $\mathcal{N}$ powers of momenta in the numerators of each diagram are external momenta. In turn, this implies that the superficial degree of divergence of each diagram is improved by $\mathcal{N}$ compared to that of the non-supersymmetric theory. For $\mathcal{N}=4$ SYM theory, this simple power counting implies the well known [516–518] ultraviolet finiteness of all of its superamplitudes [499]. For $\mathcal{N}=8$ supergravity the BCJ double copy appears to imply that individual diagrams generally have a poor power count, because the kinematic numerators are products of corresponding gauge-theory ones. However, it is difficult to draw any conclusions based on this observation because of the existence of 'enhanced

[70] In this formulation CPT-conjugate multiplets are both interpreted as being embedded in an $\mathcal{N}=4$ multiplet. This is equivalent to the formulation of [494] up to Grassmann Fourier-transform with respect to four η variables.

cancellations' which are nontrivial (and not yet fully understood) cancellations between diagrams [292].

ORCID iDs

Zvi Bern https://orcid.org/0000-0001-9075-9501
John Joseph Carrasco https://orcid.org/0000-0002-4499-8488
Marco Chiodaroli https://orcid.org/0000-0002-9010-7658
Henrik Johansson https://orcid.org/0000-0001-5236-0954
Radu Roiban https://orcid.org/0000-0002-6689-5199

References

[1] Bern Z, Carrasco J J M and Johansson H 2008 *Phys. Rev.* D **78** 085011
[2] Bern Z, Carrasco J J M and Johansson H 2010 *Phys. Rev. Lett.* **105** 061602
[3] Bern Z, Carrasco J J M, Dixon L J, Johansson H and Roiban R 2010 *Phys. Rev.* D **82** 125040
[4] Carrasco J J M and Johansson H 2012 *Phys. Rev.* D **85** 025006
[5] Oxburgh S and White C D 2013 *J. High Energy Phys.* JHEP02(2013)127
[6] Bern Z, Carrasco J J M, Dixon L J, Johansson H and Roiban R 2012 *Phys. Rev.* D **85** 105014
[7] Du Y-J and Luo H 2013 *J. High Energy Phys.* JHEP01(2013)129
[8] Yuan E Y 2013 *J. High Energy Phys.* JHEP05(2013)070
[9] Boels R H, Kniehl B A, Tarasov O V and Yang G 2013 *J. High Energy Phys.* JHEP02(2013)063
[10] Boels R H, Isermann R S, Monteiro R and O'Connell D 2013 *J. High Energy Phys.* JHEP04(2013)107
[11] Bjerrum-Bohr N E J, Dennen T, Monteiro R and O'Connell D 2013 *J. High Energy Phys.* JHEP07(2013)092
[12] Bern Z, Davies S, Dennen T, Huang Y-T and Nohle J 2015 *Phys. Rev.* D **92** 045041
[13] Ochirov A and Tourkine P 2014 *J. High Energy Phys.* JHEP05(2014)136
[14] Mafra C R and Schlotterer O 2015 *J. High Energy Phys.* JHEP10(2015)124
[15] Mogull G and O'Connell D 2015 *J. High Energy Phys.* JHEP12(2015)135
[16] He S, Monteiro R and Schlotterer O 2016 *J. High Energy Phys.* JHEP01(2016)171
[17] Yang G 2016 *Phys. Rev. Lett.* **117** 271602
[18] Boels R H, Huber T and Yang G 2017 *Phys. Rev. Lett.* **119** 201601
[19] He S, Schlotterer O and Zhang Y 2018 *Nucl. Phys.* B **930** 328–83
[20] Johansson H, Kälin G and Mogull G 2017 *J. High Energy Phys.* JHEP09(2017)019
[21] Jurado J L, Rodrigo G and Torres Bobadilla W J 2017 *J. High Energy Phys.* JHEP12(2017)122
[22] Boels R H, Huber T and Yang G 2018 *J. High Energy Phys.* JHEP01(2018)153
[23] Faller J and Plefka J 2019 *Phys. Rev.* D **99** 046008
[24] Bjerrum-Bohr N E J, Damgaard P H, Sondergaard T and Vanhove P 2011 *J. High Energy Phys.* JHEP01(2011)001
[25] Mafra C R, Schlotterer O and Stieberger S 2011 *J. High Energy Phys.* JHEP07(2011)092
[26] Bjerrum-Bohr N E J, Bourjaily J L, Damgaard P H, Feng B and Feng B 2016 *J. High Energy Phys.* JHEP09(2016)094
[27] Du Y-J and Teng F 2017 *J. High Energy Phys.* JHEP04(2017)033
[28] de la Cruz L, Kniss A and Weinzierl S 2018 *J. High Energy Phys.* JHEP03(2018)064
[29] Bridges E and Mafra C R 2019 arXiv:1906.12252
[30] Chiodaroli M, Jin Q and Roiban R 2014 *J. High Energy Phys.* JHEP01(2014)152
[31] Bern Z, Boucher-Veronneau C and Johansson H 2011 *Phys. Rev.* D **84** 105035
[32] Boucher-Veronneau C and Dixon L J 2011 *J. High Energy Phys.* JHEP12(2011)046
[33] Bern Z, Davies S, Dennen T and Huang Y-T 2012 *Phys. Rev. Lett.* **108** 201301
[34] Bern Z, Davies S, Dennen T and Huang Y-T 2012 *Phys. Rev.* D **86** 105014
[35] Boels R H and Isermann R S 2013 *J. High Energy Phys.* JHEP06(2013)017
[36] Bern Z, Davies S and Dennen T 2013 *Phys. Rev.* D **88** 065007
[37] Bern Z, Davies S and Dennen T 2014 arXiv:1412.2441

[38] Bern Z, Carrasco J J, Chen W M, Edison A, Johansson H, Parra-Martinez J, Roiban R and Zeng M 2018 *Phys. Rev.* D **98** 086021
[39] Herrmann E and Trnka J 2016 *J. High Energy Phys.* JHEP11(2016)136
[40] Herrmann E and Trnka J 2019 *J. High Energy Phys.* JHEP02(2019)084
[41] Bern Z, Dennen T, Huang Y-T and Kiermaier M 2010 *Phys. Rev.* D **82** 065003
[42] Tolotti M and Weinzierl S 2013 *J. High Energy Phys.* JHEP07(2013)111
[43] Monteiro R and O'Connell D 2011 *J. High Energy Phys.* JHEP07(2011)007
[44] Bjerrum-Bohr N E J, Damgaard P H, Monteiro R and O'Connell D 2012 *J. High Energy Phys.* JHEP06(2012)061
[45] Monteiro R and O'Connell D 2014 *J. High Energy Phys.* JHEP03(2014)110
[46] Ho P-M 2016 *Phys. Rev.* D **93** 044062
[47] Fu C-H and Krasnov K 2017 *J. High Energy Phys.* JHEP01(2017)075
[48] Fu C-H, Vanhove P and Wang Y 2018 *J. High Energy Phys.* JHEP09(2018)141
[49] Chen G, Johansson H, Teng F and Wang T 2019 arXiv:1906.10683
[50] Saotome R and Akhoury R 2013 *J. High Energy Phys.* JHEP01(2013)123
[51] Monteiro R, O'Connell D and White C D 2014 *J. High Energy Phys.* JHEP12(2014)056
[52] Luna A, Monteiro R, O'Connell D and White C D 2015 *Phys. Lett.* B **750** 272–7
[53] Ridgway A K and Wise M B 2016 *Phys. Rev.* D **94** 044023
[54] Luna A, Monteiro R, Nicholson I O'Connell D and White C D 2016 *J. High Energy Phys.* JHEP06(2016)023
[55] White C D 2016 *Phys. Lett.* B **763** 365–9
[56] Cardoso G, Nagy S and Nampuri S 2017 *J. High Energy Phys.* JHEP04(2017)037
[57] Goldberger W D and Ridgway A K 2017 *Phys. Rev.* D **95** 125010
[58] Luna A, Monteiro R, Nicholson I, Ochirov A, O'Connell D, Westerberg N and White C D 2017 *J. High Energy Phys.* JHEP04(2017)069
[59] Goldberger W D, Prabhu S G and Thompson J O 2017 *Phys. Rev.* D **96** 065009
[60] Adamo T, Casali E, Mason L and Nekovar S 2018 *Class. Quantum Grav.* **35** 015004
[61] De Smet P-J and White C D 2017 *Phys. Lett.* B **775** 163–7
[62] Bahjat-Abbas N, Luna A and White C D 2017 *J. High Energy Phys.* JHEP12(2017)004
[63] Carrillo-González M, Penco R and Trodden M 2018 *J. High Energy Phys.* JHEP04(2018)028
[64] Goldberger W D, Li J and Prabhu S G 2018 *Phys. Rev.* D **97** 105018
[65] Li J and Prabhu S G 2018 *Phys. Rev.* D **97** 105019
[66] Ilderton A 2018 *Phys. Lett.* B **782** 22–27
[67] Lee K 2018 *J. High Energy Phys.* JHEP10(2018)027
[68] Plefka J, Steinhoff J and Wormsbecher W 2019 *Phys. Rev.* D **99** 024021
[69] Shen C-H 2018 *J. High Energy Phys.* JHEP11(2018)162
[70] Berman D S, Chacn E, Luna A and White C D 2019 *J. High Energy Phys.* JHEP01(2019)107
[71] Gurses M and Tekin B 2018 *Phys. Rev.* D **98** 126017
[72] Adamo T, Casali E, Mason L and Nekovar S 2019 *J. High Energy Phys.* JHEP02(2019)198
[73] Bahjat-Abbas N, Stark-Muchão R and White C D 2019 *Phys. Lett.* B **788** 274–9
[74] Luna A, Monteiro R, Nicholson I and O'Connell D 2019 *Class. Quantum Grav.* **36** 065003
[75] Farrow J A, Lipstein A E and McFadden P 2019 *J. High Energy Phys.* JHEP02(2019)130
[76] Carrillo González M, Melcher B, Ratliff K, Watson S and White C D 2019 *J. High Energy Phys.* JHEP07(2019)167
[77] Athira P V and Manu A 2019 arXiv:1907.10021
[78] Cheung C, Rothstein I Z and Solon M P 2018 *Phys. Rev. Lett.* **121** 251101
[79] Kosower D A, Maybee B and O'Connell D 2019 *J. High Energy Phys.* JHEP02(2019)137
[80] Bern Z, Cheung C, Roiban R, Shen C-H, Solon M P and Zeng M 2019 *Phys. Rev. Lett.* **122** 201603
[81] Antonelli A, Buonanno A, Steinhoff J, van de Meent M and Vines J 2019 arXiv:1901.07102
[82] Bern Z, Cheung C, Roiban R, Shen C H, Solon M P and Zeng M 2019 arXiv: 1908.01493
[83] Veneziano G 1968 *Nuovo Cim.* A **57** 190–7
[84] Virasoro M A 1969 *Phys. Rev.* **177** 2309–11
[85] Shapiro J A 1969 *Phys. Rev.* **179** 1345–53
[86] Kawai H, Lewellen D C and Tye S-H H 1986 *Nucl. Phys.* B **269** 1–23
[87] Green M B, Schwarz J H and Witten E 1988 Superstring theory *Introduction Cambridge Monographs on Mathematical Physics (Cambridge Monographs On Mathematical Physics)* vol 1 (University Press) p 469
[88] Mangano M L and Parke S J 1991 *Phys. Rep.* **200** 301–67

[89] Dixon L J 1996 Calculating scattering amplitudes efficiently QCD and beyond *Proc., Theoretical Advanced Study Institute in Elementary Particle Physics, TASI-95 (Boulder, USA, 4–30 June 1995)* pp 539–84
[90] Bern Z, Dixon L J and Kosower D A 1996 *Ann. Rev. Nucl. Part. Sci.* **46** 109–48
[91] Elvang H and Huang Y T 2013 arXiv:1308.1697
[92] Cheung C 2018 TASI lectures on scattering amplitudes *Proc., Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016) (Boulder, CO, USA, 6 June–1 July 2016)* pp 571–623
[93] Bern Z, Dixon L J, Perelstein M and Rozowsky J S 1999 *Nucl. Phys.* B **546** 423–79
[94] Bern Z and Grant A K 1999 *Phys. Lett.* B **457** 23–32
[95] Cheung C and Remmen G N 2017 *J. High Energy Phys.* JHEP01(2017)104
[96] DeWitt B S 1967 *Phys. Rev.* **162** 1239–56 DeWitt B S 1967 *Phys. Rev.* **162** 307
[97] Sannan S 1986 *Phys. Rev.* D **34** 1749
[98] van de Ven A E M 1992 *Nucl. Phys.* B **378** 309–66
[99] Cheung C and Remmen G N 2017 *J. High Energy Phys.* JHEP09(2017)002
[100] Peskin M E and Schroeder D V 1995 *An Introduction to Quantum Field Theory* (Addison-Wesley)
[101] Dongpei Z 1980 *Phys. Rev.* D **22** 2266
[102] Goebel C J, Halzen F and Leveille J P 1981 *Phys. Rev.* D **23** 2682–5
[103] Harland-Lang L A 2015 *J. High Energy Phys.* JHEP05(2015)146
[104] Bjerrum-Bohr N E J, Damgaard P H and Vanhove P 2009 *Phys. Rev. Lett.* **103** 161602
[105] Stieberger S 2009 arXiv:0907.2211
[106] Sondergaard T 2009 *Nucl. Phys.* B **821** 417–30
[107] Jia Y, Huang R and Liu C-Y 2010 *Phys. Rev.* D **82** 065001
[108] Mafra C R, Schlotterer O, Stieberger S and Tsimpis D 2011 *Nucl. Phys.* B **846** 359–93
[109] Mafra C R, Schlotterer O and Stieberger S 2013 *Nucl. Phys.* B **873** 419–60
[110] Ma Q, Du Y-J and Chen Y-X 2012 *J. High Energy Phys.* JHEP02(2012)061
[111] Grassi P A, Mezzalira A and Sommovigo L 2011 arXiv:1111.0544
[112] Barreiro L A and Medina R 2014 *Nucl. Phys.* B **886** 870–951
[113] Feng B, Huang R and Jia Y 2011 *Phys. Lett.* B **695** 350–3
[114] Mafra C R and Schlotterer O 2016 *J. High Energy Phys.* JHEP03(2016)097
[115] Cachazo F 2012 arXiv:1206.5970
[116] Naculich S G 2014 *J. High Energy Phys.* JHEP09(2014)029
[117] Weinzierl S 2015 *J. High Energy Phys.* JHEP03(2015)141
[118] de la Cruz L, Kniss A and Weinzierl S 2015 *J. High Energy Phys.* JHEP09(2015)197
[119] Bargheer T, He S and McLoughlin T 2012 *Phys. Rev. Lett.* **108** 231601
[120] Chiodaroli M, Günaydin M, Johansson H and Roiban R 2015 *J. High Energy Phys.* JHEP01(2015)081
[121] Chiodaroli M, Günaydin M, Johansson H and Roiban R 2016 *Phys. Rev. Lett.* **117** 011603
[122] Chiodaroli M, Günaydin M, Johansson H and Roiban R 2017 *J. High Energy Phys.* JHEP06(2017)064
[123] Chiodaroli M, Günaydin M, Johansson H and Roiban R 2018 *Phys. Rev. Lett.* **120** 171601
[124] Cachazo F, He S and Yuan E Y 2014 *Phys. Rev. Lett.* **113** 171601
[125] Cachazo F, He S and Yuan E Y 2015 *J. High Energy Phys.* JHEP07(2015)149
[126] Chen G and Du Y-J 2014 *J. High Energy Phys.* JHEP01(2014)061
[127] Abbott B P *et al* (LIGO Scientific, Virgo) 2016 *Phys. Rev. Lett.* **116** 061102
[128] Abbott B P *et al* (LIGO Scientific, Virgo) 2017 *Phys. Rev. Lett.* **119** 161101
[129] Blanchet L 2014 *Living Rev. Relativ.* **17** 2
[130] Buonanno A and Sathyaprakash B S 2014 Sources of gravitational waves: Theory and observations *General Relativity and Gravitation: A Centennial Perspective* ed A Ashtekar, B K Berger, J Isenberg and M A H MacCallum pp 287–346
[131] Porto R A 2016 *Phys. Rep.* **633** 1–104
[132] Levi M 2018 arXiv:1807.01699
[133] Stieberger S and Taylor T R 2016 *Nucl. Phys.* B **913** 151–62
[134] Schlotterer O 2016 *J. High Energy Phys.* JHEP11(2016)074
[135] Cachazo F, He S and Yuan E Y 2014 *Phys. Rev.* D **90** 065001
[136] Cachazo F, He S and Yuan E Y 2013 *J. High Energy Phys.* JHEP10(2013)141
[137] Cachazo F, He S and Yuan E Y 2014 *J. High Energy Phys.* **07** 033
[138] Mason L and Skinner D 2014 *J. High Energy Phys.* JHEP07(2014)048

[139] Adamo T, Casali E and Skinner D 2014 *J. High Energy Phys.* JHEP04(2014)104
[140] Casali E, Geyer Y, Mason L, Monteiro R and Roehrig K A 2015 *J. High Energy Phys.* JHEP11(2015)038
[141] He S and Yuan E Y 2015 *Phys. Rev.* D **92** 105004
[142] Geyer Y, Mason L, Monteiro R and Tourkine P 2016 *J. High Energy Phys.* JHEP03(2016)114
[143] Geyer Y, Mason L, Monteiro R and Tourkine P 2015 *Phys. Rev. Lett.* **115** 121603
[144] Cachazo F, He S and Yuan E Y 2016 *J. High Energy Phys.* JHEP08(2016)008
[145] Azevedo T and Engelund O T 2017 *J. High Energy Phys.* JHEP11(2017)052
[146] Geyer Y and Monteiro R 2018 *J. High Energy Phys.* JHEP03(2018)068
[147] Geyer Y and Monteiro R 2018 *J. High Energy Phys.* JHEP11(2018)008
[148] Geyer Y, Monteiro R and Stark-Muchão R 2019 arXiv:1908.05221
[149] Kiermaier M 2010 Gravity as the square of gauge theory *Amplitudes* (Queen Mary, University of London)
[150] Vaman D and Yao Y-P 2014 *J. High Energy Phys.* JHEP12(2014)036
[151] Mastrolia P, Primo A, Schubert U and Bobadilla W J T 2016 *Phys. Lett.* B **753** 242–62
[152] Johansson H and Nohle J 2017 arXiv:1707.02965
[153] Johansson H, Mogull G and Teng F 2018 *J. High Energy Phys.* JHEP09(2018)080
[154] Du Y-J, Feng B and Fu C-H 2011 *J. High Energy Phys.* JHEP08(2011)129
[155] Stieberger S and Taylor T R 2014 *Nucl. Phys.* B **881** 269–87
[156] Chiodaroli M, Günaydin M, Johansson H and Roiban R 2017 *J. High Energy Phys.* JHEP07(2017)002
[157] Bern Z and Dennen T 2011 *Phys. Rev. Lett.* **107** 081601
[158] Du Y-J, Feng B and Fu C-H 2013 *J. High Energy Phys.* JHEP07(2013)057
[159] Fu C-H, Du Y-J and Feng B 2013 *J. High Energy Phys.* JHEP10(2013)069
[160] Du Y-J, Feng B and Fu C-H 2014 *J. High Energy Phys.* JHEP06(2014)157
[161] Naculich S G 2014 *J. High Energy Phys.* JHEP07(2014)143
[162] Britto R, Cachazo F, Feng B and Witten E 2005 *Phys. Rev. Lett.* **94** 181602
[163] Bern Z, Dixon L J, Dunbar D C and Kosower D A 1994 *Nucl. Phys.* B **425** 217–60
[164] Bern Z, Dixon L J, Dunbar D C and Kosower D A 1995 *Nucl. Phys.* B **435** 59–101
[165] Bern Z, Dixon L J and Kosower D A 1998 *Nucl. Phys.* B **513** 3–86
[166] Britto R, Cachazo F and Feng B 2005 *Nucl. Phys.* B **725** 275–305
[167] Mafra C R 2010 *J. High Energy Phys.* JHEP01(2010)007
[168] Mafra C R and Schlotterer O 2014 *J. High Energy Phys.* JHEP07(2014)153
[169] Carrasco J J M, Mafra C R and Schlotterer O 2017 *J. High Energy Phys.* **06** 093
[170] Mafra C R 2016 *J. High Energy Phys.* JHEP07(2016)080
[171] Mafra C R and Schlotterer O 2017 *J. High Energy Phys.* JHEP01(2017)031
[172] Garozzo L M, Queimada L and Schlotterer O 2019 *J. High Energy Phys.* JHEP02(2019)078
[173] Arkani-Hamed N, Rodina L and Trnka J 2018 *Phys. Rev. Lett.* **120** 231602
[174] Rodina L 2016 arXiv: 1612.06342
[175] Du Y-J and Zhang Y 2018 *J. High Energy Phys.* JHEP07(2018)177
[176] Plefka J and Wormsbecher W 2018 *Phys. Rev.* D **98** 026011
[177] Hou L and Du Y J 2018 arXiv:1811.12653
[178] Brown R W and Naculich S G 2016 *J. High Energy Phys.* JHEP10(2016)130
[179] Brown R W and Naculich S G 2016 *J. High Energy Phys.* JHEP11(2016)060
[180] Del Duca V, Dixon L J and Maltoni F 2000 *Nucl. Phys.* B **571** 51–70
[181] Kleiss R and Kuijf H 1989 *Nucl. Phys.* B **312** 616–44
[182] t Hooft G 1974 *Nucl. Phys.* B **72** 461
t Hooft G 1973 *Nucl. Phys.* B **72** 337
[183] Scherk J and Schwarz J H 1974 *Nucl. Phys.* B **81** 118–44
[184] Gross D J and Sloan J H 1987 *Nucl. Phys.* B **291** 41–89
[185] Bjerrum-Bohr N E J, Damgaard P H, Feng B and Sondergaard T 2010 *Phys. Rev.* D **82** 107702
[186] Bjerrum-Bohr N E J, Damgaard P H, Feng B and Sondergaard T 2010 *J. High Energy Phys.* JHEP09(2010)067
[187] Johansson H, Sabio Vera A, Serna Campillo E and Vázquez-Mozo M A 2013 *J. High Energy Phys.* JHEP10(2013)215
[188] Johansson H and Ochirov A 2015 *J. High Energy Phys.* JHEP11(2015)046
[189] Johansson H and Ochirov A 2016 *J. High Energy Phys.* JHEP01(2016)170
[190] Johansson H and Ochirov A 2019 arXiv:1906.12292

[191] Melia T 2013 *Phys. Rev.* D **88** 014020
[192] Melia T 2015 *J. High Energy Phys.* JHEP12(2015)107
[193] Ochirov A and Page B 2019 arXiv:1908.02695
[194] Bern Z, Dixon L J, Dunbar D C, Perelstein M and Rozowsky J S 1998 *Nucl. Phys.* B **530** 401–56
[195] Arkani-Hamed N, Bai Y, He S and Yan G 2018 *J. High Energy Phys.* JHEP05(2018)096
[196] Carrasco J J M and Johansson H 2011 *J. Phys. A: Math. Theor.* **44** 454004
[197] Carrasco J J M 2015 Gauge and gravity amplitude relations *Proc., Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI2014) (Boulder, Colorado 2–27 June 2014)* (WSP)) pp 477–557
[198] Rafi K and Tao J 2013 *Duke Math. J.* **162** 1833–76
[199] Kalai G 1997 Polytope skeletons and paths *Handbook of Discrete and Computational Geometry (Discrete Mathematics and Its Applications) (Cambridge Monographs on Mathematical Physics)*, ed J E Goodman and J O'Rourke (CRC Press, Inc.)
[200] Stasheff J D 1963 *Trans. Am. Math. Soc.* **108** 293
[201] Tonks A 1995 Relating the associahedron and the permutohedron *Proc. Renaissance Conf.* pp 113–32
[202] Gell-Mann M and Lévy M 1960 *Il Nuovo Cimento* **16** 705–26
[203] MacFarlane A J, Sudbery A and Weisz P H 1968 *Commun. Math. Phys.* **11** 77–90
[204] Kampf K, Novotny J and Trnka J 2013 *J. High Energy Phys.* JHEP05(2013)032
[205] Carrasco J J M and Rodina L 2019 arXiv:1908.08033
[206] Cheung C, Kampf K, Novotny J, Shen C-H, Trnka J and Wen C 2018 *Phys. Rev. Lett.* **120** 261602
[207] Low I and Yin Z 2018 *J. High Energy Phys.* JHEP10(2018)078
[208] Rodina L 2019 *Phys. Rev. Lett.* **122** 071601
[209] Adler S L 1965 *Phys. Rev.* **137** B1022–33
Adler S L 1964 *Phys. Rev.* **140** B1022–33
[210] Carrillo-Gonzalez M, Penco R and Trodden M 2019 arXiv:1908.07531
[211] Mafra C R and Schlotterer O 2015 *Fortsch. Phys.* **63** 105–31
[212] de la Cruz L, Kniss A and Weinzierl S 2017 *Phys. Lett.* B **767** 86–90
[213] Fu C-H, Du Y-J, Huang R and Feng B 2017 *J. High Energy Phys.* JHEP09(2017)021
[214] Teng F and Feng B 2017 *J. High Energy Phys.* JHEP05(2017)075
[215] Chen G and Wang T 2017 arXiv:1709.08503
[216] Du Y-J, Feng B and Teng F 2017 *J. High Energy Phys.* JHEP12(2017)038
[217] Bern Z, Carrasco J J M, Johansson H and Kosower D A 2007 *Phys. Rev.* D **76** 125020
[218] Bern Z, Carrasco J J M, Chen W M, Johansson H, Roiban R and Zeng M 2017 *Phys. Rev.* D **96** 126012
[219] Kallosh R 2017 *J. High Energy Phys.* JHEP03(2017)038
[220] Kallosh R, Karlsson A and Murli D 2017 *J. High Energy Phys.* JHEP03(2017)081
[221] Karlsson A, Luo H and Murli D 2018 *Phys. Rev.* D **97** 045019
[222] Arkani-Hamed N, Cachazo F and Kaplan J 2010 *J. High Energy Phys.* JHEP09(2010)016
[223] Volkov D V and Akulov V P 1972 *JETP Lett.* **16** 438–40 Volkov D V and Akulov V P 1972 *Pisma Zh. Eksp. Teor. Fiz.* **16** 621
[224] Chen W-M, Huang Y-T and Wen C 2015 *Phys. Rev. Lett.* **115** 021603
[225] Conde E and Mao P 2017 *J. High Energy Phys.* JHEP05(2017)060
[226] Anupam A H, Kundu A and Ray K 2018 *Phys. Rev.* D **97** 106019
[227] Distler J, Flauger R and Horn B 2018 arXiv:1808.09965
[228] Huang Y t and Wen C 2015 *J. High Energy Phys.* **12** 143
[229] Elvang H, Jones C R,T and Naculich S G 2017 *Phys. Rev. Lett.* **118** 231601
[230] Di Vecchia P, Marotta R and Mojaza M 2015 *J. High Energy Phys.* JHEP05(2015)137
[231] Di Vecchia P, Marotta R and Mojaza M 2015 *J. High Energy Phys.* **12** 150
[232] Di Vecchia P, Marotta R and Mojaza M 2016 *J. High Energy Phys.* JHEP12(2016)020
[233] Di Vecchia P, Marotta R, Mojaza M and Nohle J 2016 *Phys. Rev.* D **93** 085015
[234] Bern Z, Davies S and Nohle J 2014 *Phys. Rev.* D **90** 085015
[235] Bern Z, Davies S, Di Vecchia P and Nohle J 2014 *Phys. Rev.* D **90** 084035
[236] Guerrieri A L, Huang Y-T, Li Z and Wen C 2017 *J. High Energy Phys.* JHEP12(2017)052
[237] Vecchia P, Marotta R and Mojaza M 2019 *J. High Energy Phys.* JHEP01(2019)038
[238] Di Vecchia P, Marotta R and Mojaza M 2019 *Phys. Rev.* D **100** 041902
[239] Anastasiou A, Borsten L, Hughes M J and Nagy S 2016 *J. High Energy Phys.* JHEP01(2016)148

[240] Anastasiou A, Borsten L, Duff M J, Marrani A, Nagy S and Zoccali M 2018 *Nucl. Phys.* B **934** 606–33
[241] Carrasco J J M, Chiodaroli M, Günaydin M and Roiban R 2013 *J. High Energy Phys.* JHEP03(2013)056
[242] Tourkine P and Vanhove P 2013 *Phys. Rev.* D **87** 045001
[243] Huang Y-T and Johansson H 2013 *Phys. Rev. Lett.* **110** 171601
[244] Huang Y-T, Johansson H and Lee S 2013 *J. High Energy Phys.* JHEP11(2013)050
[245] Bagger J and Lambert N 2008 *Phys. Rev.* D **77** 065008
[246] Gustavsson A 2009 *Nucl. Phys.* B **811** 66–76
[247] Bagger J and Lambert N 2009 *Phys. Rev.* D **79** 025002
[248] Bagger J and Bruhn G 2011 *Phys. Rev.* D **83** 025003
[249] Chen F-M 2010 *J. High Energy Phys.* JHEP08(2010)077
[250] Chen F-M and Wu Y-S 2010 *Phys. Rev.* D **82** 106012
[251] Marcus N and Schwarz J H 1983 *Nucl. Phys.* B **228** 145
[252] Borsten L, Duff M J, Hughes L J and Nagy S 2014 *Phys. Rev. Lett.* **112** 131601
[253] Anastasiou A, Borsten L, Duff M J, Hughes L J and Nagy S 2014 *Phys. Rev. Lett.* **113** 231606
[254] Nagy S 2016 *J. High Energy Phys.* JHEP07(2016)142
[255] Cardoso G L, Nagy S and Nampuri S 2016 *J. High Energy Phys.* JHEP10(2016)127
[256] Anastasiou A, Borsten L, Duff M J, Nagy S and Zoccali M 2018 *Phys. Rev. Lett.* **121** 211601
[257] Bern Z, De Freitas A and Wong H L 2000 *Phys. Rev. Lett.* **84** 3531
[258] Cremmer E and Julia B 1979 *Nucl. Phys.* B **159** 141–212
[259] Julia B 1980 *Conf. Proc.* C **8006162** 331–50
[260] Carrasco J J M, Kallosh R, Roiban R and Tseytlin A A 2013 *J. High Energy Phys.* JHEP07(2013)029
[261] Boels R, Larsen K J, Obers N A and Vonk M 2008 *J. High Energy Phys.* **11** 015
[262] Roiban R and Tseytlin A A 2012 *J. High Energy Phys.* JHEP10(2012)099
[263] Marcus N 1985 *Phys. Lett.* **157B** 383–8
[264] Cangemi D 1997 *Nucl. Phys.* B **484** 521–37
[265] Bern Z, Parra-Martinez J and Roiban R 2018 *Phys. Rev. Lett.* **121** 101604
[266] Bern Z, Kosower D and Parra-Martinez J 2019 arXiv:1905.0515
[267] Freedman D Z and Van Proeyen A 2012 *Supergravity* (Cambridge University Press)
[268] Anastasiou A, Borsten L, Duff M J, Hughes L J and Nagy S 2014 *J. High Energy Phys.* JHEP04(2014)178
[269] Elvang H, Hadjiantonis M, Jones C R T and Paranjape S 2019 arXiv:1906.05321
[270] Bern Z and Morgan A G 1996 *Nucl. Phys.* B **467** 479–509
[271] Hillmann C 2010 *J. High Energy Phys.* JHEP04(2010)010
[272] Bossard G, Hillmann C and Nicolai H 2010 *J. High Energy Phys.* JHEP12(2010)052
[273] Brödel J and Dixon L J 2010 *J. High Energy Phys.* JHEP05(2010)003
[274] Beisert N, Elvang H, Freedman D Z, Kiermaier M, Morales A and Stieberger S 2011 *Phys. Lett.* B **694** 265–71
[275] Bondi H, van der Burg M G J and Metzner A W K 1962 *Proc. R. Soc.* A **269** 21–52
[276] Sachs R K 1962 *Proc. R. Soc.* A **270** 103–26
[277] Sachs R 1962 *Phys. Rev.* **128** 2851–64
[278] Strominger A 2014 *J. High Energy Phys.* JHEP07(2014)152
[279] He T, Lysov V, Mitra P and Strominger A 2015 *J. High Energy Phys.* JHEP05(2015)151
[280] Kapec D, Lysov V, Pasterski S and Strominger A 2014 *J. High Energy Phys.* JHEP08(2014)058
[281] Larkoski A J, Neill D and Stewart I W 2015 *J. High Energy Phys.* JHEP06(2015)077
[282] de Wit B and Van Proeyen A 1992 *Commun. Math. Phys.* **149** 307–34
[283] Cachazo F, He S and Yuan E Y 2015 *J. High Energy Phys.* JHEP01(2015)121
[284] Chiodaroli M, Günaydin M, Johansson H and Roiban R 2018 (arXiv:1812.10434)
[285] Cheung C, Shen C-H and Wen C 2018 *J. High Energy Phys.* JHEP02(2018)095
[286] Cheung C, Kampf K, Novotny J, Shen C-H and Trnka J 2017 *J. High Energy Phys.* JHEP02(2017)020
[287] Freedman D Z, Ferrara S and van Nieuwenhuizen P 1976 *Phys. Rev.* D **13** 3214–8
[288] Samtleben H 2008 *Class. Quantum Grav.* **25** 214002
[289] Nandan D, Plefka J, Schlotterer O and Wen C 2016 *J. High Energy Phys.* JHEP10(2016)070
[290] Stieberger S and Taylor T R 2015 *Phys. Lett.* B **744** 160–2
[291] Bern Z, Carrasco J J, Dixon L J, Johansson H and Roiban R 2009 *Phys. Rev. Lett.* **103** 081301

[292] Bern Z, Davies S and Dennen T 2014 *Phys. Rev.* D **90** 105011
[293] Bern Z, Davies S, Dennen T, Smirnov A V and Smirnov V A 2013 *Phys. Rev. Lett.* **111** 231302
[294] Ben-Shahar M and Chiodaroli M 2019 *J. High Energy Phys.* JHEP03(2019)153
[295] Damgaard P H, Huang R, Sondergaard T and Zhang Y 2012 *J. High Energy Phys.* JHEP08(2012)101
[296] Broedel J and Dixon L J 2012 *J. High Energy Phys.* JHEP10(2012)091
[297] Sivaramakrishnan A 2017 *Int. J. Mod. Phys.* A **32** 1750002
[298] Du Y-J and Fu C-H 2016 *J. High Energy Phys.* JHEP09(2016)174
[299] Chen G, Du Y-J, Li S and Liu H 2015 *J. High Energy Phys.* JHEP03(2015)156
[300] Chen G, Li S and Liu H 2016 arXiv:1609.01832
[301] Cheung C, Remmen G N, Shen C-H and Wen C 2018 *J. High Energy Phys.* JHEP04(2018)129
[302] Bergshoeff E and Townsend P K 1997 *Nucl. Phys.* B **490** 145–62
[303] Bergshoeff E, Kallosh R, Ortin T and Papadopoulos G 1997 *Nucl. Phys.* B **502** 149–69
[304] Kallosh R 1997 *Lect. Notes Phys.* **509** 49
[305] Bergshoeff E, Coomans F, Kallosh R, Shahbazi C S and Van Proeyen A 2013 *J. High Energy Phys.* JHEP08(2013)100
[306] Cachazo F, Cha P and Mizera S 2016 *J. High Energy Phys.* JHEP06(2016)170
[307] He S and Schlotterer O 2017 *Phys. Rev. Lett.* **118** 161601
[308] Elvang H, Hadjiantonis M, Jones C R T and Paranjape S 2019 *J. High Energy Phys.* JHEP01(2019)195
[309] Cheung C and Shen C-H 2017 *Phys. Rev. Lett.* **118** 121601
[310] Carrasco J J M, Mafra C R and Schlotterer O 2017 *J. High Energy Phys.* JHEP08(2017)135
[311] Bershadsky M and Johansen A 1998 *Nucl. Phys.* B **536** 141–8
[312] Bershadsky M, Kakushadze Z and Vafa C 1998 *Nucl. Phys.* B **523** 59–72
[313] Cremmer E, Scherk J and Ferrara S 1978 *Phys. Lett.* **74B** 61–64
[314] Das A K 1977 *Phys. Rev.* D **15** 2805
[315] de Wit B and Freedman D Z 1977 *Nucl. Phys.* B **130** 105–13
[316] Fischler M 1979 *Phys. Rev.* D **20** 396–402
[317] Freedman D Z, Kallosh R and Yamada Y 2018 *Fortsch. Phys.* **66** 1800054
[318] Kallosh R 2019 *J. High Energy Phys.* JHEP05(2019)109
[319] Günaydin M and Kallosh R 2018 arXiv:1812.08758
[320] Bossard G, Howe P S and Stelle K S 2011 *J. High Energy Phys.* JHEP01(2011)020
[321] Bjornsson J and Green M B 2010 *J. High Energy Phys.* JHEP08(2010)132
[322] Vanhove P 2010 arXiv:1004.1392
[323] Howe P S and Lindstrom U 1981 *Nucl. Phys.* B **181** 487–501
[324] Kallosh R E 1981 *Phys. Lett.* **99B** 122–7
[325] Green M B, Russo J G and Vanhove P 2010 *J. High Energy Phys.* JHEP06(2010)075
[326] Bossard G, Howe P S, Stelle K S and Vanhove P 2011 *Class. Quantum Grav.* **28** 215005
[327] Chicherin D, Gehrmann T, Henn J M, Wasser P, Zhang Y and Zoia S 2019 *J. High Energy Phys.* JHEP03(2019)115
[328] Henn J M and Mistlberger B 2019 *J. High Energy Phys.* JHEP05(2019)023
[329] Günaydin M, Sierra G and Townsend P K 1984 *Nucl. Phys.* B **242** 244–68
[330] Günaydin M, Sierra G and Townsend P K 1985 *Nucl. Phys.* B **253** 573 Günaydin M, Sierra G and Townsend P K 1984 *Nucl. Phys.* B **253** 573
[331] Günaydin M, Sierra G and Townsend P K 1984 *Phys. Rev. Lett.* **53** 322
[332] Gunaydin M, Sierra G and Townsend P K 1986 *Class. Quantum Grav.* **3** 763
[333] Freedman D Z van Nieuwenhuizen P 1976 *Phys. Rev.* D **14** 912
[334] Deser S and Zumino B 1976 *Phys. Lett.* B **62** 335 Deser S and Zumino B 1976 *Phys. Lett.* B 335–7
[335] Ferrara S, van Nieuwenhuizen P and van Nieuwenhuizen P 1976 *Phys. Rev. Lett.* **37** 1669
[336] Freedman D Z 1977 *Phys. Rev. Lett.* **38** 105
[337] Ferrara S, Scherk J and Zumino B 1977 *Phys. Lett.* **66B** 35–38
[338] Bern Z, Cheung C, Chi H-H, Davies S, Dixon L and Nohle J 2015 *Phys. Rev. Lett.* **115** 211301
[339] Bern Z, Chi H-H, Dixon L and Edison A 2017 *Phys. Rev.* D **95** 046013
[340] Goroff M H and Sagnotti A 1986 *Nucl. Phys.* B **266** 709–36
[341] van de Ven Anton E.M. 1992 *Nucl. Phys.* B **378** 309–366
[342] t Hooft G and Veltman M J G 1974 *Ann. Inst. H. Poincare Phys. Theor.* A **20** 69–94
[343] Grisaru M T 1977 *Phys. Lett.* **66B** 75–76
[344] Tomboulis E 1977 *Phys. Lett.* **67B** 417–20

[345] Cecotti S, Ferrara S and Girardello L 1989 *Int. J. Mod. Phys.* A **4** 2475
[346] Schreiber A 2016 arXiv:1601.03028
[347] Anastasiou A, Borsten L, Duff M J, Hughes M J, Marrani A, Nagy S and Zoccali M 2017 *Phys. Rev.* D **96** 026013
[348] Nandan D, Plefka J and Travaglini G 2018 *J. High Energy Phys.* JHEP09(2018)011
[349] Butter D, Ciceri F, Sahoo B and de Wit B 2017 *Phys. Rev. Lett.* **118** 081602
[350] Berkovits N and Witten E 2004 *J. High Energy Phys.* JHEP08(2004)009
[351] de Roo M and de Roo M 1992 *Nucl. Phys.* B **372** 243–69
[352] Fradkin E S and Tseytlin A A 1984 *Phys. Lett.* **134B** 187
[353] Tseytlin A A 2017 *J. Phys. A: Math. Theor.* **50** 48LT01
[354] Adamo T, Nakach S and Tseytlin A A 2018 *J. High Energy Phys.* JHEP07(2018)016
[355] Broedel J, Schlotterer O and Stieberger S 2013 *Fortsch. Phys.* **61** 812–70
[356] Mizera S 2017 *J. High Energy Phys.* JHEP06(2017)084
[357] Stieberger S and Taylor T R 2013 *Nucl. Phys.* B **873** 65–91
[358] Fradkin E S and Tseytlin A A 1985 *Phys. Lett.* **163B** 123–30
[359] Metsaev R R, Rahmanov M and Tseytlin A A 1987 *Phys. Lett.* B **193** 207–12
[360] Aganagic M, Popescu C and Schwarz J H 1997 *Phys. Lett.* B **393** 311–5
[361] Aganagic M, Popescu C and Schwarz J H 1997 *Nucl. Phys.* B **495** 99–126
[362] Rocek M and Tseytlin A A 1999 *Phys. Rev.* D **59** 106001
[363] Schlotterer O and Schnetz O 2019 *J. Phys. A: Math. Theor.* **52** 045401
[364] Vanhove P and Zerbini F 2018 arXiv:1812.03018
[365] Brown F and Dupont C 2018 arXiv:1810.07682
[366] Schnetz O 2014 *Commun. Num. Theor. Phys.* **08** 589–675
[367] Brown F 2014 *SIGMA* **2** e25
[368] Schlotterer O and Stieberger S 2013 *J. Phys.* A **46** 475401
[369] Stieberger S 2014 *J. Phys. A: Math. Theor.* **47** 155401
[370] Azevedo T, Chiodaroli M, Johansson H and Schlotterer O 2018 *J. High Energy Phys.* JHEP10(2018)012
[371] Mafra C R and Schlotterer O 2018 *Phys. Rev. Lett.* **121** 011601
[372] Mafra C R and Schlotterer O 2014 *J. High Energy Phys.* JHEP08(2014)099
[373] Mafra C R and Schlotterer O 2018 arXiv:1812.10969
[374] Mafra C R and Schlotterer O 2018 arXiv:1812.10970
[375] Mafra C R and Schlotterer O 2018 arXiv:1812.10971
[376] Mafra C R and Schlotterer O 2019 arXiv:1908.09848
[377] Mafra C R and Schlotterer O 2019 arXiv:1908.10830
[378] He S, Teng F and Zhang Y 2019 *Phys. Rev. Lett.* **122** 211603
[379] He S, Teng F and Zhang Y 2019 arXiv:1907.06041
[380] Bianchi L and Bianchi M S 2014 *Phys. Rev.* D **89** 125002
[381] Gustavsson A 2008 *J. High Energy Phys.* JHEP04(2008)083
[382] Aharony O, Bergman O, Jafferis D L and Maldacena J 2008 *J. High Energy Phys.* JHEP10(2008)091
[383] Ferrara S and Lüst D 2018 *J. High Energy Phys.* JHEP07(2018)114
[384] Borsten L, Duff M J and Marrani A 2019 *J. High Energy Phys.* JHEP03(2019)112
[385] Ponomarev D 2017 *J. High Energy Phys.* JHEP12(2017)141
[386] de la Cruz L, Kniss A and Weinzierl S 2016 *Phys. Rev. Lett.* **116** 201601
[387] Anastasiou A, Borsten L, Duff M J, Marrani A, Nagy S and Zoccali M 2017 *The Mile High Magic Pyramid Contemporary Mathematics* vol 721 (Amer. Math. Soc.) pp 1–27
[388] Berg M, Buchberger I and Schlotterer O 2017 *J. High Energy Phys.* JHEP07(2017)138
[389] Hull C M 2000 *Nucl. Phys.* B **583** 237–59
[390] Chiodaroli M, Günaydin M and Roiban R 2012 *J. High Energy Phys.* JHEP03(2012)093
[391] Borsten L 2018 *Phys. Rev.* D **97** 066014
[392] Green M B, Schwarz J H and Brink L 1982 *Nucl. Phys.* B **198** 474–92
[393] Bern Z and Kosower D A 1992 *Nucl. Phys.* B **379** 451–561
[394] Bern Z, Dixon L J and Kosower D A 1994 *Nucl. Phys.* B **412** 751–816
[395] Bern Z, Dixon L J, Dunbar D C and Kosower D A 1997 *Phys. Lett.* B **394** 105–15
[396] Nohle J 2014 *Phys. Rev.* D **90** 025020
[397] Bern Z, Dixon L J and Kosower D A 1993 *Phys. Rev. Lett.* **70** 2677–80
[398] Kälin G, Mogull G and Ochirov A 2018 arXiv:1811.09604

[399] Bern Z, Rozowsky J S and Yan B 1997 *Phys. Lett.* B **401** 273–82
[400] Abreu S, Dixon L J, Herrmann E, Page B and Zeng M 2019 *Phys. Rev. Lett.* **122** 121603
[401] Chicherin D, Gehrmann T, Henn J M, Wasser P, Zhang Y and Zoia S 2019 *Phys. Rev. Lett.* **122** 121602
[402] Abreu S, Dixon L J, Herrmann E, Page B and Zeng M 2019
[403] Bern Z, Bjerrum-Bohr N E J and Dunbar D C 2005 *J. High Energy Phys.* **05** 056
[404] Bjerrum-Bohr N E J, Dunbar D C, Ita H, Perkins W B and Risager K 2006 *J. High Energy Phys.* **12** 072
[405] Broedel J and Carrasco J J M 2011 *Phys. Rev.* D **84** 085009
[406] Fu C-H, Du Y-J and Feng B 2014 *J. High Energy Phys.* JHEP08(2014)098
[407] Bern Z, Carrasco J J, Dixon L J, Johansson H, Kosower D A and Roiban R 2007 *Phys. Rev. Lett.* **98** 161303
[408] Bern Z, Carrasco J J M, Dixon L J, Johansson H and Roiban R 2008 *Phys. Rev.* D **78** 105019
[409] Tourkine P and Vanhove P 2016 *Phys. Rev. Lett.* **117** 211601
[410] Ochirov A, Tourkine P and Vanhove P 2017 *J. High Energy Phys.* JHEP10(2017)105
[411] Boels R H and Isermann R S 2012 *J. High Energy Phys.* JHEP03(2012)051
[412] Sabio V A and Vazquez-Mozo M A 2015 *J. High Energy Phys.* JHEP03(2015)070
[413] Chester D 2016 *Phys. Rev.* D **93** 065047
[414] Primo A and Torres Bobadilla W J 2016 *J. High Energy Phys.* JHEP04(2016)125
[415] Bern Z, Davies S and Nohle J 2016 *Phys. Rev.* D **93** 105015
[416] Bern Z, Carrasco J J, Chen W M, Johansson H and Roiban R 2017 *Phys. Rev. Lett.* **118** 181602
[417] Tourkine P 2019 arXiv:1901.02432
[418] Henry Tye S-H and Zhang Y 2010 *J. High Energy Phys.* JHEP06(2010)071
Henry Tye S H and Zhang Y 2011 *J. High Energy Phys.* **04** 114 (erratum)
[419] Bjerrum-Bohr N E J, Damgaard P H, Sondergaard T and Vanhove P 2010 *J. High Energy Phys.* JHEP06(2010)003
[420] Vaman D and Yao Y-P 2010 *J. High Energy Phys.* JHEP11(2010)028
[421] Holstein B R and Donoghue J F 2004 *Phys. Rev. Lett.* **93** 201602
[422] Neill D and Rothstein I Z 2013 *Nucl. Phys.* B **877** 177–89
[423] Damour T 2018 *Phys. Rev.* D **97** 044038
[424] Luna A, Nicholson I, O'Connell D and White C D 2018 *J. High Energy Phys.* JHEP03(2018)044
[425] Bjerrum-Bohr N E J, Damgaard P H, Festuccia G, Planté L and Vanhove P 2018 *Phys. Rev. Lett.* **121** 171601
[426] Maybee B, O'Connell D and Vines J V 2019
[427] Boulware D G and Brown L S 1968 *Phys. Rev.* **172** 1628–31
[428] Mahlon G, Yan T-M and Dunn C 1993 *Phys. Rev.* D **48** 1337–74
[429] Mahlon G 1994 *Phys. Rev.* D **49** 4438–53
[430] Berends F A and Giele W T 1988 *Nucl. Phys.* B **306** 759–808
[431] Selivanov K G 1997 *Mod. Phys. Lett.* A **12** 3087–90
[432] Selivanov K G 1998 *Phys. Lett.* B **420** 274–8
[433] Mafra C R and Schlotterer O 2015 *Phys. Rev.* D **92** 066001
[434] Mizera S and Skrzypek B 2018 *J. High Energy Phys.* JHEP10(2018)018
[435] Lee S, Mafra C R and Schlotterer O 2016 *J. High Energy Phys.* JHEP03(2016)090
[436] Duff M J 1973 *Phys. Rev.* D **7** 2317–26
[437] Sardelis D A 1973 The tree graphs of quantum gravity and the Reissner-Nordstrom solution IC/73/186
[438] Günaydin M 2010 *The Attractor Mechanism* (*Springer Proc. Phys.* vol 134) (Springer) pp 31–84
[439] Dray T and 't Hooft G 1985 *Nucl. Phys.* B **253** 173–88
[440] Stephani H, Kramer D, MacCallum M A H, Hoenselaers C and Herlt E 2003 *Exact Solutions of Einstein's Field Equations (Cambridge Monographs on Mathematical Physics*) (Cambridge University Press)
[441] Chong Z-W, Gibbons G W, Lu H and Pope C N 2005 *Phys. Lett.* B **609** 124–32
[442] Stephani H 1982 *General Relativity. An Introduction to the Theory of the Gravitational Field* (University Press) Transl. From German By M. Pollock and J. Stewart) p 298
[443] Cho W and Lee K 2019 *J. High Energy Phys.* JHEP07(2019)030
[444] Barack L *et al* (LIGO) 2019 *Class. Quantum Grav.* **36** 143001
[445] Chester D 2018 *Phys. Rev.* D **97** 084025
[446] Laenen E, Stavenga G and White C D 2009 *J. High Energy Phys.* **03** 054

[447] White C D 2011 *J. High Energy Phys.* JHEP05(2011)060
[448] Damour T, Jaranowski P and Schäfer G 2014 *Phys. Rev.* D **89** 064058
[449] Jaranowski P and Schäfer G 2015 *Phys. Rev.* D **92** 124043
[450] Foffa S, Mastrolia P, Sturani R and Sturm C 2017 *Phys. Rev.* D **95** 104009
[451] Foffa S, Porto R A, Rothstein I and Sturani R 2019 arXiv:1410.7590
[452] Bjerrum-Bohr N E J, Donoghue J F, Holstein B R, Planté L and Vanhove P 2015 *Phys. Rev. Lett.* **114** 061301
[453] Bjerrum-Bohr N E J, Donoghue J F, Holstein B R, Plante L and Vanhove P 2016 *J. High Energy Phys.* JHEP11(2016)117
[454] Bjerrum-Bohr N E J, Holstein B R, Donoghue J F, Planté L and Vanhove P 2017 *PoS (CORFU2016)* p 077
[455] Goldberger W D and Rothstein I Z 2006 *Phys. Rev.* D **73** 104029
[456] Arkani-Hamed N, Huang Y T and O'Connell D 2019 arXiv:1906.10100
[457] Engelund O T and Roiban R 2013 *J. High Energy Phys.* JHEP03(2013)172
[458] Adamo T and Ilderton A 2019 *J. High Energy Phys.* JHEP06(2019)015
[459] Sachs I and Tran T 2019 arXiv:1902.08409
[460] Raju S 2012 *Phys. Rev.* D **85** 126009
[461] Raju S 2012 *Phys. Rev.* D **85** 126008
[462] Maldacena J M and Pimentel G L 2011 *J. High Energy Phys.* JHEP09(2011)045
[463] Arkani-Hamed N and Maldacena J 2015 arXiv:1503.08043
[464] Arkani-Hamed N and Benincasa P 2018 arXiv:1811.01125
[465] Arkani-Hamed N, Baumann D, Lee H and Pimentel G L 2018 arXiv:1811.00024
[466] Bourjaily J L, Herrmann E and Trnka J 2019 *Phys. Rev.* D **99** 066006
[467] Hohm O 2011 *J. High Energy Phys.* JHEP04(2011)103
[468] Bern Z, Enciso M, Parra-Martinez J and Zeng M 2017 *J. High Energy Phys.* JHEP05(2017)137
[469] Arkani-Hamed N and Trnka J 2014 *J. High Energy Phys.* JHEP10(2014)030
[470] Bern Z, Herrmann E, Litsey S, Stankowicz J and Trnka J 2016 *J. High Energy Phys.* JHEP06(2016)098
[471] Weinberg S and Witten E 1980 *Phys. Lett.* **96B** 59–62
[472] Siegel W 1994 *Phys. Rev.* D **49** 4144–53
[473] Siegel W 2003 arXiv:hep-th/0309093
[474] Lee K and Siegel W 2003 *Nucl. Phys.* B **665** 179–88
[475] Einstein A, Infeld L and Hoffmann B 1938 *Ann. Math.* **39** 65–100
[476] Bertotti B 1956 *Nuovo Cim.* **4** 898–906
[477] Kerr R P 1959 *I. Nuovo Cimento* **13** 469–91
[478] Bjerrum-Bohr N E J, Donoghue J F and Vanhove P 2014 *J. High Energy Phys.* JHEP02(2014)111
[479] Guevara A 2019 *J. High Energy Phys.* JHEP04(2019)033
[480] Guevara A, Ochirov A and Vines J 2018 arXiv:1812.06895
[481] Guevara A, Ochirov A and Vines J 2019 arXiv:1906.10071
[482] Westpfahl K 1985 *Fortsch. Phys.* **33** 417
[483] Damour T 2016 *Phys. Rev.* D **94** 104015
[484] Berends F A, Kleiss R, De Causmaecker P, Gastmans R and Wu T T 1981 *Phys. Lett.* **103B** 124–8
[485] Kleiss R and Stirling W J 1985 *Nucl. Phys.* B **262** 235–62
[486] Gunion J F and Kunszt Z 1985 *Phys. Lett.* **161B** 333
[487] Xu Z, Zhang D H and Chang L 1984
[488] Xu Z, Zhang D-H and Chang L 1987 *Nucl. Phys.* B **291** 392–428
[489] Nair V P 1988 *Phys. Lett.* B **214** 215–8
[490] Bern Z, De Freitas A, Dixon L J and Wong H L 2002 *Phys. Rev.* D **66** 085002
[491] Craig N, Elvang H, Kiermaier M and Slatyer T 2011 *J. High Energy Phys.* JHEP12(2011)097
[492] Arkani-Hamed N, Huang T C and Huang Y T 2017 arXiv:1709.04891
[493] Cheung C and O'Connell D 2009 *J. High Energy Phys.* JHEP07(2009)075
[494] Elvang H, Huang Y-T and Peng C 2011 *J. High Energy Phys.* JHEP09(2011)031
[495] Bianchi M, Elvang H and Freedman D Z 2008 *J. High Energy Phys.* JHEP09(2008)063
[496] Elvang H, Freedman D Z and Kiermaier M 2010 *J. High Energy Phys.* JHEP10(2010)103
[497] Brandhuber A, Spence B and Travaglini G 2011 *J. Phys. A: Math. Theor.* **44** 454002
[498] Drummond J M, Henn J, Korchemsky G P and Sokatchev E 2013 *Nucl. Phys.* B **869** 452–92
[499] Bern Z, Carrasco J J M, Ita H, Johansson H and Roiban R 2009 *Phys. Rev.* D **80** 065029
[500] Huang Y T 2011 arXiv:1104.2021

[501] Drummond J M, Henn J, Korchemsky G P and Sokatchev E 2010 *Nucl. Phys.* B **828** 317–74
[502] Drummond J M and Henn J M 2009 *J. High Energy Phys.* **04** 018
[503] Feng B and He S 2010 *J. High Energy Phys.* JHEP09(2010)043
[504] Dixon L J, Henn J M, Plefka J and Schuster T 2011 *J. High Energy Phys.* JHEP01(2011)035
[505] Cachazo F, Svrcek P and Witten E 2004 *J. High Energy Phys.* **09** 006
[506] Georgiou G, Glover E W N and Khoze V V 2004 *J. High Energy Phys.* **07** 048
[507] Bern Z and Huang Y T 2011 *J. Phys. A: Math. Theor.* **44** 454003
[508] Britto R 2011 *J. Phys. A: Math. Theor.* **44** 454006
[509] Bourjaily J L, Herrmann E and Trnka J 2017 *J. High Energy Phys.* JHEP06(2017)059
[510] Eden R J, Landshoff P V, Olive D I and Polkinghorne J C 1966 *The Analytic S-Matrix* (Cambridge University Press)
[511] Berger C F, Bern Z, Dixon L J, Febres Cordero F, Forde D, Ita H, Kosower D A and Maitre D 2008 *Phys. Rev.* D **78** 036003
[512] Bern Z, Dixon L J, Febres Cordero F, Höche S, Ita H, Kosower D A, Maître D and Ozeren K J 2013 *Phys. Rev.* D **88** 014025
[513] Koemans Collado A, Di Vecchia P and Russo R 2019 arXiv:1904.02667
[514] Bern Z, Carrasco J J M, Johansson H and Roiban R 2012 *Phys. Rev. Lett.* **109** 241602
[515] Elvang H, Freedman D Z and Kiermaier M 2009 *J. High Energy Phys.* **04** 009
[516] Mandelstam S 1983 *Nucl. Phys.* B **213** 149–68
[517] Brink L, Lindgren O and Nilsson B E W 1983 *Phys. Lett.* **123B** 323–8
[518] Howe P S, Stelle K S and Townsend P K 1983 *Nucl. Phys.* B **214** 519–31