UUITP-44/22

Tree-level amplitudes from the pure spinor superstring

Carlos R. Mafra[a], Oliver Schlotterer[b]

[a]*STAG Research Centre and Mathematical Sciences, University of Southampton, Highfield, Southampton SO17 1BJ, UK*
[b]*Department of Physics and Astronomy, Uppsala University, 75108 Uppsala, Sweden*

Abstract

We give a comprehensive review of recent developments on using the pure spinor formalism to compute massless superstring scattering amplitudes at tree level. The main results of the pure spinor computations are placed into the context of related topics including the color-kinematics duality in field theory and the mathematical structure of α'-corrections.

Keywords: Pure spinors, scattering amplitudes, multiparticle superfields

Contents

Email addresses: c.r.mafra@soton.ac.uk (Carlos R. Mafra), oliver.schlotterer@physics.uu.se (Oliver Schlotterer)

1. Introduction

Superstring theories offer ultraviolet completions of supersymmetric gauge theories and supergravity in $D \leq 10$ spacetime dimensions. Since gauge and gravity supermultiplets are realized through the massless vibration modes of open and closed superstrings, respectively, their interactions are naturally unified: superstring scattering amplitudes are computed from a topological expansion in terms of random fluctuating surfaces dubbed *worldsheets* that automatically incorporate the interplay of open and closed strings via splitting and joining.

The worldsheet origin of string amplitudes is a rich source of structure and information. First, it realizes the connection between gauge theories and gravity through the Bern-Carrasco-Johansson (BCJ) double copy in a geometrically intuitive manner. Second, the computation of string-corrections to field-theory amplitudes from moduli spaces of punctured worldsheets reveals intriguing mathematical structures and cross-fertilizes with string dualities. In order to bring these appealing implications of string amplitudes to their full fruition, it is important to have detailed control over their explicit form and hence efficient methods to organize their computation.

The worldsheet degrees of freedom underlying superstring theories and their amplitudes admit a variety of formulations. The more recent pure spinor formalism developed by Berkovits since the year 2000 [1, 2, 3] led to the first manifestly super-Poincaré invariant quantization of the superstring. The more traditional Ramond-Neveu-Schwarz (RNS) [4, 5, 6, 7] and Green-Schwarz (GS) [8, 9] formalisms are for instance described in textbooks on string theory including [10, 11, 12, 13, 14, 15, 16, 17] and differ in the implementation of worldsheet and spacetime supersymmetry. The equivalence of these formalisms is widely expected based on [18, 19, 20] and explicitly confirmed for leading orders in string perturbation theory but in general a subject of ongoing research.

This is a comprehensive review of the state of the art regarding the computation of massless superstring tree-level amplitudes in Minkowski spacetime with the pure spinor formalism. We will illustrate in detail how the manifest spacetime supersymmetry of the pure spinor formalism simplifies computations and efficiently organizes the information on the external gauge and gravity multiplets. The main results of this review include compact expressions for superstring tree-level amplitudes with an arbitrary number of massless external states revealed by a pure spinor computation in 2011 [21]. These expressions will be shown to elegantly resonate with a web of double-copy relations between a wide range of string- and field theories as well as number-theoretic properties of the low-energy expansion.

1.1. Summary of the main results

Throughout this review, the topics are presented in an order which emphasizes completeness rather than brevity. As such, the topics are developed to a depth higher than what is usually necessary for a brief application of certain parts of the formalism.[1] This is unavoidable in a comprehensive review but it can be mitigated by jumping to the topic of interest and choosing to pick up the minimal background as one goes along. This section aims to give an overview of the main results in this review along with pointers that facilitate the identification of key passages on a given topic. References to original work can be found in the main text.

1.1.1. Basics of the pure spinor formalism

We start by summarizing the worldsheet variables in the (minimal)[2] pure spinor formalism in ten-dimensional Minkowski spacetime, based on selected aspects of section 3. Center stage is taken by the worldsheet action in (3.49)

$$S_{\rm PS} = \frac{1}{\pi}\int d^2z \left(\frac{1}{2}\partial X^m \overline{\partial} X_m + p_\alpha \overline{\partial}\theta^\alpha - w_\alpha \overline{\partial}\lambda^\alpha\right) \tag{1.1}$$

[1]We welcome the readers' help in spotting typos or technical mistakes. Every correction that is firstly brought to our attention will be rewarded with 20 Euro Cent per numbered equation, to be paid in cash during the next in-person encounter with one of the authors.

[2]See [3] for the "non-minimal" pure spinor formalism with additional worldsheet variables.

with $\partial = \frac{\partial}{\partial z}$ and $\overline{\partial} = \frac{\partial}{\partial \bar{z}}$, and we shall now give a brief characterization of its main ingredients. Just like the bosonic string and the RNS or GS formulation of the superstring, the embedding coordinates X^m (with vector indices $m, n, \ldots = 0, 1, \ldots, 9$) enter (1.1) as free worldsheet bosons. In parallel to Siegel's reformulation of the GS formalism [22], the matter sector of the pure spinor worldsheet action (1.1) also features a pair of anticommuting spacetime spinors $(p_\alpha, \theta^\alpha)$ of holomorphic conformal weights $h_p = 1$ and $h_\theta = 0$ (with Weyl-spinor indices $\alpha, \beta, \ldots = 1, 2, \ldots, 16$ of $SO(1,9)$).

The main characteristic of the pure spinor formalism is the pair of commuting ghost variables $(w_\alpha, \lambda^\alpha)$ of holomorphic conformal weights $h_w = 1$ and $h_\lambda = 0$. They are spacetime spinors in contradistinction to the anticommuting scalar (b, c)-system of bosonic strings or RNS superstrings. Cancellation of conformal anomalies and nilpotency of the BRST charge $Q_{\rm BRST}$ below requires λ^α to obey the pure spinor constraint

$$(\lambda\gamma^m\lambda) = 0\,, \qquad Q_{\rm BRST} = \oint dz\, \lambda^\alpha d_\alpha\,, \qquad d_\alpha = p_\alpha - \tfrac{1}{2}\partial X^m(\gamma_m\theta)_\alpha - \tfrac{1}{8}(\theta\gamma^m\partial\theta)(\gamma_m\theta)_\alpha \tag{1.2}$$

with 16×16 Pauli matrices $\gamma^m_{\alpha\beta} = \gamma^m_{\beta\alpha}$ of $SO(1,9)$. Further details on the worldsheet ghosts and their contributions N^{mn} to the Lorentz currents can be found in section 3.3.

We have only displayed the left-moving spacetime spinors in (1.1). The pure spinor formulation of type II superstrings involves right-moving counterparts $(\tilde{p}_{\tilde{\alpha}}, \tilde{\theta}^{\tilde{\alpha}})$ and $(\tilde{w}_{\tilde{\alpha}}, \tilde{\lambda}^{\tilde{\alpha}})$ with $\overline{\partial}$ in the place of ∂ in the action. The Weyl-spinor indices $\tilde{\alpha}$ are of the same (opposite) chirality as the indices α of the left-movers in (1.1) in case of the type IIB (type IIA) theory. One can also construct a pure spinor version of heterotic strings by incorporating right-moving bosons for compactified 16 extra dimensions into (1.1) instead of $(\tilde{p}_{\tilde{\alpha}}, \tilde{\theta}^{\tilde{\alpha}})$ and $(\tilde{w}_{\tilde{\alpha}}, \tilde{\lambda}^{\tilde{\alpha}})$, see section 7.5 for a brief discussion of its amplitudes.

1.1.2. The prescription for disk amplitudes

The physical spectrum of the pure spinor superstring is constructed from the cohomology of the BRST charge in (1.2). As usual in worldsheet approaches to string theories, physical states are associated with vertex operators V and U, conformal primaries of weight $h_V = 0$ and $h_U = 1$ in the BRST cohomology. For massless states of the open superstring, the integrated and unintegrated representatives of the vertex operators are

$$V = \lambda^\alpha A_\alpha\,, \qquad \int dz\, U = \int dz\, \big(\partial\theta^\alpha A_\alpha + A_m \Pi^m + d_\alpha W^\alpha + \tfrac{1}{2}N_{mn}F^{mn}\big)\,, \tag{1.3}$$

see the discussion around (3.59) and (3.62). They combine the worldsheet variables in (1.1), (1.2) and $\Pi^m = \partial X^m + \frac{1}{2}(\theta\gamma^m\partial\theta)$ with linearized superfields $A_\alpha, A^m, W^\alpha, F^{mn}$ of ten-dimensional super Yang–Mills (SYM) reviewed in section 2, depending on the worldsheet variables X^m and θ^α but not on their derivatives.

The main subject of this review are the superstring disk amplitudes obtained from the vertex operators in (1.3) through the prescription [1]

$$\mathcal{A}(1,2,\ldots,n) = \int\limits_{-\infty<z_j<z_{j+1}<\infty} dz_2\, dz_3 \ldots dz_{n-2}\, \langle\!\langle V_1(z_1)U_2(z_2)U_3(z_3)\ldots U_{n-2}(z_{n-2})V_{n-1}(z_{n-1})V_n(z_n)\rangle\!\rangle\,, \tag{1.4}$$

see section 3.4.2. The integration domain informally refers to an ordering of the vertex-operator insertions on the disk boundary parameterized by $-\infty < z_1 < z_2 < \ldots < z_n < \infty$ which is associated with the Chan–Paton trace in the cyclic ordering ${\rm Tr}(t^{a_1}t^{a_2}\ldots t^{a_n})$. The correlators $\langle\!\langle\ldots\rangle\!\rangle$ arise from the path integral over the worldsheet variables, and the contributions from their non-zero modes can be evaluated from the OPEs encoded by (1.1). The zero modes of the variables $\lambda^\alpha, \theta^\alpha$ with conformal weight $h_\theta = h_\lambda = 0$ require a separate prescription

$$\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{mnp}\theta)\rangle = 2880 \tag{1.5}$$

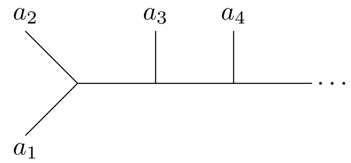
which automatically fixes zero-mode correlators of arbitrary tensor contractions of $\lambda^\alpha\lambda^\beta\lambda^\gamma\theta^{\delta_1}\theta^{\delta_2}\theta^{\delta_3}\theta^{\delta_4}\theta^{\delta_5}$ via simple group-theoretic considerations. Consistency conditions on (1.4) and the extraction of three-point component results for external gluons and gluinos are reviewed in detail in section 3.4. Massless n-point tree-level amplitudes of type II superstrings and heterotic strings are obtained by integrating double copies of the correlator in (1.4) over the sphere, see sections 7.2 and 7.5.

1.1.3. The multiparticle formalism

The driving force for the simplification of the n-point disk amplitude (1.4) is the organization of the OPEs among the vertex operators in (1.3) through multiparticle superfields. The local incarnations $A_\alpha^P, A_P^m, W_P^\alpha, F_P^{mn}$ of multiparticle superfields (with words or ordered sequences $P = p_1 p_2 \ldots$ in external-state labels p_i) are obtained from nested OPEs by systematically discarding total derivatives and BRST-exact terms in each step. The construction of section 4.1.6 in so-called BCJ gauge implements a specific scheme of discarding spurious terms and leads to multiparticle superfields with generalized Jacobi identities under permutations of P, e.g.

$$A_{12}^m = -A_{21}^m\,, \qquad A_{123}^m = -A_{213}^m\,, \qquad A_{123}^m + A_{231}^m + A_{312}^m = 0\,. \tag{1.6}$$

More generally, the symmetries of $A_{1234\ldots}^m$ and all the other local multiparticle superfields in BCJ gauge are those of contracted structure constants $f^{a_1 a_2 b} f^{b a_3 c} f^{c a_4 d} \ldots$ corresponding to the half-ladder graph



Accordingly, the multiparticle superfields inherit a diagrammatic interpretation that resonates with the BCJ duality between color and kinematics in gauge theories and can be described in a combinatorial framework based on planar binary trees, see section 4.3.

By dressing the $A_\alpha^P, A_P^m, W_P^\alpha, F_P^{mn}$ with the propagators of the associated cubic-vertex diagrams, one is led to non-local superfields or Berends–Giele currents $\mathcal{A}_\alpha^P, \mathcal{A}_P^m, \mathcal{W}_P^\alpha, \mathcal{F}_P^{mn}$ in *BCJ gauge.* This is an alternative to the construction of Berends–Giele currents in Lorenz gauge via perturbiner methods [23, 24, 25, 26]: the wave equations (4.92) of ten-dimensional SYM encode recursions for Lorenz-gauge currents such as

$$\mathcal{A}_\alpha^{12\ldots p} = \frac{1}{k_{12\ldots p}^2} \sum_{j=1}^{p-1} \big[\mathcal{A}_\alpha^{j+1\ldots p}(k_{j+1\ldots p} \cdot \mathcal{A}_{12\ldots j}) + \mathcal{A}_{j+1\ldots p}^m (\gamma_m \mathcal{W}_{12\ldots j})_\alpha - (12\ldots j \leftrightarrow j{+}1\ldots p)\big] \tag{1.7}$$

with $k_{ij\ldots} = k_i + k_j + \ldots$ and similar ones for $\mathcal{A}_P^m, \mathcal{W}_P^\alpha$ and $\mathcal{F}_P^{mn}$, see (4.95). These recursions terminate with the linearized superfields in the vertex operators (1.3) in the single-particle case, e.g. $\mathcal{A}_\alpha^j = A_\alpha^j$. Berends–Giele currents in Lorenz and BCJ gauge obey the same multiparticle equations of motion (4.97) such as

$$D_\alpha \mathcal{A}_\beta^{12\ldots p} + D_\beta \mathcal{A}_\alpha^{12\ldots p} = \gamma_{\alpha\beta}^m \mathcal{A}_m^{12\ldots p} + \sum_{j=1}^{p-1} (\mathcal{A}_\alpha^{12\ldots j} \mathcal{A}_\beta^{j+1\ldots p} - \mathcal{A}_\alpha^{j+1\ldots p} \mathcal{A}_\beta^{12\ldots j}) \tag{1.8}$$

and their BRST-invariant combinations yield the same amplitudes.

1.1.4. SYM tree-level amplitudes

Multiparticle superfields turned out to be invaluable to determine tree and loop amplitudes in string and field theory from first principles including locality and BRST invariance. As a simple manifestation thereof, color-ordered n-point tree amplitudes of ten-dimensional SYM obey the compact formula presented in (5.13),

$$A(1,2,\ldots,n) = \sum_{j=1}^{n-2} \langle M_{12\ldots j} M_{j+1\ldots n-1} M_n \rangle\,. \tag{1.9}$$

The $M_{12\ldots j}$ may be viewed as non-local multiparticle vertex operators defined by the spinorial Berends–Giele currents in (1.7) and whose BRST variation follows from the multiparticle equations of motion (1.8)

$$M_P = \lambda^\alpha \mathcal{A}_\alpha^P\,, \qquad Q_{\mathrm{BRST}} M_{12\ldots p} = \sum_{j=1}^{p-1} M_{12\ldots j} M_{j+1\ldots p}\,. \tag{1.10}$$

Since the superfields in the zero-mode bracket of (1.9) are easily checked to be BRST invariant via (1.10), the component amplitudes following from the zero-mode prescription (1.5) are guaranteed to be gauge invariant and supersymmetric. An efficient Berends–Giele organization of the component amplitudes is described in section 5.2.2 which follows from (1.9) and a combination of Lorenz and Harnad–Shnider gauge for $\mathcal{A}^P_\alpha$. In particular, this implies the bosonic components of (1.9) to reproduce the Berends–Giele formula [27] for n-gluon tree amplitudes.

The superspace formula (1.9) is a convenient starting point to prove the Kleiss–Kuijf (KK) and BCJ relations between SYM tree amplitudes in different color orderings, see sections 5.2.4 and 5.2.5. The KK relations [28] can be written as $A(P \sqcup\!\sqcup Q, n) = 0 \ \forall \ P, Q \neq \emptyset$ with the shuffle operation defined in (C.5) and follow from the shuffle properties $M_{P \sqcup\!\sqcup Q} = 0 \ \forall \ P, Q \neq \emptyset$ of the currents in (1.10). The BCJ relations [29] take the form $A(\{P, Q\}, n) = 0 \ \forall \ P, Q \neq \emptyset$ with the so-called S-bracket $\{\cdot, \cdot\}$ defined in (4.142) and are derived from multiparticle superfields in BCJ gauge using the vanishing of BRST-exact expressions under the zero-mode prescription (1.5), $\langle Q(\ldots)\rangle = 0$.

1.1.5. Superstring disk amplitudes

As a key result of this review, color-ordered superstring disk amplitudes $\mathcal{A}(P)$ with any number of external gauge multiplets are reduced to SYM tree amplitudes $A(Q)$ in a BCJ basis of color orderings Q,

$$\mathcal{A}(1, P, n{-}1, n; \alpha') = \sum_{Q \in S_{n-3}} F_P{}^Q(\alpha') A(1, Q, n{-}1, n) \,. \tag{1.11}$$

In this simplified form of the string amplitudes, the entire α'-dependence resides in the disk integrals $F_P{}^Q$ (indexed by permutations P, Q of legs $2, 3, \ldots, n{-}2$) which are defined in (6.51) and depend on external momenta. The SYM amplitudes $A(Q)$ in turn carry the complete polarization dependence of (1.11) for any combination of external bosons and fermions. Remarkably, the superspace structure of string tree amplitudes is captured by field-theory building blocks $A(Q)$ and separated from the string effects in the scalar disk integrals $F_P{}^Q$.

As detailed in sections 6.1 to 6.3, the derivation of (1.11) starts from the opening line (1.4) and relies on the local multiparticle superfields to perform the OPEs among the vertex operators. After integration-by-parts reduction of the disk integrals, the SYM amplitudes are identified through their superspace representation (1.9) in BCJ gauge.

The expression (1.11) for n-point superstring disk amplitudes turns out to line up with the Kawai–Lewellen–Tye (KLT) formula for supergravity tree amplitudes $M_n^{\rm grav}$ once the integrals $F_P{}^Q(\alpha')$ are rewritten in a Parke–Taylor basis of Z-integrals defined in (6.62):

$$\begin{aligned} M_n^{\rm grav} &= -\sum_{Q,R \in S_{n-3}} \tilde{A}(1, Q, n, n{-}1) S(Q|R)_1 A(1, R, n{-}1, n) \\ \leftrightarrow \quad \mathcal{A}(P) &= -\sum_{Q,R \in S_{n-3}} Z(P|1, Q, n, n{-}1) S(Q|R)_1 A(1, R, n{-}1, n) \,. \end{aligned} \tag{1.12}$$

The entries of the $(n{-}3)! \times (n{-}3)!$ KLT kernel $S(Q|R)_1$ are degree-$(n{-}3)$ polynomials in $k_i \cdot k_j$, see (4.160). Since the KLT formula for $M_n^{\rm grav}$ reflects the tree-level double copy of supergravity as a square of SYM, we interpret (1.12) as a field-theory double-copy construction of the open superstring from SYM and disk integrals $Z(P|Q)$.

As detailed in section 6.4.3, the disk integrals $Z(P|Q)$ at fixed color ordering P obey field-theory KK and BCJ relations between different Parke–Taylor integrands specified by Q. By these relations and their appearance in a field-theory KLT relation (1.12), the $Z(P|Q)$ are proposed to furnish (single-trace) tree-level amplitudes in a ultraviolet-completed theory of bi-colored scalars dubbed Z-theory. This is furthermore supported by the emergence of doubly-partial amplitudes of bi-adjoint scalars in the field-theory limit

$$\lim_{\alpha' \to 0} Z(P|Q) = m(P|Q) \,, \tag{1.13}$$

see section 6.4.4 and in particular (6.80) for the definition of $m(P|Q)$.

1.1.6. Color-kinematics duality and double copy

Another main result of section 6 is the manifestly local $(n-2)!$-term representation (6.64) of superstring disk amplitudes. By the field-theory limit (1.13) of the disk integrals therein, we obtain SYM amplitudes from a sum over permutations Q of legs $2,3,\ldots,n-1$,

$$A(P) = \sum_{Q \in S_{n-2}} m(P|1,Q,n) N_{1|Q|n} \,. \tag{1.14}$$

The kinematic factors $N_{1|\ldots|n}$ are trilinears in local multiparticle superfields A^P_α in BCJ gauge,

$$N_{1|P(n-1)Q|n} = (-1)^{|Q|+1} \langle V_{1P} V_{n\tilde{Q}} V_{n-1} \rangle \,, \qquad V_P = \lambda^\alpha A^P_\alpha \,. \tag{1.15}$$

The appearance of $N_{1|\ldots|n}$ in (1.14) identifies them as BCJ master numerators that manifest the color-kinematics duality of SYM at all multiplicities for any combination of external bosons and fermions. More precisely, by the discussion in section 7.1, the kinematic numerators in (1.15) are associated with the half-ladder diagrams in figure 8 and generate all other cubic-diagram numerators by kinematic Jacobi identities.

In the same way as the open superstring manifests the color-kinematics duality of n-point SYM tree amplitudes, section 7.2.3 reviews the derivation of the gravitational double copy in its cubic-diagram formulation from closed superstrings,

$$M^{\rm grav}_n = \sum_{P,Q \in S_{n-2}} \tilde{N}_{1|P|n} m(1,P,n|1,Q,n) N_{1|Q|n} \,, \tag{1.16}$$

which is equivalent to the KLT formula for supergravity tree amplitudes $M^{\rm grav}_n$ in (1.12). In both (1.14) and (1.16), the key to realize the BCJ duality and double copy with manifest locality is the simplification of the correlator in (1.4) to the $(n-2)!$-term combination (7.64) of local multiparticle superfields and Parke–Taylor factors. Moreover, the construction relies on doubly-partial amplitudes $m(P|Q)$ from the field-theory limit (1.13) of disk integrals and closely related sphere integrals (7.61).

Similarly, we shall construct explicit BCJ numerators for the non-linear sigma model (NLSM) of Goldstone bosons in section 7.4 reflected in the amplitude representation

$$A_{\rm NLSM}(P) = i^{n-2} \sum_{Q \in S_{n-2}} m(P|1,Q,n) S(Q|Q)_1 \tag{1.17}$$

analogous to (1.14), with $S(Q|Q)_1$ the diagonal entries of the KLT kernel in (1.12). Section 7.5 in turn is dedicated to double-copy representations of Einstein–Yang–Mills tree amplitudes similar to (1.16) that are derived from the heterotic version of the pure spinor superstring.

1.1.7. α'-expansions of open- and closed-superstring tree amplitudes

The low-energy expansion of the n-point disk integrals $F_P{}^Q$ and $Z(P|Q)$ in the open-string amplitudes (1.11) and (1.12) yields infinite series in dimensionless Mandelstam invariants $\alpha' k_i \cdot k_j$ with multiple zeta values (MZVs) in their coefficients,

$$\zeta_{n_1,n_2,\ldots,n_r} = \sum_{0<k_1<k_2<\ldots<k_r} k_1^{-n_1} k_2^{-n_2} \ldots k_r^{-n_r} \,, \qquad n_1,n_2,\ldots,n_r \in \mathbb{N} \,, \qquad n_r \geq 2 \,. \tag{1.18}$$

After a brief review of mathematical background in section 8.2, the f-alphabet description of (motivic) MZVs is shown to determine the entire α'-expansion from the coefficients of the Riemann zeta values ζ_w (i.e. (1.18) at depth $r=1$), see (8.37). This reflects a kind of closure of disk integrals under the motivic coaction Δ of MZVs which has also been observed in other areas of high-energy physics and can be expressed in terms of another KLT formula (8.40) for $\Delta Z(P|Q)$.

On top of these structural results, we review two recursive methods to explicitly determine the polynomials in $k_i \cdot k_j$ within the α'-expansion of n-point disk integrals. In section 8.5, matrix representations of the Drinfeld associator relate the $(n-1)$-point and n-point versions of the $F_P{}^Q$ basis in (1.11). Section 8.6 is dedicated to a Berends–Giele recursion for the $Z(P|Q)$ integrals in (1.12) which is generated by a non-linear field equation of bi-colored scalars in α'-expanded form and supports the interpretation of $Z(P|Q)$ as Z-theory amplitudes.

The α'-expansion of closed-string tree amplitudes only features the subclass of MZVs obtained from the so-called "single-valued" map sv. Even though the notion of single-valued MZVs is only well-defined in a motivic setting, we informally write the main result of section 8.7.1 as

$$\mathcal{M}_n^{\rm closed}(\alpha') = -\sum_{P,Q,R\in S_{n-3}} \tilde{A}(1,P,n,n-1)S(P|Q)_1 {\rm sv}\, F_Q{}^R(\alpha')A(1,R,n-1,n)\,. \tag{1.19}$$

The single-valued map acts on the MZVs order by order in α', for instance ${\rm sv}\,\zeta_{2k}=0$ and ${\rm sv}\,\zeta_{2k+1}=2\zeta_{2k+1}$ at depth one, but leaves the external polarizations and momenta inert. Similar to the expression (1.11) for superstring disk amplitudes, the α'-dependence of (1.19) is isolated in a scalar quantity ${\rm sv}\,F_Q{}^R(\alpha')$ while all the superfield-polarizations are carried by SYM amplitudes $\tilde{A}(P)$ and $A(R)$. With the SYM amplitudes in (1.9) one can access all multiplet components of type IIA and IIB amplitudes via (1.19). Moreover, the low-energy expansion of (1.19) can be made fully explicit through the single-valued map of disk integrals $F_Q{}^R$ within the reach of the expansion methods in sections 8.5 and 8.6.

1.1.8. A web of field-theory double copies for string amplitudes

There is a steadily growing web of double-copy relations among field theories of different spins [30, 31, 32] which can be formulated in terms of the KLT formula (1.12) with kernel $S(P|Q)_1$. In case of supergravity and Einstein–Yang–Mills, such double-copy relations can be derived from the string-theory KLT formula reviewed in section 7.2.1. It expresses closed-string tree-level amplitudes via bilinears in color-ordered open-string tree amplitudes with an α'-dependent kernel $\mathcal{S}_{\alpha'}(P|Q)_1$ that depends trigonometrically on the external momenta.

The representations in (1.12) and (1.19) for open- and closed-string tree-level amplitudes in turn involve the field-theory KLT kernel $S(P|Q)_1=\lim_{\alpha'\to 0}\mathcal{S}_{\alpha'}(P|Q)_1$ and are still exact in α'. The emergence of a field-theory double-copy in a string-theory context can be traced back to the KLT form (6.73) of the n-point correlation function of vertex operators in the pure spinor formalism. This correlator including the field-theory KLT kernel therein also enters the tree amplitudes of type II and heterotic superstrings upon pairing with right movers and for instance explains the factor of $S(P|Q)_1$ in (1.19). The latter can in fact be written as

$$\mathcal{M}_n^{\rm closed} = -\sum_{P,Q\in S_{n-3}} \tilde{A}(1,P,n,n-1)S(P|Q)_1 {\rm sv}\,\mathcal{A}(1,Q,n-1,n)\,. \tag{1.20}$$

identifying type II superstrings as a field-theory double copy of SYM with the single-valued open superstring. Field-theory KLT formulae require BCJ relations of both double-copy constituents as a consistency condition which is met for the ${\rm sv}\,\mathcal{A}(Q)$ in (1.20) to all orders in α', see the discussion in section 8.4.1. As a commonality of (1.20) with the KLT form (1.12) of open-superstring amplitudes, SYM building blocks are double-copied through the field-theory KLT kernel with one string-theoretic object – disk integrals or single-valued open-superstring amplitudes.

Also for heterotic strings reviewed in section 7.5, n-point tree amplitudes obey a field-theory KLT formula

$$\mathcal{M}_n^{\rm het} = -\sum_{P,Q\in S_{n-3}} \tilde{A}_{(DF)^2+{\rm YM}+\phi^3}(1,P,n,n-1)S(P|Q)_1 {\rm sv}\,\mathcal{A}(1,Q,n-1,n) \tag{1.21}$$

with one quantum-field-theory component $\tilde{A}_{(DF)^2+{\rm YM}+\phi^3}$ and again ${\rm sv}\,\mathcal{A}(Q)$ as a string-theoretic component. However, the Lagrangian and tree amplitudes of the $(DF)^2+{\rm YM}+\phi^3$ field theory are more complicated than those of SYM by the lack of supersymmetry and the two types of massive internal states, see section 7.5.3. On the basis of (1.21), the tree amplitudes of heterotic strings with external gauge and gravity supermultiplets reveal a field-theory double copy of $(DF)^2+{\rm YM}+\phi^3$ with single-valued open superstrings.

Together with similar field-theory KLT formulae (8.106) for open- and closed-string amplitudes of the bosonic theories, we arrive at the web of double-copy relations summarized in table 3: tree amplitudes in various perturbative string theories are intertwined with field-theory amplitudes and string-theoretic building blocks that share the KK and BCJ relations of gauge theories.

1.1.9. Example of a possible shortcut

The above summary of selected main results in this review together with the pointers to equations and sections may offer shortcuts to extract the key information on specific topics of interest. For instance, the expression (1.9) for SYM tree amplitudes can already be defined through the Berends–Giele currents M_P in Lorenz gauge. In this case, the consistency conditions and component evaluations can already be understood from the non-linear theory of ten-dimensional SYM using only non-local superfields, i.e. independently of string-theory methods and the local multiparticle superfields in BCJ gauge in section 4.1.

However, important aspects such that the BCJ amplitude relations or the kinematic Jacobi identities among SYM numerators require the notions of local multiparticle superfields and BCJ gauge as well as the associated formalism. Therefore the theory of multiparticle superfields is given an exhaustive discussion in section 4 before their applications in scattering amplitudes. Here and in other contexts, the reader should be aware that the years of development led to a healthy growth in the amount of connections between a variety of subjects which caused the review to grow beyond the page limits envisioned in earlier stages.

1.2. Related topics beyond the scope of this review

There is a variety of related topics that fruitfully resonate with superstring tree amplitudes but could not be covered in this review. As a small sample, we shall make a few comments on the Cachazo–He–Yuan (CHY) formalism, string field theory, the hybrid formalism and strings in AdS spacetimes here, and a more detailed account on loop-level string amplitudes covering references up to fall 2022 can be found in section 9.

1.2.1. The CHY formalism

An alternative worldsheet approach to double-copy representations of field-theory amplitudes is offered by the CHY formalism [33, 34, 35]. It may be viewed as an uplift of the Witten-RSV [36, 37] and Cachazo-Skinner [38] formulae to generic spacetime dimensions $D \neq 4$ and is underpinned by ambitwistor string theories in RNS [39, 40] and pure spinor [41, 42] formulations. The reader is referred to the review [43] and the white paper [32] for the wealth of developments in the CHY formalism and its interplay with double copy and superstring amplitudes.

CHY formulae directly compute field-theory amplitudes from moduli-space integrals for punctured Riemann surfaces similar to those in superstring amplitudes. These CHY integrals are completely localized via so-called scattering equations and in case of Parke–Taylor integrands seen in the main formulae for superstring tree amplitudes such as (1.12) coincide with the field-theory limits of disk and sphere integrals, see for instance (1.13). In fact, our main result in (6.73) or (7.64) for the n-point correlation function of massless vertex operators in the pure spinor superstring can be readily exported to the pure spinor version of the ambitwistor string [44].

1.2.2. String field theory

Perturbative string theories admit an alternative formulation in terms of string field theory where scattering amplitudes including their exact α'-dependence are computed from Feynman-type rules for a string field. The wavefunction of the string field depends on the zero and non-zero modes of the worldsheet variables and may guide an extension of the multiparticle formalism for massless vertex operators to the entire string spectrum. Recent lecture notes on string field theory can for instance be found in [45, 46, 47].

On the one hand, string field theory may face more technical complications in a detailed evaluation of string amplitudes than the worldsheet techniques described in this work. On the other hand, string field theory is widely considered more promising to describe non-perturbative features of superstring theory including duality symmetries or background independence. In particular, string field theory turned out to be a successful approach to tachyon condensation [48, 49, 50] or mass renormalization [51, 52, 53] and is conjectured to provide an understanding of the AdS/CFT correspondence [54, 55, 56, 57].

1.2.3. The hybrid formalism and strings in AdS spacetimes

As an alternative to the RNS, GS and pure spinor descriptions of the superstring, the so-called hybrid formalism admits manifestly $SO(1,3)$- or $SO(1,5)$-super-Poincaré invariant quantization. The hybrid formalism was constructed in the 90's from a series of field redefinitions in the RNS formalism to GS-like variables which manifest half- or quarter-maximal spacetime supersymmetry [58, 59, 60, 61, 62].

Apart from manifestly supersymmetric amplitude computations in flat spacetime [63], a major appeal of the hybrid formalism is its aptitude for the description of superstrings in $AdS_3 \times S^3$ backgrounds [61] (also see [64] for $AdS_2 \times S^2$). The intricate physical-state conditions for the $AdS_3 \times S^3$ superstring are for instance discussed in [65, 66, 67, 68, 69], also see [70] for a three-graviton amplitude. The hybrid formulation in [61] became a driving force for recent progress on type II superstrings in $AdS_3 \times S^3 \times T^4$ spacetime with NS flux and clarified the gauge-theory dual under the AdS/CFT correspondence [71, 72, 68, 73].

For superstrings in $AdS_5 \times S^5$ with finite radius, the RNS formulation faces difficulties in incorporating Ramond flux backgrounds. The pure spinor formalism in turn preserves the full $PSU(2,2|4)$ symmetry of the coset description of $AdS_5 \times S^5$ upon quantization [74], though a larger amount of computations has been performed in the GS formalism [75]. The reader is referred to the comprehensive review [76] and the white paper [77] for further references on both the pure-spinor and GS approach to superstrings in $AdS_5 \times S^5$; also see the white paper [78] for progress on relating string perturbation theory with conformal correlators through the AdS/CFT correspondence.

1.3. Conventions and notation

Ten-dimensional superspace. The ten-dimensional superspace coordinates are denoted $\{X^m, \theta^\alpha\}$, where $m = 0, \ldots, 9$ are the vector indices and $\alpha = 1, \ldots, 16$ denote the spinor indices of the Lorentz group $SO(10)$. The spinor representation is based on the 16×16 Pauli matrices $\gamma^m_{\alpha\beta} = \gamma^m_{\beta\alpha}$ satisfying the Clifford algebra $\gamma^{(m}_{\alpha\beta}\gamma^{n)\beta\gamma} = 2\delta^{mn}\delta^\gamma_\alpha$. In this review, unless stated otherwise, the (anti)symmetrization of k indices does not include a factor of $\frac{1}{k!}$. For more details on gamma matrices, see Appendix A.

Multiparticle index notation. In this review we will use a notation based on words to label multiparticle states. More precisely, let $\mathbb{N} = \{1, 2, 3, \ldots\}$ be the alphabet of external-particle labels. We will consider the vector space generated by linear combinations of *words* $P = p_1 p_2 \ldots$ with *letters* p_i from the alphabet $\mathbb{N}$. Capital letters from the Latin alphabet are used to represent words (e.g. $P = 1423$) while their composing letters are represented in lower case (e.g. $p = 3$). The length of a word $P = p_1 p_2 \ldots p_k$ is denoted by $|P| = k$ and it is given by the total number of letters contained in it. The empty word is denoted by $P = \emptyset$ has length $|P| = 0$. The reversal of a word $P = p_1 p_2 \ldots p_k$ is $\tilde{P} = p_k \ldots p_2 p_1$. The deconcatenation of a word P into two words X and Y is denoted by $\sum_{P=XY}$, and it represents all the possible ways to concatenate two words X and Y (including the empty word) such that $XY = P$. This operation will be often used when the words are labels of other objects (usually superfields such as M_P), for instance

$$\sum_{XY=123} M_X M_Y = M_\emptyset M_{123} + M_1 M_{23} + M_{12} M_3 + M_{123} M_\emptyset. \tag{1.22}$$

More definitions can be found in the Appendix C.

The multiparticle momentum k^m_P for a word P with letters i from massless particles $(k_i \cdot k_i) = 0$ and its associated Mandelstam invariant are given by

$$k^m_P := k^m_{p_1} + \cdots + k^m_{p_{|P|}}\,, \qquad s_P := \frac{1}{2}(k_P \cdot k_P)\,. \tag{1.23}$$

For example $k^m_{123} := k^m_1 + k^m_2 + k^m_3$ and $s_{123} = s_{12} + s_{13} + s_{23}$.

2. Super Yang–Mills in ten dimensions

Super Yang–Mills (SYM) theory in ten dimensions is the simplest among D-dimensional SYM theories; its spectrum contains just the gluon and gluino, related by sixteen supercharges [79] that form a Majorana-Weyl spinor of $SO(1,9)$. It is perhaps not a coincidence that it is also the theory relevant to the low-energy limit of superstring theory [80]. Its super-Poincaré covariant formulation [81, 82] is, in particular, one of the pillars supporting the pure spinor description of massless states of the open superstring. And indeed the SYM superfields of [81, 82] and their multiparticle generalization [83, 84, 85] reviewed in section 4 played an essential role in the calculation of the general n-point superstring disk amplitude. It is therefore beneficial to start this review by giving a detailed account of this beautiful field theory.

On top of the original superfields of [81, 82] we will define additional superfields of arbitrary mass dimension and study their non-linear equations of motion. This framework simplifies the θ-expansions of multiparticle superfields as detailed in Appendix F and the expressions of kinematic factors in higher-loop scattering amplitudes, including the D^6R^4 interaction in the superstring three-loop amplitude [86] as discussed in [87].

It is also well-known that the dimensional reduction of the simple ten-dimensional SYM theory gives rise to various maximally supersymmetric Yang–Mills theories in lower dimensions, including the celebrated $\mathcal{N}=4$ theory in $D=4$ [79]. Therefore a better understanding of the $D=10$ theory propagates to a variety of applications[3] to any lower dimension.

2.1. Ten-dimensional SYM

To describe the gluon and gluino states of ten-dimensional SYM, one introduces Lie algebra-valued superfield connections $\mathbb{A}_\alpha = \mathbb{A}_\alpha(X,\theta)$ and $\mathbb{A}_m = \mathbb{A}_m(X,\theta)$, the supercovariant derivatives,

$$\nabla_\alpha := D_\alpha - \mathbb{A}_\alpha\,, \qquad \nabla_m := \partial_m - \mathbb{A}_m\,, \tag{2.1}$$

and imposes the constraint [81, 82]

$$\big\{\nabla_\alpha, \nabla_\beta\big\} = \gamma^m_{\alpha\beta}\nabla_m. \tag{2.2}$$

Note that $\partial_m = \frac{\partial}{\partial X^m}$, and the superspace derivative

$$D_\alpha := \frac{\partial}{\partial\theta^\alpha} + \frac{1}{2}(\gamma^m\theta)_\alpha\partial_m \tag{2.3}$$

satisfies $\{D_\alpha, D_\beta\} = \gamma^m_{\alpha\beta}\partial_m$, see Appendix A.1 for our conventions for the 16×16 Pauli matrices $\gamma^m_{\alpha\beta}$.

Non-linear equations of motion. The constraint (2.2) and the associated Bianchi identities imply the following non-linear equations of motion [81, 82]

$$\begin{aligned} \big\{\nabla_\alpha, \nabla_\beta\big\} &= \gamma^m_{\alpha\beta}\nabla_m\,, & \big[\nabla_\alpha, \nabla_m\big] &= -(\gamma_m\mathbb{W})_\alpha\,, \\ \big\{\nabla_\alpha, \mathbb{W}^\beta\big\} &= \frac{1}{4}(\gamma^{mn})_\alpha{}^\beta\mathbb{F}_{mn}\,, & \big[\nabla_\alpha, \mathbb{F}^{mn}\big] &= (\mathbb{W}^{[m}\gamma^{n]})_\alpha\,, \end{aligned} \tag{2.4}$$

where

$$\mathbb{F}_{mn} := -\big[\nabla_m, \nabla_n\big]\,, \qquad \mathbb{W}^\alpha_m := \big[\nabla_m, \mathbb{W}^\alpha\big]\,, \tag{2.5}$$

and we recall that our conventions for (anti-)symmetrizing k indices do not contain factors of $\frac{1}{k!}$, e.g. $T^{[\mu\nu]} = T^{\mu\nu} - T^{\nu\mu}$. The superfields $\mathbb{F}^{mn}$ and $\mathbb{W}^\alpha$ are the field strengths of the gluon and gluino, respectively.

Gauge invariance. The equations (2.4) are invariant under the infinitesimal gauge transformations of the superfield connections under a Lie algebra-valued gauge parameter superfield $\Omega = \Omega(X,\theta)$

$$\delta_\Omega\mathbb{A}_\alpha = \big[\nabla_\alpha, \Omega\big]\,, \qquad \delta_\Omega\mathbb{A}_m = \big[\nabla_m, \Omega\big] \tag{2.6}$$

which in turn induce the gauge transformations of their field-strengths

$$\delta_\Omega\mathbb{W}^\alpha = \big[\Omega, \mathbb{W}^\alpha\big]\,, \qquad \delta_\Omega\mathbb{F}^{mn} = \big[\Omega, \mathbb{F}^{mn}\big]\,, \qquad \delta_\Omega\mathbb{W}^\alpha_m = \big[\Omega, \mathbb{W}^\alpha_m\big]\,. \tag{2.7}$$

[3] The dimensional reduction of the multiparticle superfields appears to be unexplored territory so far.

Lemma 1. *The equations (2.4) imply the (massless) Dirac and Yang–Mills equations,*

$$\gamma^m_{\alpha\beta}\big[\nabla_m, \mathbb{W}^\beta\big] = 0\,, \qquad \big[\nabla_m, \mathbb{F}^{mn}\big] = \gamma^n_{\alpha\beta}\big\{\mathbb{W}^\alpha, \mathbb{W}^\beta\big\}\,. \tag{2.8}$$

Proof. To obtain the Dirac equation, we use the constraint equation (2.2) to get

$$\begin{aligned}
\gamma^m_{\alpha\beta}\big[\nabla_m, \mathbb{W}^\beta\big] &= \big[\{\nabla_\alpha, \nabla_\beta\}, \mathbb{W}^\beta\big] = -\big[\{\mathbb{W}^\beta, \nabla_\alpha\}, \nabla_\beta\big] - \big[\{\nabla_\beta, \mathbb{W}^\beta\}, \nabla_\alpha\big] \\
&= -\frac{1}{4}(\gamma^{mn})_\alpha{}^\beta\big[\mathbb{F}_{mn}, \nabla_\beta\big] = \frac{1}{4}(\gamma^{mn}\gamma^n\mathbb{W}^m)_\alpha - \frac{1}{4}(\gamma^{mn}\gamma^m\mathbb{W}^n)_\alpha \\
&= \frac{9}{2}\gamma^m_{\alpha\beta}\big[\nabla_m, \mathbb{W}^\beta\big]\,,
\end{aligned} \tag{2.9}$$

where we used $\gamma^{mn}\gamma^n = 9\gamma^m$, $\gamma^{mn}{}_\beta{}^\beta = 0$ and (2.5) to arrive at the last line, implying that $\gamma^m_{\alpha\beta}\big[\nabla_m, \mathbb{W}^\beta\big] = 0$. To obtain the Yang–Mills equation, one evaluates the anti-commutator of the Dirac equation with $\gamma^{\alpha\delta}_n\nabla_\delta$ and uses the Bianchi (or Jacobi) identity,

$$\begin{aligned}
0 &= \gamma^{\alpha\delta}_n\gamma^m_{\alpha\beta}\big\{\nabla_\delta, [\nabla_m, \mathbb{W}^\beta]\big\} = \gamma^{\alpha\delta}_n\gamma^m_{\alpha\beta}\big\{\mathbb{W}^\beta, [\nabla_\delta, \nabla_m]\big\} + \gamma^{\alpha\delta}_n\gamma^m_{\alpha\beta}\big[\nabla_m, \{\mathbb{W}^\beta, \nabla_\delta\}\big] \\
&= -\gamma^{\alpha\delta}_n\gamma^m_{\alpha\beta}(\gamma_m)_{\delta\sigma}\big\{\mathbb{W}^\beta, \mathbb{W}^\sigma\big\} + \frac{1}{4}\gamma^{\alpha\delta}_n\gamma^m_{\alpha\beta}(\gamma^{rs})_\delta{}^\beta\big[\nabla_m, \mathbb{F}_{rs}\big] \\
&= 8\gamma^n_{\beta\sigma}\big\{\mathbb{W}^\beta, \mathbb{W}^\sigma\big\} - 8\big[\nabla_m, \mathbb{F}^{mn}\big]\,,
\end{aligned} \tag{2.10}$$

where to arrive in the last line we used the Clifford algebra (A.28) and $\gamma^m\gamma_m = 10$ to obtain $-(\gamma^m\gamma^n\gamma_m)_{\beta\sigma} = 8\gamma^n_{\beta\sigma}$ and used the trace relation (A.24) to get $\frac{1}{4}\mathrm{Tr}\big(\gamma_m\gamma_n\gamma^{rs}\big) = 4(\delta^r_n\delta^s_m - \delta^r_m\delta^s_n)$. □

Non-linear equations of motion. The equations of motion (2.4) can also be rewritten as

$$\begin{aligned}
\{\nabla_\alpha, \mathbb{A}_\beta\} + \{\nabla_\beta, \mathbb{A}_\alpha\} &= \gamma^m_{\alpha\beta}\mathbb{A}_m - \{\mathbb{A}_\alpha, \mathbb{A}_\beta\}\,, &\quad [\nabla_\alpha, \mathbb{A}_m] &= [\partial_m, \mathbb{A}_\alpha] + (\gamma_m\mathbb{W})_\alpha\,, \\
\{\nabla_\alpha, \mathbb{W}^\beta\} &= \frac{1}{4}(\gamma^{mn})_\alpha{}^\beta\mathbb{F}_{mn}\,, & [\nabla_\alpha, \mathbb{F}^{mn}] &= (\mathbb{W}^{[m}\gamma^{n]})_\alpha\,.
\end{aligned} \tag{2.11}$$

After using the definitions (2.1) these become

$$\begin{aligned}
\{D_\alpha, \mathbb{A}_\beta\} + \{D_\beta, \mathbb{A}_\alpha\} &= \gamma^m_{\alpha\beta}\mathbb{A}_m + \{\mathbb{A}_\alpha, \mathbb{A}_\beta\}\,, &\quad [D_\alpha, \mathbb{A}_m] &= [\partial_m, \mathbb{A}_\alpha] + (\gamma_m\mathbb{W})_\alpha + [\mathbb{A}_\alpha, \mathbb{A}_m]\,, \\
\{D_\alpha, \mathbb{W}^\beta\} &= \frac{1}{4}(\gamma^{mn})_\alpha{}^\beta\mathbb{F}_{mn} + \{\mathbb{A}_\alpha, \mathbb{W}^\beta\}\,, & [D_\alpha, \mathbb{F}^{mn}] &= (\mathbb{W}^{[m}\gamma^{n]})_\alpha + [\mathbb{A}_\alpha, \mathbb{F}^{mn}]\,,
\end{aligned} \tag{2.12}$$

which will be used later in section 4.2.3 to obtain the Berends–Giele recursions for superfields from a perturbiner expansion. For later convenience, we use the collective notation $\mathbb{K}$ referring to any element of the set containing these superfields,

$$\mathbb{K} \in \{\mathbb{A}_\alpha, \mathbb{A}_m, \mathbb{W}^\alpha, \mathbb{F}^{mn}\}\,. \tag{2.13}$$

2.2. Linearized superfields

Scattering amplitudes deal with linearized perturbations, so we need the *linearized* description of ten-dimensional SYM. This is obtained by discarding the quadratic terms from the equations of motion (2.11) and yields

$$\begin{aligned}
D_\alpha A^i_\beta + D_\beta A^i_\alpha &= \gamma^m_{\alpha\beta}A^i_m\,, &\quad D_\alpha A^i_m &= (\gamma_m W_i)_\alpha + \partial_m A^i_\alpha\,, \\
D_\alpha W^\beta_i &= \frac{1}{4}(\gamma^{mn})_\alpha{}^\beta F^i_{mn}\,, & D_\alpha F^i_{mn} &= \partial_{[m}(\gamma_{n]}W_i)_\alpha\,.
\end{aligned} \tag{2.14}$$

In addition, the linearized version of the gauge transformations (2.6) are given by

$$\delta_\Omega A_\alpha = D_\alpha\Omega\,, \qquad \delta_\Omega A_m = \partial_m\Omega\,, \tag{2.15}$$

and they will play a role in the definition of massless vertices in the pure spinor formalism in section 3.4.

In the context of scattering amplitudes, the linearized superfields are labelled by natural numbers i. These numbers are the *single-particle labels* keeping track of the $i^{\rm th}$ external state taking part in the scattering process. In addition, the linearized equations (2.14) describe the motion of a single SYM particle with label i. More abstractly, i can be thought of being a *letter* from the alphabet of natural numbers. As we will discuss in section 4, the concept of labeling superfields with a single letter i has been generalized for multiparticle states labelled by *words* P, the *multiparticle superfields*. It will then be shown in section 5 that SYM scattering amplitudes involving multiple particles can be compactly written in terms of these multiparticle superfields. And in section 6 we will see how they are utilized in the computation of superstring amplitudes.

2.2.1. θ-expansions

The linearized version of the gauge transformations (2.6) can be used to attain Harnad–Shnider gauge $\theta^\alpha A^i_\alpha = 0$, where the θ dependence is known in terms of fermionic power-series expansions from [88, 89, 90]. After peeling off the dependence of linearized superfields on the bosonic coordinates X^m via plane waves[4] $e^{k_i\cdot X}$ with on-shell momentum $k_i^2 = 0$, the different orders in θ alternate between gluino wave functions χ_i^α and gluon polarization vectors e_i^m, or their associated linearized field strength

$$f_i^{mn} = k_i^m e_i^n - k_i^n e_i^m \,. \tag{2.16}$$

More precisely,

$$\begin{aligned}
A^i_\alpha(X,\theta) &= \Big\{ \frac{1}{2}(\theta\gamma_m)_\alpha e_i^m + \frac{1}{3}(\theta\gamma_m)_\alpha(\theta\gamma^m\chi_i) - \frac{1}{32}(\theta\gamma_m)^\alpha(\theta\gamma^{mnp}\theta)f^i_{np} \\
&\quad + \frac{1}{60}(\theta\gamma_m)_\alpha(\theta\gamma^{mnp}\theta)k_n(\chi^i\gamma_p\theta) + \frac{1}{1152}(\theta\gamma_m)_\alpha(\theta\gamma^{mnp}\theta)(\theta\gamma^{pqr}\theta)k_i^n f_i^{qr} + \ldots \Big\} e^{k_i\cdot X} \,, \\
A_i^m(X,\theta) &= \Big\{ e_i^m + (\theta\gamma^m\chi_i) - \frac{1}{8}(\theta\gamma^{mpq}\theta)f_i^{pq} + \frac{1}{12}(\theta\gamma^{mnp}\theta)k_i^n(\chi_i\gamma^p\theta) \\
&\quad + \frac{1}{192}(\theta\gamma^m{}_{nr}\theta)(\theta\gamma^r{}_{pq}\theta)k_i^n f_i^{pq} - \frac{1}{480}(\theta\gamma^m{}_{nr}\theta)(\theta\gamma^r{}_{pq}\theta)k_i^n k_i^p(\chi_i\gamma^q\theta) + \ldots \Big\} e^{k_i\cdot X} \,, \\
W_i^\alpha(X,\theta) &= \Big\{ \chi_i^\alpha + \frac{1}{4}(\theta\gamma^{mn})^\alpha f^i_{mn} - \frac{1}{4}(\theta\gamma_{mn})^\alpha k_i^m(\chi_i\gamma^n\theta) - \frac{1}{48}(\theta\gamma_m{}^q)^\alpha(\theta\gamma_{qnp}\theta)k_i^m f_i^{np} \\
&\quad + \frac{1}{96}(\theta\gamma_m{}^q)^\alpha(\theta\gamma_{qnp}\theta)k_i^m k_i^n(\chi_i\gamma^p\theta) - \frac{1}{1920}(\theta\gamma_m{}^r)^\alpha(\theta\gamma_{nr}{}^s\theta)(\theta\gamma_{spq}\theta)k_i^m k_i^n f_i^{pq} + \ldots \Big\} e^{k_i\cdot X} \,, \\
F_i^{mn}(X,\theta) &= \Big\{ f_i^{mn} - k_i^{[m}(\chi_i\gamma^{n]}\theta) + \frac{1}{8}(\theta\gamma_{pq}{}^{[m}\theta)k_i^{n]}f_i^{pq} - \frac{1}{12}(\theta\gamma_{pq}{}^{[m}\theta)k_i^{n]}k_i^p(\chi_i\gamma^q\theta) \\
&\quad - \frac{1}{192}(\theta\gamma_{ps}{}^{[m}\theta)k_i^{n]}k_i^p f_i^{qr}(\theta\gamma^s{}_{qr}\theta) + \frac{1}{480}(\theta\gamma^{[m}{}_{ps}\theta)k_i^{n]}k_i^p k_i^q(\chi_i\gamma^r\theta)(\theta\gamma^s{}_{qr}\theta) + \ldots \Big\} e^{k_i\cdot X} \,,
\end{aligned} \tag{2.17}$$

see (F.7) for the analogous θ-expansions of the non-linear fields $\mathbb{K}$ in (2.13). Terms in the ellipsis involve six or higher orders in θ which won't be needed for the purpose of this review but can be obtained in closed form via expressions such as [89]

$$A_i^m(X,\theta) = \Big\{ (\cosh\sqrt{\mathcal{O}})^m{}_q e_i^q + \Big(\frac{\sinh\sqrt{\mathcal{O}}}{\sqrt{\mathcal{O}}}\Big)^m{}_q(\theta\gamma^q\chi_i) \Big\} e^{k_i\cdot X} \,, \tag{2.18}$$

where

$$\mathcal{O}^m{}_q = \frac{1}{2}(\theta\gamma^m{}_{qn}\theta)k_i^n \,. \tag{2.19}$$

[4]We absorb factors of i into momentum factors k^m in order to attain plane-wave factors of $e^{k\cdot X}$ subject to the simple conversion $\partial_m \to k_m$ instead of the more conventional $e^{ik\cdot X}$ with $\partial_m \to ik_m$. The traditional conventions can be retrieved by replacing $k_m \to ik_m$ and $s_{ij} \to -s_{ij}$ for Mandelstam variables defined in (1.23).

2.3. Superfields of higher mass dimension

As the loop order of SYM amplitudes increases so does the mass dimension of the associated kinematic factors. In the pure spinor formalism the maximum mass dimension for a four-point amplitude using only the standard SYM superfields in (2.13) is k^2F^4 obtained from the pure spinor superspace expression $\langle(\lambda\gamma^{mnpqr}\lambda)(\lambda\gamma^s W)F_{mn}F_{pq}F_{rs}\rangle$ at genus two [91].

Therefore it would be convenient to define SYM superfields of higher mass dimension as compared to the standard ones in $\mathbb{K}$. The obvious candidates of the form $\partial_m\partial_n\ldots\mathbb{K}$ are inadequate because the ordinary derivatives ∂_m do not preserve gauge covariance at a non-linear level probed by higher-point amplitudes. So, instead, the connection ∇_m in (2.1) guides the subsequent definitions [87]

$$\begin{aligned}\mathbb{W}^{m_1\ldots m_k\alpha} &:= \big[\nabla^{m_1}, \mathbb{W}^{m_2\ldots m_k\alpha}\big]\,,\\ \mathbb{F}^{m_1\ldots m_k|pq} &:= \big[\nabla^{m_1}, \mathbb{F}^{m_2\ldots m_k|pq}\big]\,,\end{aligned}\tag{2.20}$$

where the vertical bar separates the antisymmetric pair of indices present in the recursion start $\mathbb{F}^{pq}$.

2.3.1. Equations of motion at higher mass dimension

Similarly as in the standard SYM superfields of [81, 82], the equations of motion for the superfields of higher mass dimension (2.20) follow from $\big[\nabla_\alpha, \nabla_m\big] = -(\gamma_m\mathbb{W})_\alpha$ and $\big[\nabla_m, \nabla_n\big] = -\mathbb{F}_{mn}$ together with Jacobi identities among iterated brackets. In general, one can prove by induction that

$$\begin{aligned}\big\{\nabla_\alpha, \mathbb{W}^{N\beta}\big\} &= \tfrac{1}{4}(\gamma_{pq})_\alpha{}^\beta\mathbb{F}^{N|pq} - \sum_{\substack{\delta(N)=R\otimes S\\ R\neq\emptyset}}\big\{(\mathbb{W}\gamma)^R_\alpha, \mathbb{W}^{S\beta}\big\}\,,\\ \big[\nabla_\alpha, \mathbb{F}^{N|pq}\big] &= (\mathbb{W}^{N[p}\gamma^{q]})_\alpha - \sum_{\substack{\delta(N)=R\otimes S\\ R\neq\emptyset}}\big[(\mathbb{W}\gamma)^R_\alpha, \mathbb{F}^{S|pq}\big]\,.\end{aligned}\tag{2.21}$$

The vector indices have been gathered to a multi-index $N := n_1n_2\ldots n_k$ with $(\mathbb{W}\gamma)^N := (\mathbb{W}^{n_1\ldots n_{k-1}}\gamma^{n_k})$ and $\delta(N)$ denotes the deshuffle map defined in (C.10). The simplest examples of (2.21) are given by

$$\begin{aligned}\big\{\nabla_\alpha, \mathbb{W}^{m\beta}\big\} &= \tfrac{1}{4}(\gamma_{pq})_\alpha{}^\beta\mathbb{F}^{m|pq} - \big\{(\mathbb{W}\gamma^m)_\alpha, \mathbb{W}^\beta\big\}\,,\\ \big[\nabla_\alpha, \mathbb{F}^{m|pq}\big] &= (\mathbb{W}^{m[p}\gamma^{q]})_\alpha - \big[(\mathbb{W}\gamma^m)_\alpha, \mathbb{F}^{pq}\big]\,,\\ \big\{\nabla_\alpha, \mathbb{W}^{mn\beta}\big\} &= \tfrac{1}{4}(\gamma_{pq})_\alpha{}^\beta\mathbb{F}^{mn|pq} - \big\{(\mathbb{W}^m\gamma^n)_\alpha, \mathbb{W}^\beta\big\} - \big\{(\mathbb{W}\gamma^m)_\alpha, \mathbb{W}^{n\beta}\big\} - \big\{(\mathbb{W}\gamma^n)_\alpha, \mathbb{W}^{m\beta}\big\}\,,\\ \big[\nabla_\alpha, \mathbb{F}^{mn|pq}\big] &= (\mathbb{W}^{mn[p}\gamma^{q]})_\alpha - \big[(\mathbb{W}^m\gamma^n)_\alpha, \mathbb{F}^{pq}\big] - \big[(\mathbb{W}\gamma^m)_\alpha, \mathbb{F}^{n|pq}\big] - \big[(\mathbb{W}\gamma^n)_\alpha, \mathbb{F}^{m|pq}\big]\,,\\ \big\{\nabla_\alpha, \mathbb{W}^{mnp\,\beta}\big\} &= \tfrac{1}{4}(\gamma_{rs})_\alpha{}^\beta\mathbb{F}^{mnp|rs} - \big\{(\mathbb{W}^{mn}\gamma^p)_\alpha, \mathbb{W}^\beta\big\} - \big\{(\mathbb{W}^m\gamma^n)_\alpha, \mathbb{W}^{p\,\beta}\big\} - \big\{(\mathbb{W}^m\gamma^p)_\alpha, \mathbb{W}^{n\,\beta}\big\}\\ &\quad - \big\{(\mathbb{W}^n\gamma^p)_\alpha, \mathbb{W}^{m\,\beta}\big\} - \big\{(\mathbb{W}\gamma^p)_\alpha, \mathbb{W}^{mn\,\beta}\big\} - \big\{(\mathbb{W}\gamma^n)_\alpha, \mathbb{W}^{mp\,\beta}\big\} - \big\{(\mathbb{W}\gamma^m)_\alpha, \mathbb{W}^{np\,\beta}\big\}\,,\end{aligned}\tag{2.22}$$

where we used $\delta(mnp) = mnp\otimes\emptyset + mn\otimes p + mp\otimes n + np\otimes m + p\otimes mn + n\otimes mp + m\otimes np + \emptyset\otimes mnp$. One can also show inductively that the Dirac- and Yang–Mills equations (2.8) generalize as follows at higher mass dimension:

$$[\nabla_m, (\gamma^m\mathbb{W}^N)_\alpha] = \sum_{\substack{\delta(N)=R\otimes S\\ R\neq\emptyset}}\big[\mathbb{F}^{Rm}, (\gamma_m\mathbb{W}^S)_\alpha\big]\,,\tag{2.23}$$

$$[\nabla_m, \mathbb{F}^{N|pm}] = \delta_{mn}\sum_{\substack{\delta(N)=R\otimes S\\ R\neq\emptyset}}\big[\mathbb{F}^{Rm}, \mathbb{F}^{S|pn}\big] - \sum_{\delta(N)=R\otimes S}\big\{\mathbb{W}^{R\alpha}, (\gamma^p\mathbb{W}^S)_\alpha\big\}\,,\tag{2.24}$$

where $\mathbb{F}^{Rr}$ for non-empty $R := Qq$ is defined as $\mathbb{F}^{Qqr} := \mathbb{F}^{Q|qr}$. For example,

$$\begin{aligned}[\nabla_m, (\gamma^m\mathbb{W}^n)_\alpha] &= \big[\mathbb{F}^{nr}, (\gamma_r\mathbb{W})_\alpha\big]\,,\\ [\nabla_m, (\gamma^m\mathbb{W}^{np})_\alpha] &= \big[\mathbb{F}^{n|pr}, (\gamma_r\mathbb{W})_\alpha\big] + \big[\mathbb{F}^{nr}, (\gamma_r\mathbb{W}^p)_\alpha\big] + \big[\mathbb{F}^{pr}, (\gamma_r\mathbb{W}^n)_\alpha\big]\,,\end{aligned}\tag{2.25}$$

$$[\nabla_m, \mathbb{F}^{n|pm}] = [\mathbb{F}^{nm}, \mathbb{F}^p{}_m] - \{\mathbb{W}^{n\alpha}, (\gamma^p \mathbb{W})_\alpha\} - \{\mathbb{W}^\alpha, (\gamma^p \mathbb{W}^n)_\alpha\}\,,$$

where we used the deshuffle map $\delta(np) = np \otimes \emptyset + n \otimes p + p \otimes n + \emptyset \otimes np$.

Note that the linearized versions of higher-mass dimension superfields are simply the outer products of derivatives

$$W_i^{m_1 \ldots m_k \alpha} = \partial^{m_1} \ldots \partial^{m_k} W_i^\alpha\,, \qquad F_i^{m_1 \ldots m_k|pq} = \partial^{m_1} \ldots \partial^{m_k} F_i^{pq}\,, \tag{2.26}$$

where i denotes a single-particle label. In this case, the equations of motion (2.23) and (2.24) translate into

$$\partial_m (\gamma^m W_i^N)_\alpha = 0\,, \qquad \partial_m F_i^{N|mp} = 0\,. \tag{2.27}$$

In case of an empty multi-index $N \to \emptyset$, this includes the linearized Dirac and Yang–Mills equations $\partial_m (\gamma^m W_i)_\alpha = 0$ and $\partial_m F_i^{mp} = 0$.

The higher-mass-dimension superfields obey further relations which can be derived from Jacobi identities of nested (anti)commutators. For example, (2.5) determines their antisymmetrized components

$$\begin{aligned} \mathbb{W}^{[n_1 n_2] n_3 \ldots n_k \beta} &= \left[\mathbb{W}^{n_3 \ldots n_k \beta}, \mathbb{F}^{n_1 n_2}\right]\,, \\ \mathbb{F}^{[n_1 n_2] n_3 \ldots n_k|pq} &= \left[\mathbb{F}^{n_3 \ldots n_k|pq}, \mathbb{F}^{n_1 n_2}\right]\,. \end{aligned} \tag{2.28}$$

Similarly, more antisymmetrized indices give rise to nested commutators, for instance

$$\begin{aligned} \mathbb{W}^{[mn]\beta} &= [\mathbb{W}^\beta, \mathbb{F}^{mn}]\,, \\ \mathbb{W}^{[mnp]\beta} &= [\mathbb{W}^{m\beta}, \mathbb{F}^{np}] + [\mathbb{W}^{n\beta}, \mathbb{F}^{pm}] + [\mathbb{W}^{p\beta}, \mathbb{F}^{mn}]\,, \\ \mathbb{W}^{[mnpq]\beta} &= [[\mathbb{W}^\beta, \mathbb{F}^{mn}], \mathbb{F}^{pq}] - [[\mathbb{W}^\beta, \mathbb{F}^{mp}], \mathbb{F}^{nq}] + [[\mathbb{W}^\beta, \mathbb{F}^{mq}], \mathbb{F}^{np}] \\ &\quad + [[\mathbb{W}^\beta, \mathbb{F}^{np}], \mathbb{F}^{mq}] - [[\mathbb{W}^\beta, \mathbb{F}^{nq}], \mathbb{F}^{mp}] + [[\mathbb{W}^\beta, \mathbb{F}^{pq}], \mathbb{F}^{mn}]\,, \end{aligned} \tag{2.29}$$

with similar expressions at higher multiplicities.

Moreover, the definitions (2.20) via iterated commutators imply the generalized Jacobi identities of section 4.1.5 on the set of vector indices, of which first instances are

$$\mathbb{F}^{[m|np]} = 0\,, \quad \mathbb{F}^{[mn]|pq} + \mathbb{F}^{[pq]|mn} = 0\,. \tag{2.30}$$

3. Pure spinor formalism and disk amplitudes

The discovery of the pure spinor formalism by Berkovits in [1] gave to the world an efficient tool to compute superstring scattering amplitudes in a manifestly supersymmetric manner. It combined numerous convenient aspects of the Ramond–Neveu–Schwarz (RNS) [4, 5, 6, 7] and Green–Schwarz (GS) formulations [8, 9] in a way that allowed computations of various amplitudes previously out of reach.

In this section we will review the basic aspects of the formalism with a view towards the prescription to compute disk amplitudes in the superstring; multi-loop aspects will not be covered, but a path through the recent literature can be found in section 9. The presentation will follow the ICTP lectures by Berkovits [92] as well as a combination of the PhD theses of the present authors [93, 94].

We will now present some of the motivations that led to the development of the pure spinor formalism.

3.1. Difficulties with the covariant quantization of the Green-Schwarz string

Type I superstrings [95], type II superstrings [96] and heterotic strings [97] are supersymmetric in ten-dimensional space-time and therefore it is natural to seek a manifestly 10d supersymmetric description of their worldsheet action. This is traditionally achieved with the GS formalism [8, 9] but unfortunately the classical action cannot be quantized while maintaining Lorentz covariance.

The GS action for heterotic superstrings (or a chiral half of type II superstrings) in conformal gauge is given by [8]

$$S_{\mathrm{GS}} = \frac{1}{\pi} \int d^2 z \left[\frac{1}{2} \Pi^m \overline{\Pi}_m + \frac{1}{4} \Pi_m (\theta \gamma^m \overline{\partial} \theta) - \frac{1}{4} \overline{\Pi}_m (\theta \gamma^m \partial \theta)\right] \tag{3.1}$$

$$= \frac{1}{\pi} \int d^2 z \left[\frac{1}{2} \partial X^m \overline{\partial} X_m + \frac{1}{2} \partial X_m (\theta \gamma^m \overline{\partial} \theta) + \frac{1}{8} (\theta \gamma^m \partial \theta)(\theta \gamma_m \overline{\partial} \theta) \right] ,$$

where we employ supersymmetric momenta

$$\Pi^m = \partial X^m + \frac{1}{2}(\theta\gamma^m \partial\theta) \, , \qquad \overline{\Pi}^m = \overline{\partial} X^m + \frac{1}{2}(\theta\gamma^m \overline{\partial}\theta) \, . \tag{3.2}$$

The dependence of X^m, θ on the worldsheet coordinates $z, \overline{z}$ as well as the action of the gauge sector of the heterotic string is suppressed. Throughout this review, the integration measure is $d^2 z = d\,\mathrm{Re}\, z \wedge d\,\mathrm{Im}\, z = \frac{i}{2} dz \wedge d\overline{z}$, and derivatives are denoted by the shorthands $\partial = \partial_z$ and $\overline{\partial} = \partial_{\overline{z}}$. Holomorphic and antiholomorphic derivatives are related to those w.r.t. worldsheet coordinates $\sigma^0 = \frac{1}{2}(z+\overline{z})$ and $\sigma^1 = \frac{1}{2}(z-\overline{z})$ via $\partial_0 = \partial + \overline{\partial}$ and $\partial_1 = \partial - \overline{\partial}$. Following the standard closed-string conventions, we are setting $\alpha' = 2$ in sections 3 to 5 (but will reinstate it in sections 6 to 8).[5]

Covariant quantization of (3.1) is hindered by a technical challenge: the conjugate momentum to θ^α

$$p_\alpha = 2\pi \frac{\delta S_{\mathrm{GS}}}{\delta(\partial_0 \theta^\alpha)} = \frac{1}{2} \Big(\Pi^m - \frac{1}{4} (\theta \gamma^m \partial_1 \theta) \Big) (\gamma_m \theta)_\alpha \tag{3.3}$$

depends on θ^α itself, so it gives rise to the GS constraint $d_\alpha = 0$ with

$$d_\alpha = p_\alpha - \frac{1}{2} \Big(\Pi^m - \frac{1}{4} (\theta \gamma^m \partial_1 \theta) \Big) (\gamma_m \theta)_\alpha \, . \tag{3.4}$$

The variable d_α associated with the GS constraint satisfies the Poisson brackets

$$\{ d_\alpha, d_\beta \} = i \gamma^m_{\alpha\beta} \Pi_m \, . \tag{3.5}$$

Due to the Virasoro constraint $\Pi_m \Pi^m = 0$, the relation (3.5) mixes first- and second-class types of constraints in a way that is difficult to disentangle covariantly[6]. The standard way to deal with this situation is to go to the light-cone gauge [95, 99, 100, 96], where the two types of constraints can be treated separately and quantization can be achieved. However, one obviously loses manifest Lorentz covariance in the process. These difficulties are universal to heterotic and type II string theories in their GS formulations.

3.2. Siegel's reformulation of the Green–Schwarz formulation

In 1986 Siegel [22] proposed a new approach to deal with the covariant quantization of the GS formalism. His idea was to treat the conjugate momenta for θ^α as an independent variable, proposing the following action for the left-moving variables

$$S_{\mathrm{Siegel}} = \frac{1}{\pi} \int d^2 z \left[\frac{1}{2} \partial X^m \overline{\partial} X_m + p_\alpha \overline{\partial} \theta^\alpha \right] \tag{3.6}$$

in which the variable d_α

$$d_\alpha = p_\alpha - \frac{1}{2} \Big(\partial X^m + \frac{1}{4} (\theta \gamma^m \partial \theta) \Big) (\gamma_m \theta)_\alpha \tag{3.7}$$

was assumed to be independent and not a constraint (the difference between the expressions (3.4) and (3.7) for d_α is proportional to $\overline{\partial}\theta^\alpha$ and vanishes by the equations of motion for p_α). In this way, the mixing (3.5) of first- and second-class constraints of the GS formulation is not an issue in Siegel's approach.

[5] The α'-dependence of the worldsheet CFT and the associated scattering amplitudes can be reinstated based on dimensional analysis. For instance, demanding worldsheet actions to be dimensionless and $X^m, \sqrt{\alpha'}$ to have dimensions of a length, we retrieve $S_{\mathrm{GS}} \to \frac{1}{\pi\alpha'} \int d^2 z \partial X^m \overline{\partial} X_m + \ldots$.

[6] Recall that first-class (second-class) constraints are defined by the vanishing (non-vanishing) of their Poisson bracket [98]. The constraint $\Pi^2 = 0$ then implies that one half of the Poisson brackets (3.5) vanishes. We are grateful to Max Guillen for discussions on this point.

Lorentz currents and energy-momentum tensor. The action (3.6) is easily checked to yield a Lorentz current of the spinor variables[7]

$$\Sigma^{mn} = -\frac{1}{2}(p\gamma^{mn}\theta) \tag{3.8}$$

and a holomorphic component $T := T(z)$ of the energy-momentum tensor

$$T = -\frac{1}{2}\partial X^m \partial X_m - p_\alpha \partial\theta^\alpha = -\frac{1}{2}\Pi^m \Pi_m - d_\alpha \partial\theta^\alpha \, . \tag{3.9}$$

The supersymmetric momentum $\Pi^m = \partial X^m + \frac{1}{2}(\theta\gamma^m\partial\theta)$ is defined as in section 3.1 though its right-moving counterpart $\overline{\Pi}^m$ relevant for type II superstrings departs from (3.2) and is defined with separate θ-variables. For example, under the Lorentz transformation with parameters ε_{mn},

$$\delta p_\alpha = \frac{1}{4}\varepsilon_{mn}(\gamma^{mn})_\alpha{}^\beta p_\beta \, , \qquad \delta\theta^\alpha = \frac{1}{4}\varepsilon_{mn}(\gamma^{mn})^\alpha{}_\beta\theta^\beta \, , \tag{3.10}$$

we define the variation of (3.6) to be $\delta S_{\text{Siegel}} = -\frac{1}{\pi}\int \frac{1}{2}\Sigma^{mn}\overline{\partial}\varepsilon_{mn}$. The calculation using Noether's method is straightforward

$$\begin{aligned}
\delta S_{\text{Siegel}} &= \frac{1}{\pi}\int d^2z\, \delta(p_\alpha\overline{\partial}\theta^\alpha) = \frac{1}{\pi}\int d^2z \left[\frac{1}{4}\varepsilon_{mn}(\gamma^{mn})_\alpha{}^\beta p_\beta\overline{\partial}\theta^\alpha + \frac{1}{4}p_\alpha\overline{\partial}(\varepsilon_{mn}(\gamma^{mn}\theta)^\alpha)\right] \\
&= \frac{1}{\pi}\int d^2z \left[\frac{1}{4}\overline{\partial}\varepsilon_{mn}p_\alpha(\gamma^{mn})^\alpha{}_\beta\theta^\beta\right] = -\frac{1}{\pi}\int d^2z \frac{1}{2}\Sigma^{mn}\overline{\partial}\varepsilon_{mn} \, ,
\end{aligned} \tag{3.11}$$

where we used the antisymmetry $(\gamma^{mn})_\alpha{}^\beta = -(\gamma^{mn})^\beta{}_\alpha$, see (A.16).

CFT. The action (3.6) defines a conformal field theory in which the holomorphic conformal weights of $\partial X^m, p_\alpha$ and θ^α are $h_{\partial X} = h_p = 1$ and $h_\theta = 0$, respectively. See [101] for an in-depth review of conformal field theory. The operator product expansions (OPEs) among the variables in S_{Siegel} follow from standard path-integral methods [22]

$$X^m(z,\overline{z})X^n(w,\overline{w}) \sim -\delta^{mn}\ln|z-w|^2 \, , \qquad p_\alpha(z)\theta^\beta(w) \sim \frac{\delta_\alpha^\beta}{z-w} \, , \tag{3.12}$$

$$d_\alpha(z)d_\beta(w) \sim -\frac{\gamma^m_{\alpha\beta}\Pi_m(w)}{z-w} \, , \qquad d_\alpha(z)\Pi^m(w) \sim \frac{\big(\gamma^m\partial\theta(w)\big)_\alpha}{z-w} \, , \tag{3.13}$$

where here and throughout this review, $\sim$ indicates that regular terms as $z \to w$ are dropped on the right-hand side.

Vertex operator. Siegel also proposed a supersymmetric integrated vertex operator for massless open-string states labeled by i as follows

$$U_i^{\text{Siegel}} = \int dz \left(\partial\theta^\alpha A^i_\alpha(X,\theta) + A^i_m(X,\theta)\Pi^m + d_\alpha W_i^\alpha(X,\theta)\right) , \tag{3.14}$$

where $\{A^i_\alpha, A^m_i, W^\alpha_i\}$ are the linearized SYM superfields reviewed in section 2.2.

3.2.1. Difficulties with Siegel's approach

There are three types of difficulties with Siegel's approach which will be addressed by the pure spinor formalism to be introduced in section 3.3 below.

[7] The double-colon notation for normal ordering of coincident operators, $:A(z)B(z):$, will be left implicit in this review.

Non-vanishing central charge. According to the bc-system calculations [4] with conformal weight $h_p = 1$, each spinor component of the fermionic pair $(p_\alpha, \theta^\alpha)$ in the energy-momentum tensor (3.9) contributes $-3(2h_p-1)^2+1 = -2$ to the central charge for a total of $16\times(-2) = -32$ while the X^m contribute $+10$ [12]. Therefore the central charge of the energy-momentum tensor (3.9) is $c_X + c_{p\theta} = 10-32 = -22$. This non-vanishing result for the central charge leads to an anomaly when quantizing the theory, raising a first major difficulty in Siegel's approach to the GS formalism.

Inequivalence of massless vertex operators. As emphasized in [1], the vertex operator (3.14) cannot reproduce the same results for amplitudes computed in the RNS formalism as it does not satisfy the same OPEs. More explicitly, after using the θ-expansions (2.17) of the linearized SYM superfields, the gluon vertex operator obtained from (3.14) is

$$U^{\rm Siegel}_{i,\rm gluon} = \int dz\,\Big(e_i^m \partial X_m - \frac{1}{4}(p\gamma^{mn}\theta)f^i_{mn} + \dots\Big)e^{k_i\cdot X} \tag{3.15}$$

up to terms of order θ^3 in the ellipsis. The vertex operator for a gluon with polarization vector e_i^m in the RNS formalism, on the other hand, is given by (see (7.3.25) in [10][8])

$$U^{\rm RNS}_{i,\rm gluon} = \int dz\,\Big(e_i^m \partial X_m + \frac{1}{2}\psi^m\psi^n f^i_{mn}\Big)e^{k_i\cdot X}\,, \tag{3.16}$$

where ψ^m are the RNS worldsheet spinors of conformal weight $h_\phi = \frac{1}{2}$, and $f^i_{mn} = k^i_m e^i_n - k^i_n e^i_m$ denotes the linearized field strength of the gluon.

Comparing (3.16) with (3.15) one notices that the operator multiplying $\frac{1}{2}f^i_{mn}$ is the Lorentz current for the fermionic variables in each formalism,

$$\Sigma^{mn}_{\rm RNS} = -\psi^m\psi^n, \qquad \Sigma^{mn}_{\rm Siegel} = -\frac{1}{2}(p\gamma^{mn}\theta)\,. \tag{3.17}$$

The difficulty arises because their OPEs are different. On the one hand, in the RNS formalism we get

$$\Sigma^{mn}_{\rm RNS}(z)\Sigma^{pq}_{\rm RNS}(w) \sim \frac{\delta^{p[m}\Sigma^{n]q}_{\rm RNS}(w) - \delta^{q[m}\Sigma^{n]p}_{\rm RNS}(w)}{z-w} + \frac{\delta^{m[q}\delta^{p]n}}{(z-w)^2}\,, \tag{3.18}$$

where the double-pole term has coefficient $+1$ which can be identified with the level of the Kac–Moody current algebra. On the other hand, using the OPE (3.12) we get

$$\begin{aligned}\Sigma^{mn}_{\rm Siegel}(z)\Sigma^{pq}_{\rm Siegel}(w) &\sim \frac{1}{4}\frac{p_\alpha(w)(\gamma^{mn}\gamma^{pq}-\gamma^{pq}\gamma^{mn})^\alpha{}_\beta\theta^\beta(w)}{z-w} + \frac{1}{4}\frac{{\rm Tr}(\gamma^{mn}\gamma^{pq})}{(z-w)^2}\\ &= \frac{\delta^{p[m}\Sigma^{n]q}(w) - \delta^{q[m}\Sigma^{n]p}(w)}{z-w} + 4\frac{\delta^{m[q}\delta^{p]n}}{(z-w)^2}\,,\end{aligned} \tag{3.19}$$

where in the second line we used $\gamma^{mn}\gamma^{pq} - \gamma^{pq}\gamma^{mn} = 2\delta^{np}\gamma^{mq} - 2\delta^{nq}\gamma^{mp} + 2\delta^{mq}\gamma^{np} - 2\delta^{mp}\gamma^{nq}$ following from (A.32) and ${\rm Tr}(\gamma^{mn}\gamma_{pq}) = 16(\delta^m_q\delta^n_p - \delta^m_p\delta^n_q)$ from (A.24).

The discrepancy in the coefficient of the double pole between (3.18) and (3.19) leads to analogous discrepancies in the computations of gluon scattering amplitudes using the RNS vertex operators (3.16) and those of Siegel in (3.15).

[8]The difference in the sign of the second term is due to our cavalier attitude towards factors of i and can be understood as follows: for plane wave $e^{ik\cdot X}$, the RNS vertex operator is proportional to $e^m(i\partial X_m + (k\cdot\psi)\psi_m)$, so rescaling $k\to -ik$ to our convention and dividing by i yields $e_i^m\partial X_m - \frac{1}{2}\psi^m\psi^n f^i_{mn}$.

Missing constraints. Finally in Siegel's formulation (3.6) one would need to include an appropriate set of first-class constraints to reproduce the superstring spectrum: the Virasoro constraint T and the kappa symmetry generator G of the GS formalism

$$T = -\frac{1}{2}\Pi^m\Pi_m - d_\alpha\partial\theta^\alpha\,, \qquad G^\alpha = \Pi^m(\gamma_m d)^\alpha \tag{3.20}$$

in terms of the supersymmetric momentum and GS constraints should certainly be elements of that set of constraints. Even though there was a successful description of the superparticle using Siegel's approach [102, 103], the whole set of constraints was never found for the superstring case. Nevertheless, Siegel's idea was not lost as it was used by Berkovits in his proposal for the pure spinor formalism [1].

3.3. Fundamentals of the pure spinor formalism

We have seen above that while Siegel's approach circumvented the difficulties associated to the GS constraint, the non-vanishing central charge $c_X + c_{p\theta} = -22$ and the level $+4$ of the Lorentz current algebra presented serious challenges to this new formulation. This motivated Berkovits to modify Siegel's approach by introducing *pure spinor* ghost variables contributing $+22$ to the central charge of the energy momentum tensor and -3 to the double pole in the OPE of the Lorentz currents, thereby fixing the most pressing issues with the formulation by Siegel and leading to Berkovits' *pure spinor formalism* [1]. Let us briefly review below some of the central elements in this reformulation.

Lorentz currents for the ghosts. Berkovits' idea was to modify the Lorentz currents (3.8) by the addition of a contribution N^{mn} coming from ghosts,

$$M^{mn} = \Sigma^{mn} + N^{mn}. \tag{3.21}$$

The newly defined M^{mn} would satisfy the same OPE (3.18) as in the RNS formalism if the contribution to the double pole arising from the ghosts N^{mn} had a coefficient -3

$$\begin{aligned} N^{mn}(z)N^{pq}(w) &\sim \frac{\delta^{p[m}N^{n]q}(w) - \delta^{q[m}N^{n]p}(w)}{z-w} - 3\frac{\delta^{m[q}\delta^{p]n}}{(z-w)^2}\,, \\ \Sigma^{mn}(z)N^{pq}(w) &\sim \text{regular}\,. \end{aligned} \tag{3.22}$$

This would fix the issue with the Lorentz current OPE and set the level of the overall Lorentz currents M^{mn} to $4-3=1$, in lines with the level of the RNS currents in (3.18).

Energy-momentum tensor for the ghosts. To fix the problem with the non-vanishing central charge of the energy-momentum tensor in Siegel's approach, one would need these same ghosts to have a central charge $c_\lambda = +22$. Fortunately, the right solution to both problems was found when a proposal for the BRST charge was put forward and the need for *pure spinors* became evident.

The BRST operator. The next step in the line of reasoning which led to the pure spinor formalism is the proposal of the BRST operator

$$Q_{\text{BRST}} = \oint dz\, \lambda^\alpha(z) d_\alpha(z)\,, \tag{3.23}$$

where λ^α are bosonic spinors and the Siegel variable d_α corresponding to the GS constraint has been defined in (3.7). The BRST charge (3.23) must satisfy the consistency condition $Q^2_{\text{BRST}} = 0$, otherwise the BRST charge itself would not be invariant under a variation of the gauge constraint [12]. Using (3.23) and the OPE (3.13) we obtain

$$Q^2_{\text{BRST}} = \frac{1}{2}\{Q_{\text{BRST}}, Q_{\text{BRST}}\} = -\frac{1}{2}\oint dz\,(\lambda\gamma^m\lambda)\Pi_m\,. \tag{3.24}$$

Therefore imposing that the BRST charge is nilpotent

$$Q^2_{\text{BRST}} = 0 \tag{3.25}$$

implies that the bosonic fields λ^α must satisfy the *pure spinor* constraints

$$\lambda\gamma^m\lambda = 0\,, \tag{3.26}$$

which were first studied by Cartan from a geometrical perspective [104].

3.3.1. $U(5)$ decompositions

The formalism discovered by Berkovits is based on the properties of the pure spinor λ^α, and it is important to identify the number of degrees of freedom which survive the constraints (3.26). Naively, one could think that those ten constraints associated with $m = 0, 1, \ldots, 9$ would imply a pure spinor of $SO(1, 9)$ to have only $16 - 10 = 6$ degrees of freedom. However, this not the case; we will see below that a pure spinor has eleven degrees of freedom.

$U(5)$ decomposition of pure spinors. In order to see that a pure spinor has eleven degrees of freedom, it is convenient to Wick rotate $SO(1, 9)$ to $SO(10)$ and to break manifest $SO(10)$ symmetry to its $U(5)$ subgroup [1]. The explicit calculations are shown in Appendix B, with the result that the pure spinor decomposes into irreducible $U(5)$ representations as $\mathbf{16} \longrightarrow (\mathbf{1}, \overline{\mathbf{10}}, \mathbf{5})$, or more explicitly as

$$\lambda^\alpha \longrightarrow (\lambda^+, \lambda_{ab}, \lambda^a)\,,\ \text{with } \lambda_{ba} = -\lambda_{ab} \text{ and } \lambda^a = \frac{1}{8\lambda^+}\epsilon^{abcde}\lambda_{bc}\lambda_{de}\,,\quad a, b, c, d, e = 1, \ldots, 5 \tag{3.27}$$

for $\lambda^+ \neq 0$, where ϵ^{abcde} is totally antisymmetric with $\epsilon^{12345} = 1$. In this language, λ_{ab} parameterize a $SO(10)/U(5)$ coset. The pure spinor constraint (3.26) only eliminates the $\mathbf{5} \ni \lambda^a$ in favor of $\lambda^+ \in \mathbf{1}$ and $\lambda_{ab} \in \overline{\mathbf{10}}$. Hence, there remains $1 + 10 = 11$ degrees of freedom in a pure spinor of $SO(10)$.

Note that in absence of Wick rotation the λ_{ab} parameterize the compact space $SO(1, 9)/(U(4) \times \mathcal{R}^9)$ with $\mathcal{R}^9$ representing nine light-like boosts [105, 18, 106].

$U(5)$ decomposition of the Lorenz currents. To solve the pure spinor constraint (3.26) it was convenient to break the manifest $SO(10)$ symmetry to its subgroup $U(5)$, so a pure spinor is written in terms of $U(5) = SU(5) \otimes U(1)$ variables. Consequently, the Lorentz currents must also be decomposed to their irreducible $SU(5)$ representations

$$N^{mn} \longrightarrow (n, n^b_a, n_{ab}, n^{ab})\,, \tag{3.28}$$

in a manner specified in Appendix B. In the remainder of this section, these $SU(5)$ Lorentz currents will be constructed out of elementary ghost variables to be denoted by $s(z)$, $u_{ab}(z)$ and their conjugate momenta $t(z), v^{ab}(z)$ such that the required condition (3.22) is met. To do this we will first state how the OPE (3.22) decomposes under $SO(10) \to SU(5) \otimes U(1)$ given by (3.28):

Proposition 1. *The SO(10)-covariant OPE of the Lorentz currents*

$$N^{mn}(z)N^{pq}(w) \sim \frac{\delta^{mp}N^{nq}(w) - \delta^{np}N^{mq}(w) - \delta^{mq}N^{np}(w) + \delta^{nq}N^{mp}(w)}{z-w} - 3\frac{\left(\delta^{mq}\delta^{np} - \delta^{mp}\delta^{nq}\right)}{(z-w)^2}\,, \tag{3.29}$$

implies that the $SU(5) \otimes U(1)$ currents $(n, n^b_a, n_{ab}, n^{ab})$ satisfy the following OPEs:

$$n_{ab}(z)n_{cd}(w) \sim \text{regular}\,, \qquad n^{ab}(z)n^{cd}(w) \sim \text{regular}\,, \tag{3.30}$$

$$n_{ab}(z)n^{cd}(w) \sim \frac{-\delta^c_{[a}n^d_{b]}(w) + \delta^d_{[a}n^c_{b]}(w) - \frac{2}{\sqrt{5}}\delta^c_{[a}\delta^d_{b]}n(w)}{z-w} - 3\frac{\delta^c_b\delta^d_a - \delta^c_a\delta^d_b}{(z-w)^2}\,, \qquad n(z)n^a_b(w) \sim \text{regular}\,,$$

$$n^a_b(z)n^c_d(w) \sim \frac{-\delta^c_b n^a_d(w) + \delta^a_d n^c_b(w)}{z-w} - 3\frac{\delta^a_d\delta^c_b - \frac{1}{5}\delta^a_b\delta^c_d}{(z-w)^2}\,, \qquad n(z)n_{ab}(w) \sim +\frac{2}{\sqrt{5}}\frac{n_{ab}(w)}{z-w}\,,$$

$$n^{ab}(z)n^c_d(w) \sim \frac{-\delta^a_d n^{bc}(w) + \delta^b_d n^{ac}(w) - \frac{2}{5}\delta^c_d n^{ab}(w)}{z-w}\,, \qquad n(z)n^{ab}(w) \sim -\frac{2}{\sqrt{5}}\frac{n^{ab}(w)}{z-w}\,,$$

$$n_{ab}(z)n^c_d(w) \sim \frac{-\delta^c_b n_{ad}(w) + \delta^c_a n_{bd}(w) + \frac{2}{5}\delta^c_d n_{ab}(w)}{z-w}\,, \qquad n(z)n(w) \sim -\frac{3}{(z-w)^2}\,.$$

Proof. See Appendix B.2 and also [107, 108]. □

$U(5)$ decomposition of spinors. There is one more consistency condition to be obeyed when constructing the $U(5)$ Lorentz currents. The pure spinor λ^α must obviously transform as a spinor under the action of the total Lorentz current M^{mn} in (3.21),

$$\delta\lambda^\alpha = \frac{1}{2}\left[\oint dz\, \varepsilon_{mn} M^{mn}, \lambda^\alpha\right] = \frac{1}{4}\varepsilon_{mn}(\gamma^{mn}\lambda)^\alpha \,. \tag{3.31}$$

Since the OPE of λ^α with the Lorentz currents Σ^{mn} of (3.8) is regular we conclude that the pure spinor must satisfy (3.32) given below. Given the solution of the pure spinor constraint (3.27) in $U(5)$ variables we need to know the group-theoretic decomposition of how a $SO(10)$ spinor transforms in terms of its $U(5)$ representations.

Proposition 2. *The $SO(10)$-covariant transformation of a spinor*

$$N^{mn}(z)\lambda^\alpha(w) \sim \frac{1}{2}\frac{(\gamma^{mn})^\alpha{}_\beta \lambda^\beta(w)}{(z-w)} \,, \tag{3.32}$$

implies that the OPEs among the $SU(5)$ representations $(n, n^a_b, n_{ab}, n^{ab})$ and $(\lambda^+, \lambda_{cd}, \lambda^c)$ are given by

$$\begin{aligned}
n(z)\lambda^+(w) &\sim -\frac{\sqrt{5}}{2}\frac{\lambda^+(w)}{z-w}\,, & n(z)\lambda_{cd}(w) &\sim -\frac{1}{2\sqrt{5}}\frac{\lambda_{cd}(w)}{z-w}\,, \\
n(z)\lambda^c(w) &\sim \frac{3}{2\sqrt{5}}\frac{\lambda^c(w)}{z-w}\,, & n^a_b(z)\lambda^+(w) &\sim \text{regular}\,, \\
n^a_b(z)\lambda_{cd}(w) &\sim \frac{\delta^a_d\lambda_{cb}(w) - \delta^a_c\lambda_{db}(w)}{(z-w)} - \frac{2}{5}\frac{\delta^a_b\lambda_{cd}(w)}{(z-w)}\,, & n^a_b(z)\lambda^c(w) &\sim \frac{1}{5}\frac{\delta^a_b\lambda^c(w)}{(z-w)} - \frac{\delta^c_b\lambda^a(w)}{(z-w)}\,, \\
n_{ab}(z)\lambda^+(w) &\sim \frac{\lambda_{ab}(w)}{z-w}\,, & n_{ab}(z)\lambda_{cd}(w) &\sim \frac{\epsilon_{abcde}\lambda^e(w)}{z-w}\,, \\
n_{ab}(z)\lambda^c(w) &\sim \text{regular}\,, & n^{ab}(z)\lambda^+(w) &\sim \text{regular}\,, \\
n^{ab}(z)\lambda_{cd}(w) &\sim -\frac{\delta^{[a}_c\delta^{b]}_d\lambda^+(w)}{z-w}\,, & n^{ab}(z)\lambda^c(w) &\sim -\frac{\epsilon^{abcde}\lambda_{de}(w)}{2(z-w)}\,.
\end{aligned} \tag{3.33}$$

Proof. See Appendix B and also [107]. □

It turns out that all these OPEs can be reproduced from an action involving the ghost variables $s(z)$, $u_{ab}(z)$, $t(z)$ and $v^{cd}(z)$ below that serve as the ingredients of the Lorentz currents $(n, n^b_a, n_{ab}, n^{ab})$ and pure spinor $(\lambda^+, \lambda_{cd}, \lambda^c)$. The pure spinor formalism crucially hinges on the existence of such a construction.

Before moving on, note the consistency between the OPE (3.32) and the simple pole of (3.29) arising from a twofold application of the spinorial transformation. That is, if $[N^{mn}, \lambda^\alpha] = \frac{1}{2}(\gamma^{mn})^\alpha{}_\beta\lambda^\beta$ then $[N^{pq}, [N^{mn}, \lambda^\alpha]] = \frac{1}{4}(\gamma^{mn})^\alpha{}_\beta(\gamma^{pq})^\beta{}_\delta\lambda^\delta$ which implies,

$$\begin{aligned}
[[N^{mn}, N^{pq}], \lambda^\alpha] &= [N^{mn}, [N^{pq}, \lambda^\alpha]] - [N^{pq}, [N^{mn}, \lambda^\alpha]] = \frac{1}{4}\big[(\gamma^{pq}\gamma^{mn})^\alpha{}_\beta - (\gamma^{mn}\gamma^{pq})^\alpha{}_\beta\big]\lambda^\beta \\
&= \delta^{mp}[N^{nq}, \lambda^\alpha] - \delta^{np}[N^{mq}, \lambda^\alpha] - \delta^{mq}[N^{np}, \lambda^\alpha] + \delta^{nq}[N^{mp}, \lambda^\alpha]\,,
\end{aligned} \tag{3.34}$$

where we used the gamma-matrix identity $\gamma^{pq}\gamma^{mn} - \gamma^{mn}\gamma^{pq} = 2\delta^{mp}\gamma^{nq} - 2\delta^{np}\gamma^{mq} - 2\delta^{mq}\gamma^{np} + 2\delta^{nq}\gamma^{mp}$ which follows from the product relation (A.31). These OPEs play a crucial role in evaluating the CFT correlation functions for string amplitudes and will for instance be used in the derivation of the multiparticle vertex operators at multiplicity two in section 4.1 [109, 83].

3.3.2. The pure spinor ghosts

In this section we will display the solution to the above problems found by Berkovits with the introduction of a specific $U(5)$ parameterization of pure spinors, Lorentz currents and the energy-momentum tensor.

The action for the ghosts appearing in the pure spinor constraint is given by [18, 1, 92]

$$S_\lambda = \frac{1}{2\pi}\int d^2z\Big(-\partial t\overline{\partial}s + \frac{1}{2}v^{ab}\overline{\partial}u_{ab}\Big)\,, \qquad a,b = 1,\ldots,5\,, \tag{3.35}$$

where $t(z)$ and $v^{ab}(z)$ are the conjugate momenta for $s(z)$ and $u_{ab}(z)$. Furthermore, $s(z)$ and $t(z)$ are chiral bosons, so one must impose their equations of motions by hand, $\overline{\partial}s = \overline{\partial}t = 0$. The OPEs are given by

$$\begin{aligned} t(z)s(w) &\sim \ln(z-w)\,, \\ v^{ab}(z)u_{cd}(w) &\sim \frac{\delta^a_c\delta^b_d - \delta^a_d\delta^b_c}{z-w}\,. \end{aligned} \tag{3.36}$$

Matching group theory with CFT. The fundamental result allowing the construction of the pure spinor formalism is given by the explicit construction of the $U(5)$ Lorentz currents $(n, n^a_b, n_{ab}, n^{ab})$ and pure spinors $(\lambda^+, \lambda_{ab}, \lambda^a)$ in terms of the ghost variables $s(z)$, $t(z)$, $v^{ab}(z)$ and $u_{ab}(z)$ from the action (3.35). This has to be done in such a way as that their $U(5)$ OPEs among themselves satisfy all the group-theoretic relations (3.30) and (3.33). The solution found by Berkovits is given by[9] [18, 1, 92]

$$\begin{aligned} n &= -\frac{1}{\sqrt{5}}\Big(\frac{1}{4}u_{ab}v^{ab} + \frac{5}{2}\partial t - \frac{5}{2}\partial s\Big)\,, & \lambda^+ &= e^s\,, \\ n^a_b &= -u_{bc}v^{ac} + \frac{1}{5}\delta^a_b u_{cd}v^{cd}\,, & \lambda_{ab} &= u_{ab}\,, \\ n^{ab} &= -e^s v^{ab}\,, & \lambda^a &= \frac{1}{8}e^{-s}\epsilon^{abcde}u_{bc}u_{de}\,, \\ n_{ab} &= -e^{-s}\Big(2\partial u_{ab} - u_{ab}\partial t - 2u_{ab}\partial s + u_{ac}u_{bd}v^{cd} - \frac{1}{2}u_{ab}u_{cd}v^{cd}\Big)\,. \end{aligned} \tag{3.37}$$

The unusual normalization of $n(z)$ was chosen such that the coefficient of its double pole is -3. Straightforward but long calculations show that their OPEs among themselves reproduce all OPEs in (3.30) and (3.33), provided that the ghost variables $s(z)$, $t(z)$, $v^{ab}(z)$ and $u_{ab}(z)$ satisfy the OPEs (3.36). For instance, two sample calculations are

$$\begin{aligned} n(z)n^{ab}(w) &= \frac{1}{\sqrt{5}}\Big(\frac{1}{4}u_{fg}(z)v^{fg}(z) + \frac{5}{2}\partial t(z) - \frac{5}{2}\partial s(z)\Big)e^{s(w)}v^{ab}(w) \\ &= \frac{1}{\sqrt{5}}\frac{1}{4}e^{s(w)}v^{fg}(z)u_{fg}(z)v^{ab}(w) - \frac{\sqrt{5}}{2}\partial t(z)e^{s(w)}v^{ab}(w) \\ &\sim \frac{1}{\sqrt{5}}\frac{1}{4}e^{s(w)}v^{fg}(z)\frac{(-\delta^a_f\delta^b_g + \delta^a_g\delta^b_f)}{z-w} - \frac{\sqrt{5}}{2}\frac{1}{z-w}e^{s(w)}v^{ab}(w) \\ &\sim -\frac{2}{\sqrt{5}}\frac{n^{ab}(w)}{z-w} \end{aligned} \tag{3.38}$$

and

$$\begin{aligned} n^{ab}(z)\lambda^c(w) &= -\frac{1}{8}e^{s(z)}\epsilon^{cdefg}\big(v^{ab}(z)u_{de}(w)u_{fg}(w) + u_{de}(w)v^{ab}(z)u_{fg}(w)\big)e^{-s(w)} \\ &\sim -\frac{1}{8}e^{s(z)}\frac{\big(2\epsilon^{cabfg}u_{fg}(w) + 2\epsilon^{cdeab}u_{de}(w)\big)}{z-w}e^{-s(w)} \\ &\sim -\frac{1}{2}\epsilon^{abcde}\frac{\lambda_{de}(w)}{z-w}\,, \end{aligned} \tag{3.39}$$

[9]Note the sign flip of the Lorentz generators and of λ^a with respect to [92]. This ensures that the conventions of Appendix B.2 are respected.

where we used the OPEs $u_{fg}(z)v^{ab}(w) \sim \frac{-\delta_f^a\delta_g^b+\delta_g^a\delta_f^b}{z-w}$ and $\partial t(z)e^{s(w)} \sim \frac{1}{z-w}e^{s(w)}$ that follow from (3.36). The above results reproduce two of the OPEs in (3.30) and (3.33) that were obtained from a group-theoretic decomposition of the parental $SO(10)$-covariant OPEs. All the other OPEs can be verified similarly. Therefore, even though the action for the ghosts S_λ is not manifestly Lorentz covariant, all OPEs involving N^{mn} and λ^α descend from manifestly $SO(10)$-covariant expressions. So the pure spinor formalism has manifest Lorentz covariance.

Energy-momentum tensor. We will show that the central charge of the energy-momentum tensor for the ghosts

$$T_\lambda = \frac{1}{2}v^{ab}\partial u_{ab} + \partial t\partial s + \partial^2 s\,, \tag{3.40}$$

following from the ghost action (3.35) is +22. This is indeed the required value for it to annihilate the total central charge when added to Siegel's matter variables. The derivation of (3.40) follows from Noether's procedure using

$$\delta S_\lambda = \frac{1}{2\pi}\int d^2z\,\left[\overline{\partial}\varepsilon T_\lambda(z) + \partial\overline{\varepsilon}\,\overline{T}_\lambda(\overline{z})\right], \tag{3.41}$$

under the conformal transformations of $(v^{ab}, u_{ab}, \partial s, \overline{\partial}t)$ whose conformal weights are $(1,0)$, $(0,0)$, $(1,0)$ and $(0,1)$, respectively,

$$\begin{aligned}
\delta v^{ab} &= \partial\varepsilon v^{ab} + \varepsilon\partial v^{ab} + \overline{\varepsilon}\overline{\partial}v^{ab}\,, & \delta u_{ab} &= \varepsilon\partial u_{ab} + \overline{\varepsilon}\overline{\partial}u_{ab}\,,\\
\delta\partial s &= \partial\varepsilon\partial s + \varepsilon\partial^2 s + \partial\overline{\varepsilon}\overline{\partial}s + \overline{\partial}\overline{\varepsilon}\partial s\,, & \delta\overline{\partial}t &= \varepsilon\partial\overline{\partial}t + \overline{\partial}\varepsilon\partial t + \overline{\partial}\overline{\varepsilon}\overline{\partial}t + \overline{\varepsilon}\overline{\partial}^2 t\,,
\end{aligned} \tag{3.42}$$

and requiring the Lorentz currents $(n, n^a_b, n^{ab}, n_{ab})$ to be primary fields [1, 92] (see also [107] for the explicit calculations).

Proposition 3. *The central charge of the energy-momentum tensor for the ghosts (3.40) is $c_\lambda = 22$.*

Proof. The central charge is determined from the fourth-order pole in $T_\lambda(z)T_\lambda(w) \sim \frac{(c_\lambda/2)}{(z-w)^4} + \cdots$. There are two contributions

$$\begin{aligned}
\frac{1}{4}v^{ab}(z)\partial u_{ab}(z)v^{cd}(w)\partial u_{ab}(w) &= \frac{1}{4}\frac{\delta_c^{[a}\delta_d^{b]}\delta_a^{[c}\delta_b^{d]}}{(z-w)^4} = \frac{10}{(z-w)^4}\,,\\
\partial t(z)\partial s(z)\partial t(w)\partial s(w) &= \frac{1}{(z-w)^4}\,,
\end{aligned} \tag{3.43}$$

whose sum implies that $c_\lambda = +22$. ☐

Therefore, as there are no poles between the ghosts and matter variables, the total central charge of the energy-momentum tensor in the pure spinor formalism

$$T_{\rm PS} = -\frac{1}{2}\partial X^m\partial X_m - p_\alpha\partial\theta^\alpha + \frac{1}{2}v^{ab}\partial u_{ab} + \partial t\partial s + \partial^2 s\,, \tag{3.44}$$

vanishes; $c_X + c_{p\theta} + c_\lambda = 10 - 32 + 22 = 0$. Therefore there will not be a conformal anomaly in the formalism.

3.3.3. The action of the pure spinor formalism

$U(5)$-covariant action. From the discussion above we learn that adding the pure spinor ghost action of (3.35) to the Siegel action (3.6) implies that the energy-momentum tensor of the theory has vanishing central charge, as $c_X + c_{p\theta} = -22$ from the matter variables is neutralized by $c_\lambda = 22$ from the ghosts. Furthermore, the Lorentz currents of the combined actions have the same OPE as in the RNS formalism. Berkovits then proposed that the pure spinor formalism action for the left-moving fields is given by [1]

$$S_{\rm PS} = \frac{1}{\pi}\int d^2z\left(\frac{1}{2}\partial X^m\overline{\partial}X_m + p_\alpha\overline{\partial}\theta^\alpha - \partial t\overline{\partial}s + \frac{1}{2}v^{ab}\overline{\partial}u_{ab}\right). \tag{3.45}$$

Spacetime supersymmetry transformations are generated by

$$\mathcal{Q}_\alpha = \oint dz \Big(p_\alpha + \frac{1}{2}(\gamma^m\theta)_\alpha \partial X_m + \frac{1}{24}(\gamma^m\theta)_\alpha(\theta\gamma_m\partial\theta)\Big)\,, \tag{3.46}$$

and their action on the variables in the pure spinor formalism with Weyl-spinor parameter ε^α is given by

$$\begin{aligned} \delta X^m &= \frac{1}{2}\,(\varepsilon\gamma^m\theta)\,, \quad \delta\theta^\alpha = \varepsilon^\alpha\,, \\ \delta p_\beta &= -\frac{1}{2}(\varepsilon\gamma^m)_\beta\partial X_m + \frac{1}{8}(\varepsilon\gamma_m\theta)(\partial\theta\gamma^m)_\beta\,, \\ \delta s &= \delta t = \delta u_{ab} = \delta v^{ab} = 0\,. \end{aligned} \tag{3.47}$$

The action (3.45) is found to be supersymmetric by exploiting that the total derivatives

$$\begin{aligned} \partial\big[(\varepsilon\gamma_m\theta)\overline{\partial}X^m\big] - \overline{\partial}\big[(\varepsilon\gamma_m\theta)\partial X^m\big] &= (\varepsilon\gamma_m\partial\theta)\overline{\partial}X^m - (\varepsilon\gamma_m\overline{\partial}\theta)\partial X^m \\ \partial\big[(\varepsilon\gamma_m\theta)(\theta\gamma^m\overline{\partial}\theta)\big] - \overline{\partial}\big[(\varepsilon\gamma_m\theta)(\theta\gamma^m\partial\theta)\big] &= 3(\varepsilon\gamma_m\theta)(\partial\theta\gamma^m\overline{\partial}\theta) \end{aligned} \tag{3.48}$$

from the variation of $\partial X^m\overline{\partial}X_m$ and the θ^2-contribution to δp_β integrate to zero under d^2z.

$SO(10)$-covariant action. The action (3.45) in the pure spinor formalism can be written covariantly as

$$S_{\rm PS} = \frac{1}{\pi}\int d^2z\left(\frac{1}{2}\partial X^m\overline{\partial}X_m + p_\alpha\overline{\partial}\theta^\alpha - w_\alpha\overline{\partial}\lambda^\alpha\right), \tag{3.49}$$

where w_α is the conjugate momentum to the pure spinor. The dependence on α' can be reinstated from the following length dimensions of all these variables [110, 111][10]

$$[\alpha'] = 2\,, \quad [X^m] = 1\,, \quad [\theta^\alpha] = [\lambda^\alpha] = \frac{1}{2}\,, \quad [p_\alpha] = [w_\alpha] = -\frac{1}{2}\,. \tag{3.50}$$

Inspired by the approach of Siegel, this action needs to be supplemented by the definitions of the supersymmetric momentum Π^m, the GS constraint d_α and the supersymmetric derivative D_α which we repeat here for the reader's convenience:

$$\begin{aligned} \Pi^m &= \partial X^m + \frac{1}{2}(\theta\gamma^m\partial\theta)\,, \\ d_\alpha &= p_\alpha - \frac{1}{2}\Big(\partial X^m + \frac{1}{4}(\theta\gamma^m\partial\theta)\Big)(\gamma_m\theta)_\alpha\,, \\ D_\alpha &= \frac{\partial}{\partial\theta^\alpha} + \frac{1}{2}(\gamma^m\theta)_\alpha\partial_m\,. \end{aligned} \tag{3.51}$$

In addition, the BRST charge is given by[11] (dropping the subscript $_{\rm BRST}$ henceforth)

$$Q = \oint dz\,\lambda^\alpha(z)d_\alpha(z)\,. \tag{3.52}$$

The $SO(10)$-covariant versions of the energy-momentum tensor (3.44) and the fermionic Lorentz currents derived from the action (3.49) are given by

$$T_{\rm PS} = -\frac{1}{2}\Pi^m\Pi_m - d_\alpha\partial\theta^\alpha + w_\alpha\partial\lambda^\alpha\,, \qquad M^{mn} = -\frac{1}{2}(p\gamma^{mn}\theta) + \frac{1}{2}(w\gamma^{mn}\lambda)\,. \tag{3.53}$$

[10] We omitted all factors of α' for brevity and maximum flexibility. For the open- and closed-string they can be restored from the conventions $\alpha' = 1/2$ and $\alpha' = 2$ respectively.

[11] In recent years this BRST charge has been derived from first principles [112]. For previous attempts, see [113, 114, 115, 116].

3.3.4. Operator product expansions

We shall now summarize the $SO(10)$-covariant form of the OPEs that govern the CFT of the pure spinor formalism. The basic worldsheet matter variables obey

$$\begin{aligned}
X^m(z,\overline{z})X^n(w,\overline{w}) &\sim -\delta^{mn}\ln|z-w|^2\,, & d_\alpha(z)\theta^\beta(w) &\sim \frac{\delta_\alpha^\beta}{z-w}\,, \\
d_\alpha(z)d_\beta(w) &\sim -\frac{\gamma^m_{\alpha\beta}\Pi_m(w)}{z-w}\,, & d_\alpha(z)\Pi^m(w) &\sim \frac{\big(\gamma^m\partial\theta(w)\big)_\alpha}{z-w}\,, \\
\Pi^m(z)\Pi^n(w) &\sim -\frac{\delta^{mn}}{(z-w)^2}\,,
\end{aligned} \tag{3.54}$$

the OPEs involving the Lorentz currents are

$$\begin{aligned}
M^{mn}(z)M^{pq}(w) &\sim \frac{\delta^{p[m}M^{n]q}(w)-\delta^{q[m}M^{n]p}(w)}{z-w}+\frac{\delta^{m[q}\delta^{p]n}}{(z-w)^2}\,, \\
N^{mn}(z)N^{pq}(w) &\sim \frac{\delta^{p[m}N^{n]q}(w)-\delta^{q[m}N^{n]p}(w)}{z-w}-3\frac{\delta^{m[q}\delta^{p]n}}{(z-w)^2}\,, \\
N^{mn}(z)\lambda^\alpha(w) &\sim \frac{1}{2}\frac{(\gamma^{mn})^\alpha{}_\beta\lambda^\beta(w)}{(z-w)}\,,
\end{aligned} \tag{3.55}$$

and generic superfields $K(X,\theta)$ that do not depend on any derivatives $\partial^k X^m,\ \partial^k\theta^\alpha$ with $k\geq 1$ obey

$$\begin{aligned}
d_\alpha(z)K\big(X(w,\overline{w}),\theta(w)\big) &\sim \frac{D_\alpha K\big(X(w,\overline{w}),\theta(w)\big)}{z-w}\,, \\
\Pi^m(z)K\big(X(w,\overline{w}),\theta(w)\big) &\sim -\frac{\partial^m K\big(X(w,\overline{w}),\theta(w)\big)}{z-w}\,.
\end{aligned} \tag{3.56}$$

Using these OPEs one can check that the supersymmetry currents (3.46) satisfy the supersymmetry algebra

$$\{\mathcal{Q}_\alpha,\mathcal{Q}_\beta\} = \gamma^m_{\alpha\beta}\oint\partial X_m\,, \tag{3.57}$$

and that all of $\{\partial\theta^\alpha,\Pi^m,d_\alpha,N^{mn}\}$ are conformal primary fields of weight $+1$,

$$T_{\rm PS}(z)\big\{\partial\theta^\alpha,\Pi^m,d_\alpha,N^{mn}\big\}(w) \sim \frac{\big\{\partial\theta^\alpha,\Pi^m,d_\alpha,N^{mn}\big\}(w)}{(z-w)^2}+\frac{\partial\big\{\partial\theta^\alpha,\Pi^m,d_\alpha,N^{mn}\big\}(w)}{z-w}\,, \tag{3.58}$$

a crucial fact in the construction of the integrated massless vertex operator below.

3.4. Scattering amplitudes on the disk

We shall now review the opening line for superstring disk amplitudes along with the dictionary between superspace expressions and component amplitudes.

3.4.1. Massless vertex operators

In order to compute scattering amplitudes in superstring theory using conformal-field-theory methods, first we need to describe the vertex operators containing the information about the string states. The integrated massless vertex operator (3.14) proposed by Siegel led to the discrepancy of double-pole coefficients due to the Lorentz currents of the fermionic variables. The integrated massless vertex operator proposed by Berkovits adds a correction to $U_{\rm Siegel}(z)$ proportional to the Lorentz current N_{mn} of the pure spinor ghost [1]

$$U(z) = \partial\theta^\alpha A_\alpha(X,\theta)+A_m(X,\theta)\Pi^m+d_\alpha W^\alpha(X,\theta)+\frac{1}{2}N_{mn}F^{mn}(X,\theta)\,, \tag{3.59}$$

where the linearized SYM superfields $A_\alpha(X,\theta), A_m(X,\theta), W^\alpha(X,\theta)$ and $F^{mn}(X,\theta)$ were introduced in section 2.2 and the dependence $\theta^\alpha = \theta^\alpha(z)$ and $X^m = X^m(z,\overline{z})$ on the vertex insertion points is left implicit. The superfields have the following length dimensions [110, 111]

$$[A_\alpha] = \frac{1}{2}\,, \quad [A_m] = 0\,, \quad [W^\alpha] = -\frac{1}{2}\,, \quad [F_{mn}] = -1\,, \quad [V(z)] = [U(z)] = 1\,, \tag{3.60}$$

and the superfields $K(X,\theta)$ are decomposed into plane waves as

$$K(X,\theta) = K(\theta)e^{k\cdot X}\,. \tag{3.61}$$

Using the θ-expansions (2.17) in Harnad–Shnider gauge, the gluon vertex operator following from (3.59) features the complete Lorentz current $M^{mn}(z) = \Sigma^{mn}(z) + N^{mn}(z)$ of (3.21) as the coefficient of the component field strength. In this way, the issue with the double-pole mismatch with the RNS vertex operator is absent from (3.59). Therefore it will appear in the amplitude prescription integrated over (parts of) the worldsheet boundary, i.e. in the conformally invariant combination $\int dz U(z)$.

The prescription to compute tree-level amplitudes will also require a massless vertex operator with conformal weight zero to be used at fixed locations on the Riemann surface to remove the redundancy of the Möbius transformations. The proposal by Berkovits for this unintegrated vertex is

$$V = \lambda^\alpha A_\alpha(X,\theta)\,. \tag{3.62}$$

Furthermore, the massless vertex operators represent the physical states of gluons and gluinos and must *be in the cohomology* of the BRST operator Q of (3.52).

Definition 1. *A state Ψ is said to be in the cohomology of the BRST operator if it is BRST-closed, $Q\Psi = 0$, and not BRST-exact, $\Psi \neq Q\Omega$ for some Ω.*

Recall that the BRST charge satisfies $Q^2 = 0$ due to the pure spinor condition (3.26) and the OPE (3.13).

Proposition 4. *The unintegrated vertex operator $V(z) = \lambda^\alpha(z)A_\alpha(X,\theta)$ for massless particles $k^2 = 0$ is BRST closed $QV = 0$ when the linearized superfield $A_\alpha(X,\theta)$ is on-shell and has zero conformal weight.*

Proof. An on-shell linearized superfield A_α satisfies the equations of motion (2.14). In particular $D_{(\alpha}A_{\beta)} = \gamma^m_{\alpha\beta}A_m$, so

$$QV(w) = \oint dz\, \lambda^\alpha(z)d_\alpha(z)\lambda^\beta(w)A_\beta\big(X(w),\theta(w)\big) = \lambda^\alpha\lambda^\beta D_\alpha A_\beta = \frac{1}{2}(\lambda\gamma^m\lambda)A_m = 0\,, \tag{3.63}$$

where we used the OPE (3.56) and the pure spinor constraint (3.26). To show that V has conformal weight zero, first recall that in a conformal field theory the OPE of the energy-momentum tensor with a conformal primary ϕ_h of weight h is given by [101, 12]

$$T(z)\phi_h(w) \sim \frac{h\phi_h(w)}{(z-w)^2} + \frac{\partial_w\phi_h(w)}{(z-w)}\,. \tag{3.64}$$

Using the total energy-momentum tensor T_{PS} from (3.53) and the OPEs (3.56) we get

$$T(z)V(w) \sim -\frac{1}{2}\frac{\partial^m\partial_m V}{(z-w)^2} + \frac{(\Pi^m\partial_m + \partial\theta^\alpha D_\alpha)V + \partial\lambda^\alpha A_\alpha}{(z-w)} = \frac{\partial V}{(z-w)}\,, \tag{3.65}$$

where we used the massless condition and the chain rule for ∂_w

$$(\Pi^m\partial_m + \partial\theta^\alpha D_\alpha)V + \partial\lambda^\alpha A_\alpha = \lambda^\alpha\partial A_\alpha(X,\theta) + (\partial\lambda^\alpha)A_\alpha(X,\theta) = \partial V(X(w),\theta(w)) = \partial V(w) \tag{3.66}$$

since

$$(\partial\theta^\beta D_\beta + \Pi^m\partial_m)K(X,\theta) = (\partial\theta^\beta\partial_\beta + \partial X^m\partial_m)K(X,\theta) = \partial K(X,\theta) \tag{3.67}$$

for an arbitrary superfield $K(X,\theta)$ that is independent on λ^α and on the worldsheet derivatives of X^m, θ^α, as can easily be checked using the expressions for D_α and Π^m in (3.51). $\square$

As can be seen from (2.15), a BRST-exact vertex operator of the form $Q\Omega$ is interpreted as capturing the gauge variation of the super Yang–Mills fields $Q\Omega = \lambda^\alpha D_\alpha \Omega = \lambda^\alpha \delta_\Omega A_\alpha$. In this sense, viewing V as a representative in the cohomology of Q excludes pure-gauge superfields.

The synergy between pure spinors and the SYM equations of motion seen in (3.63) was already anticipated by Howe and Nilsson in [117, 118], also see [119, 120] for a more recent overview articles on the importance of pure spinors for off-shell supersymmetric actions. An early application of ten-dimensional pure spinors to the classical superstring can be found in [121].

Relating integrated and unintegrated vertices. In the RNS formalism, the integrated vertex operator $U_{\rm RNS}$ is related to the unintegrated vertex operator $V_{\rm RNS} = cU_{\rm RNS}$ via $QU_{\rm RNS} = \partial V_{\rm RNS}$ [122]. This can be checked by recalling that $U_{\rm RNS} = \{\oint b, V_{\rm RNS}\}$ and $T = \{Q, b\}$, where (b,c) is the ghost system used to fix the reparametrization invariance of the worldsheet. The proof then follows from the Jacobi identity

$$QU_{\rm RNS} = [Q, \{\oint b, V_{\rm RNS}\}] = -[V_{\rm RNS}, \{Q, \oint b\}] - [\oint b, \{V_{\rm RNS}, Q\}] = \partial V_{\rm RNS} \tag{3.68}$$

because the cohomology condition requires $\{V_{\rm RNS}, Q\} = 0$ and the conformal weight $h{=}0$ of $V_{\rm RNS}$ implies that $[\oint T, V_{\rm RNS}] = \partial V_{\rm RNS}$ by (3.64).

While the pure spinor formalism does not feature any direct analogue of the (b,c) system[12] – that is why the forms of the unintegrated (3.62) and integrated (3.59) vertex operators are very different – the vertex operators V, U still satisfy the relation (3.68) of their analogues in the RNS formalism (see also [127]):

Proposition 5. *The massless integrated and unintegrated vertex operators (3.59) and (3.62) are related by*

$$QU = \partial V\,. \tag{3.69}$$

Proof. Using the OPEs (3.54) and (3.56) as well as the equations of motion for the linearized SYM superfields (2.14) we get

$$\begin{aligned}
Q(\partial\theta^\alpha A_\alpha) &= (\partial\lambda^\alpha)A_\alpha - \partial\theta^\alpha \lambda^\beta D_\beta A_\alpha\,, \\
Q(\Pi^m A_m) &= (\lambda\gamma^m\partial\theta)A_m + \Pi^m\lambda^\alpha(D_\alpha A_m)\,, \\
Q(d_\alpha W^\alpha) &= -(\lambda\gamma^m W)\Pi_m - d_\beta\lambda^\alpha D_\alpha W^\beta\,, \\
Q\Big(\frac{1}{2}N_{mn}F^{mn}\Big) &= \frac{1}{4}(\gamma_{mn}\lambda)^\alpha d_\alpha F^{mn} + \frac{1}{2}N_{mn}\lambda^\alpha D_\alpha F^{mn}\,.
\end{aligned} \tag{3.70}$$

Summing them up yields

$$\begin{aligned}
QU &= (\partial\lambda^\alpha)A_\alpha - \partial\theta^\beta\lambda^\alpha(D_\alpha A_\beta - \gamma^m_{\alpha\beta}A_m) + \lambda^\alpha\Pi^m(D_\alpha A_m - (\gamma_m W)_\alpha) \\
&\quad - \lambda^\alpha d_\beta\Big(D_\alpha W^\beta + \frac{1}{4}(\gamma^{mn})_\alpha{}^\beta F_{mn}\Big) + N_{mn}(\lambda\gamma^n\partial^m W) \\
&= (\partial\lambda^\alpha)A_\alpha + \lambda^\alpha\partial\theta^\beta D_\beta A_\alpha + \lambda^\alpha\Pi^m\partial_m A_\alpha\,,
\end{aligned} \tag{3.71}$$

where $N_{mn}(\lambda\gamma^n\partial^m W)$ vanishes due to a combination of the pure spinor condition $(\lambda\gamma^n)_\alpha(\lambda\gamma_n)_\beta = 0$ proven in (A.35) and the linearized equation of motion $\gamma^m_{\alpha\beta}\partial_m W^\beta = 0$,

$$N_{mn}(\lambda\gamma^n\partial^m W) = \frac{1}{2}(w\gamma_m\gamma_n\lambda)(\lambda\gamma^n\partial^m W) - (w\lambda)(\lambda\gamma^m\partial_m W) = 0\,. \tag{3.72}$$

Therefore, using (3.66) the BRST variation QU in (3.71) becomes

$$QU = (\partial\lambda^\alpha)A_\alpha + \lambda^\alpha(\partial\theta^\beta D_\beta A_\alpha + \Pi^m\partial_m A_\alpha) = (\partial\lambda^\alpha)A_\alpha + \lambda^\alpha\partial A_\alpha = \partial(\lambda A) = \partial V\,, \tag{3.73}$$

as we wanted to show. $\square$

[12] See [3, 123, 124, 125] for a composite b ghost in the non-minimal pure spinor formalism (see also [126]).

Corollary 1. *The integrated vertex operator $\int dzU(z)$ is BRST invariant up to surface terms.*

As we will see below, surface terms do not contribute to open- or closed-string amplitudes by the so-called canceled-propagator argument. The cancellation of surface terms is also used to demonstrate linearized gauge invariance of the massless vertex operators: under the linearized variations $\delta_\Omega A_\alpha = D_\alpha \Omega$ and $\delta_\Omega A_m = \partial_m \Omega$ of (2.15) with some gauge-scalar superfield Ω, the variation $\delta_\Omega V = \lambda^\alpha D_\alpha \Omega = Q\Omega$ vanishes in the cohomology, whereas

$$\delta U = \partial\theta^\alpha D_\alpha \Omega + \Pi^m \partial_m \Omega = \partial\Omega \tag{3.74}$$

reduces to vanishing surface terms after using the chain rule (3.67).

Superspace vertex operators for massive open-string states φ can be constructed by following the same principle: identifying a BRST-closed unintegrated vertex operator V_φ of conformal weight zero and then engineering its integrated counterpart U_φ of weight one such that $QU_\varphi = \partial V_\varphi$. By the conformal weight $h(e^{k\cdot X}) = -N$ at the N^{th} mass level, the combinations of $\Pi^m, d_\alpha, \ldots$ accompanying the plane waves accumulate more and more conformal weight and Lorentz indices at growing N. That the pure spinor cohomology contains all massive states of the superstring was shown in [128, 129] (see also [130]). The vertex operators at the first mass level are known in superspace from [131, 132, 133], and it is an open problem to pinpoint their explicit form at higher levels.

3.4.2. Scattering-amplitude prescription at genus zero

The prescription to compute n-point tree amplitudes of open-superstring states is given by the following correlation function of vertex operators on a disk worldsheet [1]

$$\mathcal{A}(P) = \int_{D(P)} dz_2\, dz_3 \ldots dz_{n-2}\, \langle\!\langle V_1(z_1) U_2(z_2) U_3(z_3) \ldots U_{n-2}(z_{n-2}) V_{n-1}(z_{n-1}) V_n(z_n) \rangle\!\rangle\,, \tag{3.75}$$

where $\langle\!\langle \ldots \rangle\!\rangle$ refers to the path integral over the variables in the pure spinor action (3.49) and Möbius invariance of the correlator was used to fix the insertion points of the three unintegrated vertex operators. We adopt the particularly convenient choices to parameterize the disk boundary by the real line $z_j \in \mathbb{R}$ and to place the unintegrated vertex operators at

$$(z_1, z_{n-1}, z_n) \to (0, 1, \infty)\,. \tag{3.76}$$

For a general n-point disk ordering characterized by a cyclic permutation $P := p_1 p_2 \ldots p_n$, the formal definition of the *integration domain P* in (3.75) reads

$$D(P) := \{(z_1, z_2, \ldots, z_n) \in \mathbb{R}^n \mid -\infty < z_{p_1} < z_{p_2} < \ldots < z_{p_n} < \infty\}\,. \tag{3.77}$$

For example, with the three fixed positions as in (3.76), a domain specified by the canonical ordering $P = 123\ldots n$ amounts to the disk ordering $0 < z_2 < z_3 < \cdots < z_{n-2} < 1$ of the integrated vertex operators at the boundary of the disk such that the cyclic ordering $1, 2, \ldots, n$ in P is preserved.

The prescription (3.75) is tailored to color-ordered amplitudes $\mathcal{A}(P)$ and can be color-dressed by weighting each disk ordering $D(P)$ with a trace of Chan–Paton factors t^{a_i} in the same cyclic ordering,

$$\mathcal{M}(1, 2, \ldots, n) = \sum_{\rho \in S_{n-1}} \mathcal{A}\big(1, \rho(2, 3, \ldots, n)\big) \mathrm{Tr}(t^{a_1} t^{a_{\rho(2)}} t^{a_{\rho(3)}} \ldots t^{a_{\rho(n)}})\,, \tag{3.78}$$

where ρ is in the set of permutations S_{n-1} of the $n{-}1$ legs $2, 3, \ldots, n$. Cyclicity of the trace propagates to the color-ordered amplitudes, $\mathcal{A}(1, 2, \ldots, n) = \mathcal{A}(2, 3, \ldots, n, 1)$, and the prescription (3.75) furthermore implies reflection properties $\mathcal{A}(1, 2, \ldots, n) = (-1)^n \mathcal{A}(n, \ldots, 2, 1)$.

As will be reviewed from several perspectives, the $\frac{1}{2}(n{-}1)!$ cyclically and reflection inequivalent permutations of $\mathcal{A}(1, 2, \ldots, n)$ are not linearly independent. First, the monodromy relations [134, 135] in section 7.3 only leave an $(n{-}3)!$-dimensional basis of disk orderings. Second, these relations among the disk amplitudes can be refined according to the multiple zeta values in the α'-expansions, see section 8.4: Parts of the string corrections obey field-theory relations of SYM tree amplitudes [136, 137] and others obey KK-like symmetries [138] related to permutations in the (inverse) Solomon descent algebra [139, 140, 141, 142, 143].

CFT calculation and zero modes. In order to evaluate the tree-level correlation function $\langle\!\langle \ldots \rangle\!\rangle$ in (3.75), one first integrates out the non-zero modes using the OPEs (3.54) to (3.56) to obtain its dependence on the positions z_i carried by the conformal-weight-one variables $[\partial\theta^\alpha(z), \Pi^m(z), d_\alpha(z), N^{mn}(z)]$. As explained in [12], this unambiguously determines the correlator as a function of the positions z_i on a genus-zero surface. After using the OPEs in this way, the correlation function (3.75) will still contain the zero modes of λ^α and θ^α, as they are variables of conformal weight zero with a single zero mode at genus zero [5]. These zero-mode correlators are denoted by $\langle \ldots \rangle$ (as opposed to the above double brackets $\langle\!\langle \ldots \rangle\!\rangle$ including the non-zero modes), and one needs an ad-hoc rule to integrate them. Using the shorthand

$$(\lambda^3\theta^5) := (\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{mnp}\theta)\,, \tag{3.79}$$

the only non-vanishing contributions in ten-dimensional[13] zero-mode correlators is proportional to (3.79) [1]

$$\langle(\lambda^3\theta^5)\rangle = 2880\,, \tag{3.80}$$

where the normalization 2880 was chosen in [145] in order to match the RNS tree-level amplitude conventions.

Proposition 6. *The combination $(\lambda^3\theta^5)$ is in the cohomology of the pure spinor BRST operator.*

Proof. It is BRST closed

$$\begin{aligned} Q(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{mnp}\theta) &= 3(\lambda\gamma^m\lambda)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{mnp}\theta) \\ &\quad - 2(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\lambda\gamma_{mnp}\theta) = 0\,. \end{aligned} \tag{3.81}$$

The first term vanishes by the pure spinor constraint (3.26), while the vanishing of the second term $(\lambda\gamma^m)_\alpha(\lambda\gamma^n)_\beta(\lambda\gamma^p)_\gamma(\lambda\gamma_{mnp}\theta) = 0$ can be seen by decomposing $\gamma^{mnp} = \gamma^m\gamma^{np} - \delta^{mn}\gamma^p + \delta^{mp}\gamma^n$ and using (A.35), $(\lambda\gamma^m)_\alpha(\lambda\gamma_m)_\beta = 0$. Moreover, (3.79) is not BRST exact,

$$(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{mnp}\theta) \neq Q\Omega(\lambda,\theta)\,, \tag{3.82}$$

because there is no scalar built from two λs and six θs. If there was a $\Omega(\lambda,\theta)$ such that $Q\Omega = (\lambda^3\theta^5)$, it would necessarily be a scalar function containing two λs and six θs since $Q\theta^\alpha = \lambda^\alpha$ and the ∂_m-admixture of $\lambda^\alpha D_\alpha$ can be dropped for functions of only λ and θ. The $SO(10)$ representation of two pure spinors λ^α is characterized by Dynkin labels (00002) while six antisymmetric thetas are represented by $(01020)\oplus(20100)$, see Appendix D. However, their product [146]

$$(00002) \otimes \Big((01020) \oplus (20100)\Big) = (00011) \oplus (00022) \oplus 2(00120) \oplus \cdots \tag{3.83}$$

has no scalar representation (00000). This shows that the putative BRST ancestor $\Omega(\lambda,\theta)$ in (3.82) cannot be constructed, finishing the proof. □

In the formulation of the prescription (3.75), we have chosen legs $1, n{-}1$ and n to be represented by an unintegrated vertex operator V_j at fixed locations z_1, z_{n-1}, z_n. It remains to verify that any other choice of three legs to appear in the unintegrated picture leads to the same result for each color-ordered amplitude.

Proposition 7. *The disk amplitude prescription (3.75) does not depend on which triplet $\{i,j,k\}$ of the external legs enters via unintegrated vertex operators $V_i(z_i)V_j(z_j)V_k(z_k)$ at fixed punctures z_i, z_j, z_k.*

Proof. Following the strategy of [145], it is sufficient to show that the representation $V_i, \int dz_{i+1}\, U_{i+1}$ of neighboring states i and $i{+}1$ can always be swapped to $\int dz_i\, U_i, V_{i+1}$, i.e.

$$\left\langle\!\!\left\langle V_1(0)\int_0^1 dz_2\, U_2(z_2)\prod_{j=3}^{n-2}\int_{z_{j-1}}^1 dz_j\; U_j(z_j)V_{n-1}(1)V_n(\infty)\right\rangle\!\!\right\rangle$$

[13] For the dimensional reduction of the condition (3.80) to $D = 4$, see [144].

$$= \left\langle\!\!\left\langle \int_{-\infty}^{0} dy\, U_1(y) V_2(0) \prod_{j=3}^{n-2} \int_{z_{j-1}}^{1} dz_j\, U_j(z_j) V_{n-1}(1) V_n(\infty) \right\rangle\!\!\right\rangle . \tag{3.84}$$

Since $QU_j(w) = \oint dz\, \lambda^\alpha d_\alpha(z) U_j(w) = \partial V_j(w)$, we can rewrite the left-hand side via

$$V_1(0)V_n(\infty) = \int_{-\infty}^{0} dy\, \partial V_1(y) V_n(\infty) = \int_{-\infty}^{0} dy\, Q\big(U_1(y)\big) V_n(\infty)\,. \tag{3.85}$$

In the first step, we have discarded $V_1(\infty)V_n(\infty)$ by the so-called "canceled-propagator argument" [12] which states that terms with colliding vertex operators $V_i(z)V_j(z)$ or $V_i(z)U_j(z)$ identically vanish. As a next step, we deform the integration contour of the BRST current $\lambda^\alpha d_\alpha$ such that it encircles all the vertex operators apart from U_1:

$$\begin{aligned}
&\left\langle\!\!\left\langle V_1(0) \int_{0}^{1} dz_2\, U_2(z_2) \prod_{j=3}^{n} \int_{z_{j-1}}^{1} dz_j\, U_j(z_j) V_{n-1}(1) V_n(\infty) \right\rangle\!\!\right\rangle \\
&\quad = - \left\langle\!\!\left\langle \int_{-\infty}^{0} dy\, U_1(y) \int_{0}^{1} dz_2\, Q\Big[U_2(z_2) \prod_{j=3}^{n} \int_{z_{j-1}}^{1} dz_j\, U_j(z_j) V_{n-1}(1) V_n(\infty) \Big] \right\rangle\!\!\right\rangle \\
&\quad = - \left\langle\!\!\left\langle \int_{-\infty}^{0} dy\, U_1(y) \int_{0}^{1} dz_2\, \partial V_2(z_2) \prod_{j=3}^{n} \int_{z_{j-1}}^{1} dz_j\, U_j(z_j) V_{n-1}(1) V_n(\infty) \right\rangle\!\!\right\rangle \\
&\quad = + \left\langle\!\!\left\langle \int_{-\infty}^{0} dy\, U_1(y) V_2(0) \prod_{j=3}^{n} \int_{z_{j-1}}^{1} dz_j\, U_j(z_j) V_{n-1}(1) V_n(\infty) \right\rangle\!\!\right\rangle
\end{aligned} \tag{3.86}$$

In passing to the third line, terms where Q acts on the U_j vertices with $3 \leq j \leq n-2$ were discarded due to the canceled-propagator argument: it forces both boundary terms of the $\int_{z_{j-1}}^{1} dz_j\, \partial V_j(z_j)$ integrals to vanish,

$$\ldots\, U_{j-1}(z_{j-1}) \big(V_j(1) \;-\; V_j(z_{j-1}) \big) \,\ldots\, V_{n-1}(1) \,\ldots = 0\,. \tag{3.87}$$

On the other hand, the integral over $QU_2 = \partial V_2$ contributes non-trivially to the last line of (3.86): while the upper integration limit $z_2 = 1$ cancels due to $V_2(1) \ldots V_{n-1}(1) = 0$, the lower one $z_2 = 0$ generically does not coincide with the position y of U_1, i.e. $U_1(y)V_2(0) \neq 0$. $\square$

As we have seen in (3.47), the worldsheet action in the pure spinor formalism is spacetime supersymmetric. This means that the OPEs among its worldsheet fields have the appropriate transformations under the generators $\mathcal{Q}_\alpha$ in (3.46) and will not violate supersymmetry. However, one still needs to show that the zero-mode integration rule (3.80) for the disk-amplitude prescription (3.75) preserves the supersymmetric nature of the formalism.

Proposition 8. *The disk-amplitude prescription (3.75) is supersymmetric [1].*

Proof. We will show that the supersymmetry variation of the amplitude under $\delta\theta^\alpha = \epsilon^\alpha$ vanishes, i.e. that $\delta\mathcal{A}(1, \ldots, n) = 0$. Note that the only possibility of getting a non-vanishing result after the supersymmetry transformation $\delta\theta^\alpha = \epsilon^\alpha$ is if the correlator in the amplitude (3.75) contains a term of the form

$$\mathcal{A}(P) = \int\limits_{D(P)} dz_2 \cdots dz_{n-2}\, \langle (\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{mnp}\theta)(\theta^\alpha \Phi_\alpha(z, e, \chi, k) + \ldots)\rangle \tag{3.88}$$

for some $\Phi_\alpha(z, e, \chi, k)$ depending on the worldsheet positions z_i of all open-string vertex operators as well as polarizations e^m, χ^α and momenta k^m. The zero-mode integration (3.80) would then imply the supersymmetry variation to be $\delta\mathcal{A} = \int dz_2 \cdots \int dz_{n-2} \epsilon^\alpha \Phi_\alpha(z, e, \chi, k)$. To see why this variation must be zero, note that the result of the OPE calculation in the amplitude prescription (3.75) leads to an amplitude that can be written as $\int dz_2 \cdots \int dz_{n-2} \langle \lambda^\alpha \lambda^\beta \lambda^\gamma f_{\alpha\beta\gamma}(\theta, e, \chi, k)\rangle$ for some function f depending on the zero modes of θ and on the momenta and polarizations of the open-string states. However, the amplitude must be BRST invariant, so its correlator must be such that

$$\int_{D(P)} dz_2 \cdots dz_{n-2}\, \lambda^\alpha \lambda^\beta \lambda^\gamma \lambda^\delta D_\delta f_{\alpha\beta\gamma}(\theta, e, \chi, k) = 0\,. \tag{3.89}$$

Using the function f following from (3.88) and plugging it into (3.89) we conclude

$$\int_{D(P)} dz_2 \cdots dz_{n-2}\, \lambda^\alpha \lambda^\beta \lambda^\gamma \lambda^\delta \Phi_\delta(z, e, \chi, k) = 0\,. \tag{3.90}$$

This vanishing is only possible if Φ_δ is a total worldsheet derivative, $\Phi_\delta = \partial(\ldots)$, implying that the supersymmetry variation of the amplitude vanishes after integration, $\delta\mathcal{A} = 0$. □

3.4.3. The field-theory limit

Disk amplitudes of massless open-superstring states reduce to n-point tree-level amplitudes among the supermultiplet of ten-dimensional SYM [79] when the dimensionless combinations $\alpha' k_i \cdot k_j$ are taken to be small [147, 148, 149, 150]. We will refer to this low-energy regime as the field-theory limit and informally write $\alpha' \to 0$. Since the scattering energies are small in comparison to the inverse string-length scale, this limit can also be thought of as shrinking the string to a point particle.

Throughout this review, SYM tree-level amplitudes in the field-theory limit will be denoted by $A(1, \ldots, n)$ when they contain all states in the supermultiplet, i.e.

$$A(1, 2, \ldots, n) = \lim_{\alpha' \to 0} \mathcal{A}(1, 2, \ldots, n)\,. \tag{3.91}$$

As will be illustrated in section 3.4.5 below, the superstring-amplitude prescription (3.75) yields formal sums of component amplitudes with external bosons and fermions since the pure spinor formalism is manifestly supersymmetric. When restricted to bosonic external states, the field-theory tree-level amplitudes will be denoted by $A^{\rm YM}(1, 2, \ldots, n)$. The construction of SYM tree-level amplitudes $A(1, 2, \ldots, n)$ using pure spinor cohomology methods will be described in section 5.2, and the alternative derivation from the $\alpha' \to 0$ limit of the superstring amplitude will be reviewed in section 6.5 (see also (7.42)).

3.4.4. Pure spinor superspace

Superfield expressions containing the zero modes of three pure spinors define *pure spinor superspace* [1, 151]

$$\lambda^\alpha \lambda^\beta \lambda^\gamma f_{\alpha\beta\gamma}(\theta, e, \chi, k)\,, \tag{3.92}$$

where $f_{\alpha\beta\gamma}(\theta, e, \chi, k)$ represents a function containing zero modes of θ^α as well as gluon and gluino polarizations and momenta. It is easy to see that such expressions necessarily arise from the amplitude prescription (3.75) after integrating the non-zero modes via OPEs as outlined above. For example, the massless three-point disk amplitude $\mathcal{A}(1, 2, 3) = \langle V_1 V_2 V_3\rangle$ leads to the pure spinor superspace expression $f_{\alpha\beta\gamma}(\theta, e, \chi, k) = A^1_\alpha(\theta) A^2_\beta(\theta) A^3_\gamma(\theta)$, see (2.17) for the θ-expansion of the SYM superfields $A^i_\alpha(\theta)$.

As seen above, the final step in the computation of string disk amplitudes boils down to integrating out the zero modes of three pure spinors and five θs using the prescription (3.80). These zero-mode integrations result in the component expansion of the amplitude under consideration written as a scalar function of polarizations and momenta. Let us write the most general form of a pure spinor superspace expression containing five θs as

$$\lambda^\alpha \lambda^\beta \lambda^\gamma \theta^{\delta_1} \theta^{\delta_2} \theta^{\delta_3} \theta^{\delta_4} \theta^{\delta_5} f_{\alpha\beta\gamma|\delta_1\delta_2\delta_3\delta_4\delta_5}(e, \chi, k)\,, \tag{3.93}$$

where (e, χ, k) indicates a dependence on gluon and gluino polarizations as well as their momenta. We need to extract the Lorentz contractions of polarizations and momenta from pure spinor superspace expressions like (3.93). This can be done on the basis of the group-theory statement that there is only one scalar built from three pure spinors λ^α and five unconstrained θs.

Lemma 2. *There is only one scalar representation in the decomposition of three pure spinors and five unconstrained fermionic Weyl spinors of SO(10).*

Proof. This follows from the tensor product of $SO(10)$ representations (00003) corresponding to three pure spinors $\lambda^\alpha\lambda^\beta\lambda^\gamma$ and $(00030)\oplus(11010)$ corresponding to $\theta^{\delta_1}\theta^{\delta_2}\theta^{\delta_3}\theta^{\delta_4}\theta^{\delta_5}$ [146]

$$(00003)\otimes\big((00030)\oplus(11010)\big) = 1\times(00000)\oplus 2\times(00011)\oplus\cdots\,, \tag{3.94}$$

where the scalar (00000) occurs with multiplicity one. □

The above Lemma means that any expression containing three λs and five θs can be reduced to its scalar component $(\lambda^3\theta^5)$ with proportionality constants given entirely in terms of Kronecker deltas, gamma matrices and Levi–Civita ϵ_{10} tensors. This will be exploited in Appendix E to build up a catalog of various pure spinor correlators.

For example, suppose we have the pure spinor superspace expression $\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{abc}\theta)\rangle$ with free vector indices m,n,p and a,b,c. In order to use the rule (3.80) one needs to extract its scalar component $(\lambda^3\theta^5)$. Because we know from the Lemma above that there is only one scalar representation in this product, this is easily done using symmetry arguments alone. The result is

$$\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{abc}\theta)\rangle = 24\delta^{mnp}_{abc}\,, \tag{3.95}$$

where δ^{mnp}_{abc} is the generalized Kronecker delta (A.9). To see this, observe that the right-hand side is the unique term that is antisymmetric in both $[mnp]$ and $[abc]$, as required by the symmetries of the left-hand side. The proportionality constant can be fixed by contracting the vectorial indices on both sides with $\delta^a_m\delta^b_n\delta^c_p$: On the left-hand side we get $\langle(\lambda^3\theta^5)\rangle$, while the right-hand side reduces to $24\times 120 = 2880$ (using $\delta^{mnp}_{mnp} = \binom{10}{3} = 120$, see (A.11)). Therefore we recover the normalization (3.80), and the Lemma guarantees that this is the correct tensor.

For another example, consider $\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\chi\gamma_n\theta)(\psi\gamma_p\theta)\rangle$, for two arbitrary Weyl spinors χ and ψ. Based on the Fierz identity (A.18), $\theta^\alpha\theta^\beta = \frac{1}{96}\gamma^{\alpha\beta}_{rst}(\theta\gamma^{rst}\theta)$, we obtain

$$\begin{aligned}\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_n\chi)(\theta\gamma_p\psi)\rangle &= -\frac{1}{96}(\chi\gamma_n\gamma^{rst}\gamma_p\psi)\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{rst}\theta)\rangle\\ &= -\frac{1}{4}(\chi\gamma_n\gamma^{mnp}\gamma_p\psi) = 18(\chi\gamma^m\psi)\,,\end{aligned} \tag{3.96}$$

where we used (3.95) and $\gamma_n\gamma^{mnp}\gamma_p = -72\gamma^m$. For an alternative derivation, see [145]. And for a more in-depth excursion on the evaluation of pure spinor superspace zero-mode correlators, see Appendix E.

3.4.5. Component expansion from pure spinor superspace

As an illustration of the above steps, the supersymmetric three-point tree amplitude following from (3.75) is given by

$$\mathcal{A}(1,2,3) = \langle(\lambda A^1)(\lambda A^2)(\lambda A^3)\rangle = \mathcal{A}(1_b,2_b,3_b)+\mathcal{A}(1_b,2_f,3_f)+\mathcal{A}(1_f,2_b,3_f)+\mathcal{A}(1_f,2_f,3_b)\,, \tag{3.97}$$

where the subscripts b or f refer to the bosonic or fermionic component polarizations, corresponding to the gluon or gluino at the massless level of the open superstring.

Evaluating the explicit component expansion for the three-gluon amplitude is a matter of plugging in the θ-expansions (2.17) in Harnad–Shnider gauge and selecting the components with five θs which contain the gluon fields. Doing this for the spinorial superpotential A_α, the only terms in the bosonic θ-expansion that can contribute are

$$A_\alpha(X,\theta)\to\Big\{\frac{1}{2}e_m(\gamma^m\theta)_\alpha - \frac{1}{32}f_{mn}(\gamma_p\theta)_\alpha(\theta\gamma^{mnp}\theta)\Big\}e^{k\cdot X}\,. \tag{3.98}$$

In combination with the minimum of θ^2 from the two other vertex operators, the term containing θ^5 in A_α leads to a superspace expression proportional to $\lambda^3\theta^{p\geq 7}$ which is annihilated by the pure spinor bracket rule (3.80). This results in the three-gluon component amplitude

$$\mathcal{A}(1_b,2_b,3_b) = -\frac{1}{128}e^m_1 f^{pq}_2 e^n_3\langle(\lambda\gamma^m\theta)(\lambda\gamma^r\theta)(\lambda\gamma^n\theta)(\theta\gamma_{pqr}\theta)\rangle + \mathrm{cyc}(1,2,3)\,. \tag{3.99}$$

As discussed above, symmetry arguments and the normalization condition (3.80) fix all pure spinor correlators and we find ($\delta^r_r = 10$)

$$\langle (\lambda\gamma^m\theta)(\lambda\gamma^r\theta)(\lambda\gamma^n\theta)(\theta\gamma_{pqr}\theta)\rangle = 24\delta^{mrn}_{pqr} = -64\delta^{mn}_{pq} \,. \tag{3.100}$$

So the three-gluon amplitude (3.99) is given by

$$\begin{aligned}\mathcal{A}(1_b,2_b,3_b) &= \frac{1}{2}e_1^m f_2^{mn} e_3^n + \text{cyc}(1,2,3) \\ &= (e_1\cdot k_2)(e_2\cdot e_3) + \text{cyc}(1,2,3)\,,\end{aligned} \tag{3.101}$$

where we have applied transversality $e_i \cdot k_i = 0$ and momentum conservation $k_1+k_2+k_3 = 0$ in passing to the second line. To obtain the amplitude with one gluon and two gluinos distributed as $(1_b, 2_f, 3_f)$, we use the fermionic component expansion $A_\alpha(\theta) \to \frac{1}{3}(\theta\gamma_m)_\alpha(\theta\gamma^m\chi)e^{k\cdot X}$ in (2.17) for legs 2, 3 to get

$$\mathcal{A}(1_b,2_f,3_f) = \frac{1}{18}e_1^m \langle (\lambda\gamma_m\theta)(\lambda\gamma_n\theta)(\lambda\gamma_p\theta)(\theta\gamma^n\chi_2)(\theta\gamma^p\chi_3)\rangle = e_1^m(\chi_2\gamma_m\chi_3)\,, \tag{3.102}$$

where we used the pure spinor correlator (3.96) and picked the unique term $\sim \theta^1$ from A^1_α which is compatible with two θs from both A^2_α and A^3_α. Assembling all the components in the three-point amplitude (3.97) yields

$$\mathcal{A}(1,2,3) = \frac{1}{2}e_1^m f_2^{mn} e_3^n + e_1^m(\chi_2\gamma_m\chi_3) + \text{cyc}(1,2,3)\,. \tag{3.103}$$

Given the systematic nature of the above procedure, an implementation using `FORM` [152] has been written which performs these expansions automatically [153] (see also [154]).

In contrast to the three-gluon amplitude of the open bosonic string, the three-point superstring amplitude (3.103) is independent on α' and therefore coincides with the SYM amplitude,

$$A(1,2,3) = \mathcal{A}(1,2,3) = \frac{1}{2}e_1^m f_2^{mn} e_3^n + e_1^m(\chi_2\gamma_m\chi_3) + \text{cyc}(1,2,3)\,. \tag{3.104}$$

For both color-ordered and color-dressed amplitudes of ten-dimensional SYM,

$$A(1,2,\ldots,n) = \lim_{\alpha'\to 0}\mathcal{A}(1,2,\ldots,n)\,, \qquad M(1,2,\ldots,n) = \lim_{\alpha'\to 0}\mathcal{M}(1,2,\ldots,n) \tag{3.105}$$

we will use non-calligraphic letters to distinguish them from the analogous superstring quantities, see (3.78) for the color-dressed open-string amplitude.

3.4.6. Preview of higher-point SYM amplitudes

In the same way as the single term $V_1V_2V_3$ in the superspace expression for $\mathcal{A}(1,2,3)$ produces the six terms in (3.104) upon component expansion, higher-point string and SYM amplitudes take a particularly compact form in pure spinor superspace. For instance, we will see later that the six-point SYM tree-level amplitude can be written in pure spinor superspace as

$$A(1,2,\ldots,6) = \frac{1}{3}\frac{\langle V_{12}V_{34}V_{56}\rangle}{s_{12}s_{34}s_{56}} + \frac{1}{2}\Big\langle\Big(\frac{V_{123}}{s_{12}s_{123}} + \frac{V_{321}}{s_{23}s_{123}}\Big)\Big(\frac{V_{45}V_6}{s_{45}} + \frac{V_4V_{56}}{s_{56}}\Big)\Big\rangle + \text{cyc}(1,2,\ldots,6)\,, \tag{3.106}$$

in terms of multiparticle vertex operators V_P subject to beautiful combinatorial properties that will be introduced below. The expression (3.106) can be checked to be gauge invariant and supersymmetric in a couple of lines with pen and paper. Moreover, it evades BRST-exactness for purely kinematic reasons and lines up with a recursive structure of n-point SYM tree amplitudes in pure spinor superspace. Here and below, our normalization conventions for Mandelstam invariants for massless particles are

$$s_{12} = k_1\cdot k_2 = \frac{1}{2}(k_1{+}k_2)^2\,, \qquad s_{12\ldots p} = \sum_{1\le i<j}^{p} s_{ij} = \frac{1}{2}(k_1{+}k_2{+}\ldots{+}k_p)^2\,. \tag{3.107}$$

Already for the purely bosonic terms, the component expansion of (3.106) in terms of single-particle polarizations and momenta produces more than 6700 terms. Still, we will see later that the complete component expansion of (3.106) can be arranged in the compact form

$$
\begin{aligned}
A(1,2,\ldots,6) &= \frac{1}{2}\mathfrak{e}_{12}^m \mathfrak{f}_{34}^{mn}\mathfrak{e}_{56}^n + \frac{1}{4}\big[\mathfrak{e}_{123}^m \mathfrak{f}_{45}^{mn}\mathfrak{e}_6^n + \mathfrak{e}_{45}^m \mathfrak{f}_6^{mn}\mathfrak{e}_{123}^n + \mathfrak{e}_6^m \mathfrak{f}_{123}^{mn}\mathfrak{e}_{45}^n + (4\leftrightarrow 6)\big] \\
&+ (\mathcal{X}_{12}\gamma_m \mathcal{X}_{34})\mathfrak{e}_{56}^m + \frac{1}{2}\big[(\mathcal{X}_{123}\gamma_m\mathcal{X}_{45})\mathfrak{e}_6^m + (\mathcal{X}_{45}\gamma_m\mathcal{X}_6)\mathfrak{e}_{123}^m + (\mathcal{X}_6\gamma_m\mathcal{X}_{123})\mathfrak{e}_{45}^m + (4\leftrightarrow 6)\big] + \mathrm{cyc}(1,2,\ldots,6)
\end{aligned}
\tag{3.108}
$$

in terms of recursively defined multiparticle Berends–Giele polarizations $\mathfrak{e}_P^m$, $\mathcal{X}_P^\alpha$ of (4.117) and field strengths $\mathfrak{f}_P^{mn}$ of (4.120) instead of single-particle polarizations and momenta. Similar objects also drive compact representations of supersymmetric loop amplitudes in string and field-theory, and they are excellently suited for numerical computations [155].

The ten-dimensional polarization vectors in the bosonic components of expressions in pure spinor superspace can be straightforwardly dimensionally reduced. In this way, one obtains scalar and gluon amplitudes in maximally supersymmetric SYM in lower dimensions, say $\mathcal{N}=4$ in four dimensions. Upon insertion of spinor-helicity expressions, (3.106) and (3.108) then reproduce all the six-point MHV and NMHV components of $\mathcal{N}=4$ SYM at tree level. Hence, pure spinor superspace elegantly unifies all the MHV, NMHV, N^kMHV components upon reduction to four dimensions and captures all the different functional forms of color-ordered amplitudes with particles of alike helicities in neighboring or non-neighboring legs. The number of terms in the pure spinor superspace representations of n-point SYM amplitude grows moderately with n thanks to the multiparticle formalism to be introduced below.

4. Multiparticle SYM in ten dimensions

OPEs among massless vertex operators of the pure spinor superstring feature rich patterns which led to a systematic definition of *multiparticle* superfields of ten-dimensional SYM in [83, 84, 85], in both local and non-local forms. These multiparticle superfields encompass arbitrary numbers of single-particle gluon and gluino states and can be constructed independently of their OPE origins using field-theory methods, in particular Berends–Giele recursion relations [27] and perturbiner methods [23, 24, 25, 26]. Over time, the definition of multiparticle superfields led to an elegant symbiosis of an ever-increasing number of related topics: their local version is at the heart of the local BCJ-satisfying numerators, and their non-local version is used to relate the BCJ properties of the amplitudes as originating from standard finite gauge transformations. In addition, multiparticle superfields appear in connection with planar binary trees leading to a combinatorial underpinning of the KLT map [156, 157] as well as the closely related S bracket [83] and contact-term map [85]. Ultimately, the use of multiparticle superfields simplifies the construction of expressions for scattering amplitudes of both SYM field theory and superstrings.

4.1. Local superfields

The definition of local multiparticle superfields is inspired by OPE calculations of massless vertex operators (3.59) and (3.62) in the pure spinor formalism. These multiparticle superfields generalize in a natural way the single-particle description of ten-dimensional SYM theory reviewed in section 2.2. For each of the standard four types of superfields $A_\alpha^i(X,\theta)$, $A_m^i(X,\theta)$, $W_i^\alpha(X,\theta)$ and $F_{mn}^i(X,\theta)$, the single-particle label i is generalized to labels for multiple particles, characterized either by words P such as $P=1234$ or by nested Lie brackets as $[[[1,2],[3,4]],5]$. As such, it will be convenient to refer to their multiparticle counterparts collectively in a set K_P

$$
K_P \in \{A_\alpha^P(X,\theta),\ A_P^m(X,\theta),\ W_P^\alpha(X,\theta),\ F_P^{mn}(X,\theta)\}\,,
\tag{4.1}
$$

with obvious extension for $K_{[P,Q]}$ where P and Q can themselves be nested brackets.

Calculations of superstring disk correlators revealed that there is a rich set of properties obeyed by the multiparticle superfields, reflected by the symmetry properties of their multiparticle labels. The symmetries in turn are attained by various gauge transformations of the individual single-particle superfields and give rise to different definitions of multiparticle superfields, all related by gauge transformations. Of special importance is the gauge transformation leading to Jacobi identities within the nested brackets characterizing the multiparticle state. We will see in section 7.1 that this gauge leads to the color-kinematics duality of Bern, Carrasco and Johansson [29]. At the superfield level, this translates to non-linear gauge transformations which act on multiparticle superfields defined recursively in the so-called Lorenz gauge.

The construction of the two-particle superfields is inspired by string-theory methods in the following way. The insertion of a gauge-multiplet state on the boundary of an open-string worldsheet is described by the pure spinor integrated vertex operator (3.59),

$$U_i := \partial\theta^\alpha A^i_\alpha + \Pi^m A^i_m + d_\alpha W^\alpha_i + \tfrac{1}{2} N^{mn} F^i_{mn} \, . \tag{4.2}$$

In the computation of disk amplitudes with the prescription (3.75), the worldsheet fields of conformal weight one $[\partial\theta^\alpha, \Pi^m, d_\alpha, N^{mn}]$ contracting linearized superfields K_i with particle label i approach other linearized vertex operators describing other particle labels. This is captured by OPE singularities (3.54) to (3.56) and lead to composite superfields at their residues, dubbed multiparticle superfields.

The first example of a multiparticle superfield appears in [158] as the OPE residue of two massless vertex operators, an integrated U_2 describing SYM states with particle label 2 and an unintegrated V_1 with particle label 1

$$V_1(z_1)U_2(z_2) \sim z_{21}^{-k_1\cdot k_2} \frac{L_{21}(z_1)}{z_{21}} \, , \qquad z_{ij} := z_i - z_j \, . \tag{4.3}$$

In order to attain open-string conventions, we dropped the $\overline{z}_{ij}$ dependence of the factor in the OPE of plane waves $e^{k_1\cdot X(z_1)} e^{k_2\cdot X(z_2)} \sim |z_{21}|^{-2k_1\cdot k_2} e^{k_{12}\cdot X(z_2)}$ with $|z_{21}|^{-2k_1\cdot k_2} = z_{21}^{-k_1\cdot k_2} \overline{z}_{21}^{-k_1\cdot k_2}$. The corresponding n-point correlation function on the sphere

$$\langle\!\langle \prod_{j=1}^n e^{k_j\cdot X(z_j)} \rangle\!\rangle = \prod_{1\le i<j}^n |z_{ij}|^{-2k_i\cdot k_j} \tag{4.4}$$

will be referred to as the *Koba–Nielsen factor*.

The superfield structure of the OPE is captured by

$$L_{21} = -A^m_1 (\lambda\gamma_m W_2) - V_1(k_1\cdot A_2) + Q(A_1 W_2) \tag{4.5}$$

which has a simple BRST variation

$$Q L_{21} = (k_1\cdot k_2) V_1 V_2 \, , \tag{4.6}$$

where $Q = \lambda^\alpha D_\alpha$ denotes the action of the BRST operator (3.52) of the pure spinor formalism on superfields independent of $\partial^{k\ge 1}\theta$. The BRST-exact term $Q(A_1 W_2)$ in (4.5) does not contribute to the variation (4.6) and will be removed in passing from L_{21} to the two-particle superfield V_{12} below. Proceeding recursively and defining higher-rank superfield building blocks [159]

$$L_{2131\ldots(p-1)1}(z_1) U_p(z_p) \sim z_{p1}^{-(k_1+k_2+\ldots k_{p-1})\cdot k_p} \frac{L_{2131\ldots(p-1)1p1}(z_1)}{z_{p1}} \tag{4.7}$$

yields BRST transformations such as

$$\begin{aligned} QL_{2131} &= (k_{12}\cdot k_3) L_{21} V_3 + (k_1\cdot k_2)(L_{31}V_2 + V_1 L_{32}) \, , \\ QL_{213141} &= (k_{123}\cdot k_4) L_{2131} V_4 + (k_{12}\cdot k_3)(L_{21}L_{43} + L_{2141}V_3) \\ &\quad + (k_1\cdot k_2)(L_{3141}V_2 + L_{31}L_{42} + L_{41}L_{32} + V_1 L_{3242}) \end{aligned} \tag{4.8}$$

with a suggestive recursive structure. The collection of $L_{2131\ldots p1}$ is said to be *BRST covariant* since their Q-variation is expressible in terms of products of lower-rank building blocks (with V_j as their rank-one version) along with factors of $k_i\cdot k_j$. Here and below, we are using the notation

$$k_{12\ldots p} = k_1 + k_2 + \ldots + k_p \tag{4.9}$$

for multiparticle momenta. However, a major shortcoming of the OPE residues $L_{2131\ldots p1}$ defined above is their lack of symmetry under exchange of labels $1, 2, 3, \ldots, p$. Luckily, the obstructions to having symmetry properties conspire to BRST-exact terms and can be removed by redefinitions that do not affect the desired amplitudes [160, 21, 83]. As a simple example of this phenomenon, the symmetric part of the rank-two OPE is BRST exact

$$L_{21} + L_{12} = Q\Big\{(A_1 W_2) + (A_2 W_1) - (A_1 \cdot A_2)\Big\}\,. \tag{4.10}$$

The spinor and vector superfields A_α and A^m of $D = 10$ SYM can be distinguished by identifying the superfields that they contract – above these are W^α or A_m (and we only use the $\cdot$ for vector-index contractions, i.e. not for spinor indices). Using the BRST transformation properties of $L_{2131\ldots}$, these BRST-exact admixtures have been identified in [21] up to rank five, and their removal leads to a redefinition of the OPE residues that satisfy generalized Jacobi identities (see section 4.1.5 for their definition). The outcome of this removal procedure is an improved family of multiparticle superfields $V_{123\ldots p}$ dubbed BRST building blocks.

This approach was streamlined and further developed in [83] where all multiplicity-two superfields K_{12} were extracted from the OPE between two integrated vertices as the coefficients of the conformal fields in the OPE, following earlier calculations in [109]

$$\begin{aligned}
U_1(z_1)U_2(z_2) \sim z_{12}^{-k_1\cdot k_2-1}\Big(&\partial\theta^\alpha\big[(k_1\cdot A_2)A^1_\alpha - (k_2\cdot A_1)A^2_\alpha + D_\alpha A^2_\beta W_1^\beta - D_\alpha A^1_\beta W_2^\beta\big]\\
&+ \Pi^m\big[(k_1\cdot A_2)A^1_m - (k_2\cdot A_1)A^2_m + k^2_m(A_2W_1) - k^1_m(A_1W_2) - (W_1\gamma_m W_2)\big]\\
&+ d_\alpha\big[(k_1\cdot A_2)W_1^\alpha - (k_2\cdot A_1)W_2^\alpha + \tfrac{1}{4}(\gamma^{mn}W_1)^\alpha F^2_{mn} - \tfrac{1}{4}(\gamma^{mn}W_2)^\alpha F^1_{mn}\big]\\
&+ \frac{1}{2}N^{mn}\big[(k_1\cdot A_2)F^1_{mn} - (k_2\cdot A_1)F^2_{mn} - 2k^{12}_m(W_1\gamma_n W_2) + 2F^1_{ma}F^2_{na}\big]\Big)\\
&+ (1+k_1\cdot k_2)z_{12}^{-k_1\cdot k_2-2}\big[(A_1W_2) + (A_2W_1) - (A_1\cdot A_2)\big]\,.
\end{aligned} \tag{4.11}$$

Using the relation $\partial K = \partial\theta^\alpha D_\alpha K + \Pi^m k_m K$ for superfields K independent of non-zero modes $\partial\theta^\alpha$ and λ^α, cf. (3.67), we can absorb the most singular piece $\sim z_{12}^{-k_1\cdot k_2-2}$ into total z_1, z_2 derivatives and rewrite

$$\begin{aligned}
U_1(z_1)U_2(z_2) \to &-z_{12}^{-k_1\cdot k_2-1}\Big(\partial\theta^\alpha A^{12}_\alpha + \Pi^m A^{12}_m + d_\alpha W^\alpha_{12} + \frac{1}{2}N^{mn}F^{12}_{mn}\Big) \\
&+ \partial_1\Big(z_{12}^{-k_1\cdot k_2-1}\big[\tfrac{1}{2}(A_1\cdot A_2) - (A_1W_2)\big]\Big) - \partial_2\Big(z_{12}^{-k_1\cdot k_2-1}\big[\tfrac{1}{2}(A_1\cdot A_2) - (A_2W_1)\big]\Big)\,.
\end{aligned} \tag{4.12}$$

Straightforward calculations using the linearized SYM equations of motion (2.14) yield the following multiplicity-two superfields,

$$\begin{aligned}
A^{12}_\alpha &= \tfrac{1}{2}\big[A^2_\alpha(k_2\cdot A_1) + A^m_2(\gamma_m W_1)_\alpha - (1\leftrightarrow 2)\big]\,,\\
A^m_{12} &= \tfrac{1}{2}\big[A^m_2(k_2\cdot A_1) + A^1_p F^{pm}_2 + (W_1\gamma^m W_2) - (1\leftrightarrow 2)\big]\,,\\
W^\alpha_{12} &= \tfrac{1}{4}(\gamma_{mn}W_2)^\alpha F^{mn}_1 + W^\alpha_2(k_2\cdot A_1) - (1\leftrightarrow 2)\,,\\
F^{mn}_{12} &= F^{mn}_2(k_2\cdot A_1) + \tfrac{1}{2}F_2^{[m}{}_p F_1^{n]p} + k_1^{[m}(W_1\gamma^{n]}W_2) - (1\leftrightarrow 2)\,,
\end{aligned} \tag{4.13}$$

where we reiterate our conventions $F_2^{[m}{}_p F_1^{n]p} = F_2^m{}_p F_1^{np} - F_2^n{}_p F_1^{mp}$ for antisymmetrization brackets. An interesting observation is that the two-particle field strength F^{mn}_{12} admits a more conventional form

$$F^{mn}_{12} = k^m_{12}A^n_{12} - k^n_{12}A^m_{12} - (k_1\cdot k_2)(A^m_1 A^n_2 - A^n_1 A^m_2) \tag{4.14}$$

with a non-linear extension as compared to the linearized field-strength superfield $F^{mn}_i = k^m_i A^n_i - k^n_i A^m_i$. More importantly, the covariant nature of the BRST transformations observed in (4.8) generalizes to the whole set of superfields in K_{12}. In fact,

$$\begin{aligned}
D_\alpha A^{12}_\beta + D_\beta A^{12}_\alpha &= \gamma^m_{\alpha\beta}A^{12}_m + (k_1\cdot k_2)(A^1_\alpha A^2_\beta + A^1_\beta A^2_\alpha)\,,\\
D_\alpha A^m_{12} &= \gamma^m_{\alpha\beta}W^\beta_{12} + k^m_{12}A^{12}_\alpha + (k_1\cdot k_2)(A^1_\alpha A^m_2 - A^2_\alpha A^m_1)\,,
\end{aligned} \tag{4.15}$$

$$D_\alpha W_{12}^\beta = \tfrac{1}{4}(\gamma_{mn})_\alpha{}^\beta F_{12}^{mn} + (k_1\cdot k_2)(A_\alpha^1 W_2^\beta - A_\alpha^2 W_1^\beta)\,,$$
$$D_\alpha F_{12}^{mn} = k_{12}^{[m}(\gamma^{n]}W_{12})_\alpha + (k_1\cdot k_2)\big[A_\alpha^1 F_2^{mn} + A_1^{[n}(\gamma^{m]}W_2)_\alpha - (1\leftrightarrow 2)\big]\,.$$

This set of superspace derivatives for the multiparticle superfields K_{12} mimic the single-particle case (2.14). The difference involves contact-term corrections proportional to the Mandelstam invariant $k_1\cdot k_2 = s_{12}$. We will see below that these contact terms admit a generalization and can be compactly described by the so-called contact-term map acting on *Lie polynomials*. In this review a "Lie polynomial" is understood to be any linear combination of terms that can be written in terms of nested commutators. For a more mathematical definition, see [143].

The definition of multiplicity-two superfields can be formalized by

$$\begin{aligned} U_{12}(z_2) &:= -\oint dz_1\,(z_1-z_2)^{k_1\cdot k_2}U_1(z_1)U_2(z_2) \\ &= \partial\theta^\alpha A_\alpha^{12} + \Pi^m A_m^{12} + d_\alpha W_{12}^\alpha + \tfrac{1}{2}N^{mn}F_{mn}^{12}\,, \end{aligned} \tag{4.16}$$

where the contour integral extracts the singular behavior of the approaching vertex operators as $z_1\to z_2$ and annihilates the total derivatives w.r.t. z_1, z_2 spelled out in (4.12). As we shall see below, these singularities on the worldsheet translate into propagators $k_{12}^{-2} = 2s_{12}^{-1}$ of the gauge-theory amplitude after performing the field-theory limit. In other words, OPEs in string theory govern the pole structure of tree-level subdiagrams in SYM field theory obtained from the point-particle limit. In addition to the multiplicity-two integrated vertex U_{12}, we define the multiplicity-two version of the unintegrated vertex as

$$V_{12} = \lambda^\alpha A_\alpha^{12}\,. \tag{4.17}$$

The two-particle equations of motion (4.15) imply that the single-particle relations $QV_1 = 0$ and $QU_1 = \partial V_1$ generalize as follows at multiplicity two [83]:

$$\begin{aligned} QV_{12} &= (k_1\cdot k_2)V_1V_2 \\ QU_{12} &= \partial V_{12} + (k_1\cdot k_2)(V_1U_2 - V_2U_1)\,. \end{aligned} \tag{4.18}$$

Note that the total derivatives in the last line of (4.12) are in one-to-one correspondence to the BRST-exact difference

$$V_{12} = L_{21} + Q\Big\{\tfrac{1}{2}(A_1\cdot A_2) - (A_1W_2)\Big\}\,. \tag{4.19}$$

The higher-multiplicity extensions of V_{12} and U_{12} to be constructed below also enjoy covariant BRST transformations among multiparticle vertex operators such as

$$\begin{aligned} QV_{123} &= (k_1\cdot k_2)\big[V_1V_{23} + V_{13}V_2\big] + (k_{12}\cdot k_3)V_{12}V_3\,, \\ QU_{123} &= \partial V_{123} + (k_1\cdot k_2)(V_1U_{23} - V_{23}U_1 + V_{13}U_2 + V_2U_{13}) + (k_{12}\cdot k_3)(V_{12}U_3 - V_3U_{12})\,, \end{aligned} \tag{4.20}$$

whose systematics are accurately described by the contact-term map in the subsequent section.

4.1.1. The contact-term map

We will see in the discussion below that many formulae simplify if we have a general formula to associate contact terms $\sum(k_R\cdot k_S)(\ldots)$ with general nested brackets of the form $[P,Q]$. The algorithm to do this is called the *contact-term map* and it was defined for the first time in [85] and further analyzed in [157]. This map encodes in a systematic manner many properties that were implicitly used and assumed in several papers. Among its many useful properties, we will see that the contact-term map C gives rise to the various contact terms in the local equations of motion of multiparticle superfields. In addition, its combinatorial properties will allow us to prove that the later equations of motions of non-local superfields exhibit a "deconcatenation property" in their non-linear terms, based on fine-tuned cancellations of the contact terms and associated kinematic poles.

The contact-term map acting on a letter i and on Lie monomials $[P,Q]$ is defined by the following recursion [85, 157]

$$\begin{aligned} C(i) &:= 0\,, \\ C([P,Q]) &:= [C(P),Q]+[P,C(Q)]+(k_P\cdot k_Q)\big(P\otimes Q-Q\otimes P\big)\,, \end{aligned} \tag{4.21}$$

where the Lie bracket in the space $\mathcal{L}$ of all Lie polynomials is extended canonically to $\mathcal{L}\otimes\mathcal{L}$ as

$$\begin{aligned} [A\otimes B,Q] &:= [A,Q]\otimes B+A\otimes[B,Q]\,, \\ [P,A\otimes B] &:= [P,A]\otimes B+A\otimes[P,B]\,, \end{aligned} \tag{4.22}$$

and we have $k_P = k_1+k_2+\cdots+k_p$ for $P = 12\ldots p$ according to the definition (4.9) for multiparticle momenta. To illustrate the definition, some examples can be worked out to give

$$\begin{aligned} C([1,2]) &= (k_1\cdot k_2)(1\otimes 2-2\otimes 1)\,, \\ C([[1,2],3]) &= (k_1\cdot k_2)\big([1,3]\otimes 2+1\otimes[2,3]-[2,3]\otimes 1-2\otimes[1,3]\big) \\ &\quad+(k_{12}\cdot k_3)\big([1,2]\otimes 3-3\otimes[1,2]\big)\,, \\ C([1,[2,3]]) &= (k_2\cdot k_3)\big([1,2]\otimes 3+2\otimes[1,3]-[1,3]\otimes 2-3\otimes[1,2]\big) \\ &\quad+(k_1\cdot k_{23})\big(1\otimes[2,3]-[2,3]\otimes 1\big)\,, \\ C([[[1,2],3],4]) &= (k_1\cdot k_2)\big([[1,3],4]\otimes 2+[1,3]\otimes[2,4]+[1,4]\otimes[2,3]+1\otimes[[2,3],4] \\ &\qquad -[[2,3],4]\otimes 1-[2,3]\otimes[1,4]-[2,4]\otimes[1,3]-2\otimes[[1,3],4]\big) \\ &\quad+(k_{12}\cdot k_3)\big([[1,2],4]\otimes 3+[1,2]\otimes[3,4]-[3,4]\otimes[1,2]-3\otimes[[1,2],4]\big) \\ &\quad+(k_{123}\cdot k_4)\big([[1,2],3]\otimes 4-4\otimes[[1,2],3]\big)\,, \\ C([[1,2],[3,4]]) &= (k_1\cdot k_2)\big([1,[3,4]]\otimes 2+1\otimes[2,[3,4]]-[2,[3,4]]\otimes 1-2\otimes[1,[3,4]]\big) \\ &\quad+(k_3\cdot k_4)\big([[1,2],3]\otimes 4+3\otimes[[1,2],4]-[[1,2],4]\otimes 3-4\otimes[[1,2],3]\big) \\ &\quad+(k_{12}\cdot k_{34})\big([1,2]\otimes[3,4]-[3,4]\otimes[1,2]\big)\,, \end{aligned} \tag{4.23}$$

where the expression for $C([[1,2],3])$ for instance encodes the Mandelstam invariants in QU_{123} previewed in (4.20). By definition, the contact-term map produces the antisymmetrized combinations $P\otimes Q-Q\otimes P$ of Lie monomials. Therefore it is convenient to consider the image of the contact-term map as being in the wedge product of Lie polynomials

$$P\wedge Q := P\otimes Q-Q\otimes P\,, \tag{4.24}$$

which implies that (4.22) becomes $[A\wedge B,C]=[A,C]\wedge B+A\wedge[B,C]$. Using this notation streamlines the output of the contact-term map, for example

$$C([1,[2,3]]) = (k_2\cdot k_3)\big([1,2]\wedge 3+2\wedge[1,3]\big)+(k_1\cdot k_{23})\big(1\wedge[2,3]\big)\,. \tag{4.25}$$

Contact-term map and BRST charge. A definition, implicit in [85, 157], extends the action of the contact-term map to $\mathcal{L}\wedge\mathcal{L}$ as

$$C(P\wedge Q) = C(P)\wedge Q-P\wedge C(Q)\,. \tag{4.26}$$

From this definition it follows that

Lemma 3. *The contact-term map is nilpotent,*

$$C^2 = 0\,. \tag{4.27}$$

Proof. See Appendix H.

The condition (4.27) turns out to be an important consistency check, as the contact-term map will be related to the pure spinor BRST charge in the discussions below,

$$C \leftrightarrow Q_{\text{BRST}}\,. \tag{4.28}$$

In addition, when acting on a left-to-right Dynkin bracket $\ell(P) = [[\ldots[[p_1, p_2], p_3], \ldots], p_n]$ defined in (C.1), it gives rise to the deshuffle sums, as proven by induction [85]

$$C(\ell(P)) = \sum_{\substack{XjY=P\\ \delta(Y)=R\otimes S}} (k_X \cdot k_j)\big[\ell(XR)\otimes\ell(jS) - \ell(jR)\otimes\ell(XS)\big] = \sum_{\substack{XjY=P\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\ell(XR)\wedge\ell(jS) \tag{4.29}$$

where the deshuffle map $\delta(Y)$ is defined in (C.10) and the effect of the swap $X \leftrightarrow j$ is to replace $\otimes \to \wedge$ since the deshuffle map $\delta(Y) = R\otimes S$ is symmetric sum over R and S. For example,

$$C(\ell(123)) = (k_1\cdot k_2)\big(1\wedge\ell(23) + \ell(13)\wedge 2\big) + (k_{12}\cdot k_3)\ell(12)\wedge 3\,. \tag{4.30}$$

As we will see later, this is the same structure of the BRST variation of V_{123} seen in (4.73) and will play an important role in motivating the correspondence (4.176) below.

The synergy between the contact-term map and multiparticle superfields will become more natural once we define how the Lie polynomials P and Q in the image of the contact-term map become labels of generic superfields[14] K and T

$$(K\otimes T)_{P\otimes Q} := K_P T_Q\,, \qquad (K\wedge T)_{P\wedge Q} := K_P T_Q\,, \tag{4.31}$$

and extended by linearity. For example,

$$(A^m\otimes V)_{[[1,2],3]\otimes[4,[5,6]]} = A^m_{[[1,2],3]}V_{[4,[5,6]]}\,, \qquad (V\wedge V)_{s_{12}1\wedge 2} = s_{12}V_1V_2\,. \tag{4.32}$$

4.1.2. Multiparticle superfield in the Lorenz gauge

The generalization of the single-particle linearized superfields of (2.14) to an arbitrary number of labels naturally leads to a Lie-polynomial structure for the multiparticle labels. For a simplified definition sufficient for this review, P is a Lie polynomial if it is a linear combination of words generated by nested Lie brackets acting on non-commutative letters representing the particle labels. For example, $P = [[1,2],3] = 123-213-312+321$ is a Lie polynomial.

Initially defined by consistency of the resulting equations of motion in [83], the following recursive definition of multiparticle superfields was identified in [84] to correspond to a multiparticle version of superfields in the Lorenz gauge:

$$\begin{aligned}
\hat A^{[P,Q]}_\alpha &= \frac{1}{2}\big[\hat A^Q_\alpha(k_Q\cdot\hat A_P) + \hat A^m_Q(\gamma_m\hat W_P)_\alpha - (P\leftrightarrow Q)\big]\,,\\
\hat A^m_{[P,Q]} &= \frac{1}{2}\big[\hat A^m_Q(k_Q\cdot\hat A_P) + \hat A^P_n\hat F^{nm}_Q + (\hat W_P\gamma^m\hat W_Q) - (P\leftrightarrow Q)\big]\,,\\
\hat W^\alpha_{[P,Q]} &= \frac{1}{4}\hat F^{rs}_P(\gamma_{rs}\hat W_Q)^\alpha + \frac{1}{2}(k_Q\cdot\hat A_P)\hat W^\alpha_Q + \frac{1}{2}\hat W^{m\alpha}_Q\hat A^P_m - (P\leftrightarrow Q)\,,\\
\hat F^{mn}_{[P,Q]} &= \frac{1}{2}\big[\hat F^{mn}_Q(k_Q\cdot\hat A_P) + \hat F^{p|mn}_Q\hat A^P_p + \hat F^{[m}_Q{}_r\hat F^{n]r}_P - 2\gamma^{[m}_{\alpha\beta}\hat W^{n]\alpha}_P\hat W^\beta_Q - (P\leftrightarrow Q)\big]\,,
\end{aligned} \tag{4.33}$$

where the momentum indexed by a Lie polynomial is understood to be stripped of brackets, for example $k^m_{[1,[2,3]]} = k^m_{123}$. The hat in $\hat K_{[P,Q]}$ above distinguishes this definition in the Lorenz gauge from other definitions of multiparticle superfields in other gauges, as we will see shortly. In order to complete (4.33) to a recursion, we define the multiparticle instances of the higher-mass-dimension superfields in section 2.3

$$\begin{aligned}
\hat W^{m\alpha}_{[P,Q]} &= k^m_{PQ}\hat W^\alpha_{[P,Q]} - (\hat A^m\otimes\hat W^\alpha)_{C([P,Q])}\,,\\
\hat F^{m|pq}_{[P,Q]} &= k^m_{PQ}\hat F^{pq}_{[P,Q]} - (\hat A^m\otimes\hat F^{pq})_{C([P,Q])}\,,
\end{aligned} \tag{4.34}$$

where the contact-term map C is defined in (4.21) and we are using the notation (4.31), for instance

$$\hat F^{m|pq}_{[1,2]} = k^m_{12}\hat F^{pq}_{[1,2]} - (k_1\cdot k_2)(\hat A^m_1\hat F^{pq}_2 - \hat A^m_2\hat F^{pq}_1)\,,$$

[14] This notation deviates from the one used in [85].

$$
\begin{aligned}
\hat{F}^{m|pq}_{[1,[2,3]]} &= k^m_{123}\hat{F}^{pq}_{[1,[2,3]]} - (k_2\cdot k_3)\big(\hat{A}^m_{[1,2]}\hat{F}^{pq}_3 + \hat{A}^m_2\hat{F}^{pq}_{[1,3]} - \hat{A}^m_{[1,3]}\hat{F}^{pq}_2 - \hat{A}^m_3\hat{F}^{pq}_{[1,2]}\big) \\
&\quad - (k_1\cdot k_{23})\big(\hat{A}^m_1\hat{F}^{pq}_{[2,3]} - \hat{A}^m_{[2,3]}\hat{F}^{pq}_1\big)\,.
\end{aligned} \tag{4.35}
$$

Like in the multiplicity-two case (4.14), the multiparticle field strength can be rewritten in a more conventional form as

$$
\hat{F}^{mn}_{[P,Q]} = k^m_{PQ}\hat{A}^n_{[P,Q]} - k^m_{PQ}\hat{A}^m_{[P,Q]} - (\hat{A}^m\otimes\hat{A}^n)_{C([P,Q])}\,. \tag{4.36}
$$

The recursions (4.33) terminate with the single-particle superfields $\hat{K}_i = K_i$, and the resulting two-particle superfields in Lorenz gauge turn out to match the expressions (4.13) obtained from OPEs, i.e. $\hat{K}_{[i,j]} = K_{ij}$.

It is important to emphasize that the above recursions apply to arbitrary bracketing structures encompassed by P and Q. For example,

$$
\begin{aligned}
\hat{A}^m_{[1,2]} &= \frac{1}{2}\Big[\hat{A}^m_2(k_2\cdot\hat{A}_1) + \hat{A}^1_n\hat{F}^{nm}_2 + (\hat{W}_1\gamma^m\hat{W}_2) - (1\leftrightarrow 2)\Big]\,, \\
\hat{A}^m_{[[1,2],3]} &= \frac{1}{2}\Big[\hat{A}^m_3(k_3\cdot\hat{A}_{[1,2]}) + \hat{A}^{[1,2]}_n\hat{F}^{nm}_3 + (\hat{W}_{[1,2]}\gamma^m\hat{W}_3) - ([1,2]\leftrightarrow 3)\Big]\,, \\
\hat{A}^m_{[[1,2],[[3,4],5]]} &= \frac{1}{2}\Big[\hat{A}^m_{[[3,4],5]}(k_{345}\cdot\hat{A}_{[1,2]}) + \hat{A}^{[1,2]}_n\hat{F}^{nm}_{[[3,4],5]} + (\hat{W}_{[1,2]}\gamma^m\hat{W}_{[[3,4],5]}) - ([1,2]\leftrightarrow[[3,4],5])\Big]\,.
\end{aligned} \tag{4.37}
$$

In addition, from the contact-term terms in $C([[1,2],[3,4]])$ as in (4.23), namely

$$
\begin{aligned}
C([[1,2],[3,4]]) &= (k_1\cdot k_2)\big([1,[3,4]]\otimes 2 + 1\otimes[2,[3,4]] - [2,[3,4]]\otimes 1 - 2\otimes[1,[3,4]]\big) \\
&\quad + (k_3\cdot k_4)\big([[1,2],3]\otimes 4 + 3\otimes[[1,2],4] - [[1,2],4]\otimes 3 - 4\otimes[[1,2],3]\big) \\
&\quad + (k_{12}\cdot k_{34})\big([1,2]\otimes[3,4] - [3,4]\otimes[1,2]\big)\,,
\end{aligned} \tag{4.38}
$$

the field-strength (4.36) for $P=[1,2]$ and $Q=[3,4]$ becomes

$$
\hat{F}^{mn}_{[[1,2],[3,4]]} = k^m_{1234}\hat{A}^n_{[[1,2],[3,4]]} - k^n_{1234}\hat{A}^m_{[[1,2],[3,4]]} \tag{4.39}
$$

$$
\begin{aligned}
&- (k_1\cdot k_2)\big(\hat{A}^m_{[1,[3,4]]}\hat{A}^n_2 + \hat{A}^m_1\hat{A}^n_{[2,[3,4]]} - (1\leftrightarrow 2)\big) \\
&- (k_3\cdot k_4)\big(\hat{A}^m_{[[1,2],3]}\hat{A}^n_4 + \hat{A}^m_3\hat{A}_{[[1,2],4]} - (3\leftrightarrow 4)\big) \\
&- (k_{12}\cdot k_{34})\big(\hat{A}^m_{[1,2]}\hat{A}^n_{[3,4]} - \hat{A}^m_{[3,4]}\hat{A}^n_{[1,2]}\big)\,.
\end{aligned} \tag{4.40}
$$

Identifying the pair of words P and Q for the superfields on the right-hand side of the above examples leads to further applications of the recursions in (4.33) until eventually all superfields are of single-particle nature. Naturally, the number of terms in the multiparticle superfields increases very rapidly when expanded down to single-particle superfields. Fortunately, there is rarely the need for doing so as even component expansions using the top cohomology factor (3.80) of pure spinor superspace can be performed efficiently at a multiparticle level, see Appendix F.

OPE channels and Catalan numbers. In presence of more than two vertex operators, different orders of performing the OPEs lead to different multiparticle superfields. One can intuitively understand the different bracketings of the definition (4.33) of multiparticle superfields in the Lorenz gauge and the associated vertex operators

$$
\hat{V}_P := \lambda^\alpha\hat{A}^P_\alpha\,, \qquad \hat{U}_P := \partial\theta^\alpha\hat{A}^P_\alpha + \Pi^m\hat{A}^P_m + d_\alpha\hat{W}^\alpha_P + \tfrac{1}{2}N^{mn}\hat{F}^P_{mn}\,, \tag{4.41}
$$

in this way: three vertex operators $U_1(z_1)U_2(z_2)U_3(z_3)$ admit two[15] possible ways of performing two OPEs in sequence while preserving the order of z_i on the disk: $z_2\to z_1$ and $z_3\to z_1$ or $z_3\to z_2$ and $z_2\to z_1$. These two possibilities lead to two possible bracketings for the resulting multiparticle vertex at position z_1

$$
\hat{U}_{[[1,2],3]}(z_1)\,, \qquad \hat{U}_{[1,[2,3]]}(z_1)\,. \tag{4.42}
$$

[15] Of course, in a correlation function, the ordering of operators is arbitrary and will lead to index permutations of the multiparticle vertices. But in the end, the *BCJ-gauge* counterpart $U_{[2,[1,3]]}$ of $\hat{U}_{[2,[1,3]]}$ will still be expressible in terms of $U_{[1,[2,3]]}, U_{[[1,2],3]}$, see section 4.1.5.

In general, for a string of p vertex operators there will be C_{p-1} bracketings, where C_{p-1} is the $(p-1)$-th Catalan number[16]. For example, the $C_3 = 5$ bracketings corresponding to the different OPE orderings among neighboring vertices in the correlation $U_1(z_1)U_2(z_2)U_3(z_3)U_4(z_4)$ give rise to the following vertex operators at z_1:

$$\hat{U}_{[[[1,2],3],4]}(z_1)\,, \quad \hat{U}_{[[1,[2,3]],4]}(z_1)\,, \quad \hat{U}_{[[1,2],[3,4]]}(z_1)\,, \quad \hat{U}_{[1,[[2,3],4]]}(z_1)\,, \quad \hat{U}_{[1,[2,[3,4]]]}(z_1)\,. \tag{4.43}$$

For example, the first vertex in (4.43) corresponds to performing the OPEs as $z_2 \to z_1$ first, then $z_3 \to z_1$ and $z_4 \to z_1$.

4.1.3. Equations of motion of local multiparticle superfields in Lorenz gauge

The equations of motion satisfied by the local multiparticle superfields given in the recursive definition (4.33) can be written in a similar fashion as their single-particle counterparts of (2.14). To see this we define an analogue of $\nabla_\alpha := D_\alpha - \mathbb{A}_\alpha$ at the level of local multiparticle superfields as

$$\nabla^{(L)}_\alpha \hat{K}_{[P,Q]} := D_\alpha \hat{K}_{[P,Q]} - (\hat{A}_\alpha \otimes \hat{K})_{C([P,Q])} \tag{4.44}$$

in terms of the contact-term map (4.21) and the notation (4.31).

With this definition, the equations of motion for the Lorenz-gauge superfields $\hat{K}_{[P,Q]}$ can be written as

$$\begin{aligned}
\nabla^{(L)}_{(\alpha} \hat{A}^{[P,Q]}_{\beta)} &= \gamma^m_{\alpha\beta} \hat{A}^{[P,Q]}_m\,, & \nabla^{(L)}_\alpha \hat{A}^m_{[P,Q]} &= k^m_{PQ} \hat{A}^{[P,Q]}_\alpha + (\gamma^m \hat{W}_{[P,Q]})_\alpha\,, \\
\nabla^{(L)}_\alpha \hat{W}^\beta_{[P,Q]} &= \frac{1}{4} (\gamma^{mn})_\alpha{}^\beta \hat{F}^{[P,Q]}_{mn}\,, & \nabla^{(L)}_\alpha \hat{F}^{mn}_{[P,Q]} &= \big(\hat{W}^{[m}_{[P,Q]} \gamma^{n]}\big)_\alpha\,,
\end{aligned} \tag{4.45}$$

which assume exactly the same form as their non-linear counterparts (2.11). After expanding out the derivatives $\nabla^{(L)}_\alpha$ in (4.44), the local equations of motion for the Lorenz-gauge superfields (4.45) are given by

$$\begin{aligned}
D_{(\alpha} \hat{A}^{[P,Q]}_{\beta)} &= \gamma^m_{\alpha\beta} \hat{A}^{[P,Q]}_m + (\hat{A}_\alpha \otimes \hat{A}_\beta)_{C([P,Q])}\,, \\
D_\alpha \hat{A}^{[P,Q]}_m &= (\gamma_m \hat{W}^{[P,Q]})_\alpha + k^{PQ}_m \hat{A}^{[P,Q]}_\alpha + (\hat{A}_\alpha \otimes \hat{A}^m)_{C([P,Q])}\,, \\
D_\alpha \hat{W}^\beta_{[P,Q]} &= \frac{1}{4} (\gamma^{mn})_\alpha{}^\beta \hat{F}^{[P,Q]}_{mn} + (\hat{A}_\alpha \otimes \hat{W}^\beta)_{C([P,Q])}\,, \\
D_\alpha \hat{F}^{mn}_{[P,Q]} &= \big(\hat{W}^{[m}_{[P,Q]} \gamma^{n]}\big)_\alpha + (\hat{A}_\alpha \otimes \hat{F}^{mn})_{C([P,Q])}\,.
\end{aligned} \tag{4.46}$$

In the simplest case $P = 1$ and $Q = 2$, the contact-term map produces a factor of $k_1 \cdot k_2$ and we recover the two-particle equations of motion (4.15) upon noting that $\hat{K}_{[i,j]} = K_{ij}$. To illustrate the above equations, consider the equation of motion $D_\alpha \hat{A}^{[P,Q]}_m$ for $[P,Q] = [[1,2],3]$. Using the contact-term map $C([[1,2],3])$ from (4.23) leads to

$$\begin{aligned}
D_\alpha \hat{A}^m_{[[1,2],3]} &= (\gamma^m \hat{W}_{[[1,2],3]})_\alpha + k^m_{123} \hat{A}^{[[1,2],3]}_\alpha \\
&\quad + (k_1 \cdot k_2)\big(\hat{A}^{[1,3]}_\alpha \hat{A}^m_2 + \hat{A}^1_\alpha \hat{A}^m_{[2,3]} - \hat{A}^{[2,3]}_\alpha \hat{A}^m_1 - \hat{A}^2_\alpha \hat{A}^m_{[1,3]}\big) \\
&\quad + (k_{12} \cdot k_3)\big(\hat{A}^{[1,2]}_\alpha \hat{A}^m_3 - \hat{A}^3_\alpha \hat{A}^m_{[1,2]}\big)\,,
\end{aligned} \tag{4.47}$$

thus recovering equation (3.20) from [83].

4.1.4. Local multiparticle superfields in the BCJ gauge

The explicit calculations of string disk amplitudes at multiplicities five [109, 159] and six [161] revealed a truly fascinating pattern arising from a conjunction of factors: first, the double poles in the OPEs among

[16] Catalan numbers are given by $C_n = \frac{1}{n+1}\binom{2n}{n}$, and the simplest examples are $C_1 = 1$, $C_2 = 2$, $C_3 = 5$ and $C_4 = 14$.

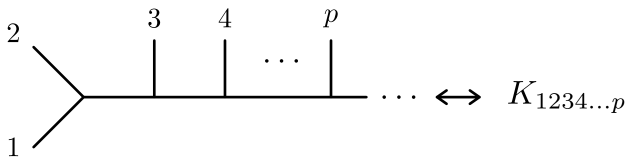


Figure 1: The symmetry mapping between a half-ladder cubic graph and the local SYM multiparticle superfields $K \in \{A_\alpha, A_m, W^\alpha, F^{mn}\}$.

massless vertex operators can be integrated by parts within the full string integrand containing the Koba–Nielsen factor $\sim \prod_{1\le i<j}^n |z_{ij}|^{-k_i\cdot k_j}$, see (4.12) for a two-particle example. This amounts to redistributing the superfields in the double-pole terms among various single-pole terms in the OPEs of the vertex operators. Second, the superfields in the numerators of the double poles have the precise form that, once redistributed to single-pole terms via integration by parts, lead to effective single-pole numerators that satisfy improved symmetry properties within their multiparticle labels – so-called generalized Jacobi identities [162]. This mechanism hinges on the fact that the BRST-exact terms in (4.19) match the total derivatives in (4.12).

Unfortunately, doing these calculations in practice is tedious, and currently the best justification for this mechanism is the total-derivative distribution seen at the two-particle OPE (4.12) and extensive explicit cancellations in the six-point disk amplitude of [161], see section 3.2 and appendix B.3 of the reference. Luckily after these patterns were understood in [83] these integration-by-parts steps could be bypassed by the recursive procedure to be described below. However, it remains a challenging open problem to rigorously prove the recursions from the OPE approach.

4.1.5. Generalized Jacobi identities

In section 4.1.6 below, we will introduce a gauge-transformed version K_P of the multiparticle superfields $\hat{K}_P$ in Lorenz gauge defined in (4.33). Before spelling out these redefinitions due to double-pole terms in OPEs, we shall here describe the resulting symmetries of the multiparticle labels in K_P. These symmetries can be summarized by

$$K_{A\ell(B)C} + K_{B\ell(A)C} = 0\,, \quad A, B \neq \emptyset\,, \quad \forall C\,, \tag{4.48}$$

where $\ell(A)$ is the left-to-right Dynkin bracket,

$$\ell(123\ldots n) := \ell(123\ldots n{-}1)n - n\ell(123\ldots n{-}1)\,, \quad \ell(i) := i\,, \quad \ell(\emptyset) := 0\,, \tag{4.49}$$

for instance $\ell(12) = 12 - 21$ and $\ell(123) = 123 - 213 - 312 + 321$. The symmetries (4.48) are known as the *generalized Jacobi identities* in the mathematics literature [162] (see also section 8.6.7 of [143]).

These are the same symmetries obtained by a string of standard structure constants [163], or equivalently, by a Dynkin bracket

$$\ell(1234\ldots p) \leftrightarrow K_{1234\ldots p} \leftrightarrow f^{12a_2} f^{a_2 3 a_3} f^{a_3 4 a_4} \ldots f^{a_{p-1} p a_p}\,, \tag{4.50}$$

see figure 1. A few examples of generalized Jacobi identities are as follows

$$\begin{aligned} K_{12C} + K_{21C} &= 0\,, \quad \forall C\,,\\ K_{123C} + K_{231C} + K_{312C} &= 0\,, \quad \forall C\,,\\ K_{1234C} + K_{2143C} + K_{3412C} + K_{4321C} &= 0\,, \quad \forall C\,. \end{aligned} \tag{4.51}$$

Let A be a word and $\ell(A)$ its Dynkin bracket defined in (4.49). The generalized Jacobi identities correspond to the elements in the kernel of ℓ. The simplest examples

$$\ell(12+21) = 0\,, \qquad \ell(123+231+312) = 0 \tag{4.52}$$

are tantamount to the antisymmetry and Jacobi identity of the Lie bracket.

Using Baker's identity $\ell(P\ell(Q)) = [\ell(P), \ell(Q)]$ [143], it is easy to see that $A\ell(B) + B\ell(A)$ is in the kernel of ℓ for any pair of words A and B. In addition, due to the recursive definition of ℓ, if $\ell(P) = 0$ it also follows that $\ell(PQ) = 0$ for the concatenation of P with any word Q. Therefore the generalized Jacobi identities can be encoded by an abstract operator $\pounds_k$

$$\pounds_k \circ K_{ABC} := K_{A\ell(B)C} + K_{B\ell(A)C}\,, \quad \forall\, A, B \neq \emptyset \text{ and } \forall\, C \text{ such that } |A| + |B| = k\,. \tag{4.53}$$

We emphasize the arbitrary partition of non-empty words A and B in the above definition (while C can be empty), leading to a non-unique operator $\pounds$. For instance

$$\begin{aligned} \pounds_3 \circ K_{123} &= K_{123} - K_{132} + K_{231}\,, \quad \text{for } A = 1,\ B = 23 \text{ and } C = \emptyset\,, \\ \pounds_3 \circ K_{123} &= K_{123} + K_{312} - K_{321}\,, \quad \text{for } A = 12,\ B = 3 \text{ and } C = \emptyset\,. \end{aligned} \tag{4.54}$$

However, if $\pounds_2 \circ K_{123} = 0$ then the right-hand sides of the above expressions agree and both are equal to $K_{123} + K_{231} + K_{312}$.

For reasons to become clear later, multiparticle superfields that satisfy $\pounds_k K_P = 0$ for all $k \leq |P|$ are said to be in the BCJ gauge. When K_P is in the BCJ gauge we use the notation

$$K_{\ell(P)} := K_P\,. \tag{4.55}$$

For example $K_{[[1,2],3]}$ in the BCJ gauge (not to be confused with the Lorenz gauge) is represented by K_{123}. In particular, with this notation we can write Baker's identity for superfields in the BCJ gauge as

$$K_{[P,Q]} = K_{P\ell(Q)}\,. \tag{4.56}$$

For example, $K_{[12,34]} = K_{1234} - K_{1243}$. The expansion of more general bracketings works similarly, and it amounts to rewriting an arbitrary Lie monomial $[\Gamma, \Sigma]$ in the Dynkin bracket basis of $\ell(1P)$, for instance:

$$K_{[[12,34],[5,67]]} = K_{1234567} - K_{1234576} - K_{1234675} + K_{1234765} - K_{1243567} + K_{1243576} + K_{1243675} - K_{1243765}\,. \tag{4.57}$$

In addition, if K_P with $P = AiB$ and i a single letter satisfies generalized Jacobi identities, then it follows from (4.48) that

$$K_{AiB} = -K_{i\ell(A)B}\,, \quad A \neq \emptyset\,, \forall\, B\,, \tag{4.58}$$

as $\ell(i) = i$ for a letter i. This relation implies that there is an $(p-1)!$ basis of permutations $K_{i_1 \ldots i_p}$ of $K_{12\ldots p}$.

4.1.6. Multiparticle superfields in the BCJ gauge

As mentioned above, explicit calculations of superstring disk amplitudes with the pure spinor formalism led to the discovery that the superfield numerators of single-pole integrands were being redefined by the double-pole terms under integration by parts.[17] The end result of these redefinitions is an improved symmetry property of the composite superfields, which turns out to be the same as the generalized Jacobi identities of the previous section.

Later computations in [83, 84, 85] showed that the multiparticle superfields in the BCJ gauge could be computed by an intrinsic recursive method independent on the (fortuitous) interference between superstring OPEs and integration by parts when multiplied by the Koba–Nielsen factor. The recursive method can be approached in two ways, each one convenient for different situations. Let us now review these.

[17] The simplicity arising from integration by parts hinges on the fact that the double-pole terms in superspace are proportional to factors of $(1 + s_{ijk\ldots})$. Explicit computations up to six points show that this is the case in the pure spinor formalism.

From a hybrid gauge to BCJ gauge. It is convenient to encode the redefinitions needed to attain the BCJ gauge by defining an intermediate set of multiparticle superfields in a *hybrid gauge*, denoted by $\check{K}_{[P,Q]}$. The definition of multiparticle superfields in the hybrid gauge $\check{K}_{[P,Q]}$ assumes that all superfields of lower multiplicities K_P and K_Q have been redefined to satisfy all the generalized Jacobi identities, i.e. $\pounds_k K_P = 0$ for $k \leq |P|$ (and similarly for Q). We then define

$$
\begin{aligned}
\check{A}_\alpha^{[P,Q]} &= \frac{1}{2}[A_\alpha^Q (k_Q \cdot A_P) + A_Q^m (\gamma_m W_P)_\alpha - (P \leftrightarrow Q)] \,, \\
\check{A}^m_{[P,Q]} &= \frac{1}{2}[A_Q^m (k_Q \cdot A_P) + A_n^P F_Q^{nm} + (W_P \gamma^m W_Q) - (P \leftrightarrow Q)] \,, \\
\check{W}^\alpha_{[P,Q]} &= \frac{1}{4} F_P^{rs} (\gamma_{rs} W_Q)^\alpha + \frac{1}{2}(k_Q \cdot A_P) W_Q^\alpha + \frac{1}{2} W_Q^{m\alpha} A_P^m - (P \leftrightarrow Q) \,, \\
\check{F}^{mn}_{[P,Q]} &= \frac{1}{2}\big[F_Q^{mn}(k_Q \cdot A_P) + F_Q^{r|mn} A_r^P + F_Q^{[m}{}_r F_P^{n]r} - 2\gamma^{[m}_{\alpha\beta} W_P^{n]\alpha} W_Q^\beta - (P \leftrightarrow Q)\big] \,,
\end{aligned}
\tag{4.59}
$$

and $\check{K}_i = K_i$, where the superfields K_P and K_Q on the right-hand side satisfy the generalized Jacobi identities $\pounds_k K_P = 0$ for $k \leq |P|$ and

$$
\begin{aligned}
W^{m\alpha}_{[P,Q]} &= k^m_{PQ} W^\alpha_{[P,Q]} - (A^m \otimes W^\alpha)_{C([P,Q])} \,, \\
F^{m|pq}_{[P,Q]} &= k^m_{PQ} F^{pq}_{[P,Q]} - (A^m \otimes F^{pq})_{C([P,Q])} \,,
\end{aligned}
\tag{4.60}
$$

are the local form of the multiparticle superfields of higher mass dimension defined in [84] involving the contact-term map $C([P,Q])$ in (4.21) and the notation (4.31).

In contrast to the definitions (4.33) of Lorenz-gauge superfields $\hat{K}_{[P,Q]}$, in the hybrid gauge their definitions (4.59) are *not* recursive: The superfields $\check{K}_{[P,Q]}$ on the left-hand side of (4.59) have to be redefined $\check{K}_{[P,Q]} \to K_{[P,Q]}$ before qualifying as input on the right-hand side in the next step of the recursion.

The hybrid gauge leads to more convenient explicit expressions to arrive at multiparticle superfields in the BCJ gauge. One can show that the following redefinitions

$$
\begin{aligned}
K_{[P,Q]} := \check{K}_{[P,Q]} &- \sum_{\substack{P=XjY \\ \delta(Y)=R\otimes S}} (k_X \cdot k_j)\big[H_{[XR,Q]}\, K_{jS} - (X \leftrightarrow j)\big] \\
&+ \sum_{\substack{Q=XjY \\ \delta(Y)=R\otimes S}} (k_X \cdot k_j)\big[H_{[XR,P]}\, K_{jS} - (X \leftrightarrow j)\big] - \begin{cases} D_\alpha H_{[P,Q]} & : \ K = A_\alpha \\ k^m_{PQ} H_{[P,Q]} & : \ K = A^m \\ 0 & : \ K = W^\alpha \end{cases}
\end{aligned}
\tag{4.61}
$$

imply that the left-hand side satisfies all generalized Jacobi identities. Note that $\delta(Y)$ in (4.61) denotes the deshuffle map defined in (C.10) and the superfields H will be defined below. To illustrate the above redefinitions we write down explicit examples for $A^m_{[P,Q]}$ up to multiplicity five (recall that $\check{A}^m_i := A^m_i$ and $\check{A}^m_{[i,j]} := A^m_{ij}$)

$$
\begin{aligned}
A^m_{[12,3]} &= \check{A}^m_{[12,3]} - k^m_{123} H_{[12,3]} \,, \\
A^m_{[12,34]} &= \check{A}^m_{[12,34]} - k^m_{1234} H_{[12,34]} \\
&\quad - (k_1 \cdot k_2)\Big[H_{[1,34]} A^m_2 - H_{[2,34]} A^m_1\Big] \\
&\quad + (k_3 \cdot k_4)\Big[H_{[3,12]} A^m_4 - H_{[4,12]} A^m_3\Big] \,, \\
A^m_{[123,4]} &= \check{A}^m_{[123,4]} - k^m_{1234} H_{[123,4]} \\
&\quad - (k_1 \cdot k_2)\Big[H_{[13,4]} A^m_2 - H_{[23,4]} A^m_1\Big] \\
&\quad - (k_{12} \cdot k_3) H_{[12,4]} A^m_3 \,,
\end{aligned}
\tag{4.62}
$$

$$\begin{aligned}
A^m_{[1234,5]} &= \check{A}^m_{[1234,5]} - k^m_{12345} H_{[1234,5]} \\
&\quad - (k_1 \cdot k_2)\big[H_{[134,5]} A^m_2 + H_{[14,5]} A^m_{23} + H_{[13,5]} A^m_{24} - (1 \leftrightarrow 2)\big] \\
&\quad - (k_{12} \cdot k_3)\big[H_{[124,5]} A^m_3 + H_{[12,5]} A^m_{34} - (12 \leftrightarrow 3)\big] \\
&\quad - (k_{123} \cdot k_4) H_{[123,5]} A^m_4 \,, \\
A^m_{[123,45]} &= \check{A}^m_{[123,45]} - k^m_{12345} H_{[123,45]} \\
&\quad - (k_1 \cdot k_2)\Big[H_{[13,45]} A^m_2 + H_{[1,45]} A^m_{23} - (1 \leftrightarrow 2)\Big] \\
&\quad - (k_{12} \cdot k_3)\Big[H_{[12,45]} A^m_3 - (12 \leftrightarrow 3)\Big] \\
&\quad + (k_4 \cdot k_5)\Big[H_{[4,123]} A^m_5 - (4 \leftrightarrow 5)\Big] \,.
\end{aligned}$$

The explicit expressions for the new superfields $H_{[P,Q]}$ were obtained up to multiplicity five in [84] and for arbitrary multiplicity in [85]:

$$H_{[i,j]} = 0 \,, \qquad H_{[A,B]} = (-1)^{|B|} \frac{|A|}{|A| + |B|} \sum_{XjY = \dot{a}\tilde{B}} (-1)^{|Y|} H'_{\tilde{Y},j,X} - (A \leftrightarrow B) \,, \tag{4.63}$$

where $\dot{a}$ and $\dot{b}$ denote the letterifications of A and B as defined in (C.12) and

$$\begin{aligned}
H'_{A,B,C} &:= H_{A,B,C} + \Big[\frac{1}{2} H_{[A,B]}(k_{AB} \cdot A_C) + \text{cyc}(A,B,C)\Big] \\
&\quad - \Big[\sum_{\substack{XjY = A \\ \delta(Y) = R \otimes S}} (k_X \cdot k_j)\big[H_{[XR,B]} H_{[jS,C]} - (X \leftrightarrow j)\big] + \text{cyc}(A,B,C)\Big] \,,
\end{aligned} \tag{4.64}$$

$$H_{A,B,C} := -\frac{1}{4} A^m_A A^n_B F^{mn}_C + \frac{1}{2} (W_A \gamma_m W_B) A^m_C + \text{cyc}(A,B,C) \,. \tag{4.65}$$

It is straightforward to check that the superfields $H_{[A,B]}$ satisfy generalized Jacobi identities within A and B. This justifies using the notation where nested brackets are flattened, e.g. $H_{[[[1,2],3],4]} = H_{[123,4]}$ in accordance with the notation (4.55). The combination of (4.63), (4.64) and (4.65) reduces all the redefinitions (4.61) from hybrid gauge to BCJ gauge to products of building blocks $H_{A,B,C}$ and $(k_{AB} \cdot A_C)$.

The superfields $H_{[P,Q]}$ in (4.63) up to multiplicity seven are given by

$$\begin{aligned}
H_{[12,3]} &= \frac{1}{3}\big(H'_{1,2,3}\big) \,, \\
H_{[123,4]} &= \frac{1}{4}\big(H'_{12,3,4} - H'_{1,2,43}\big) \,, \\
H_{[12,34]} &= \frac{1}{4}\big(2H'_{1,2,34} - 2H'_{3,4,12}\big) \,, \\
H_{[1234,5]} &= \frac{1}{5}\big(H'_{123,4,5} - H'_{12,3,54} + H'_{1,2,543}\big) \,, \\
H_{[123,45]} &= \frac{1}{5}\big(2H'_{12,3,45} - 2H'_{1,2,453} - 3H'_{4,5,123}\big) \,, \\
H_{[12345,6]} &= \frac{1}{6}\big(H'_{1234,5,6} - H'_{123,4,65} + H'_{12,3,654} - H'_{1,2,6543}\big) \,, \\
H_{[1234,56]} &= \frac{1}{6}\big(2H'_{123,4,56} - 2H'_{12,3,564} + 2H'_{1,2,5643} - 4H'_{5,6,1234}\big) \,, \\
H_{[123,456]} &= \frac{1}{6}\big(3H'_{12,3,456} - 3H'_{1,2,4563} - 3H'_{45,6,123} + 3H'_{4,5,1236}\big) \,, \\
H_{[123456,7]} &= \frac{1}{7}\big(H'_{12345,6,7} - H'_{1234,5,76} + H'_{123,4,765} - H'_{12,3,7654} + H'_{1,2,76543}\big) \,,
\end{aligned} \tag{4.66}$$

$$H_{[12345,67]} = \frac{1}{7}\big(2H'_{1234,5,67} - 2H'_{123,4,675} + 2H'_{12,3,6754} - 2H'_{1,2,67543} - 5H'_{6,7,12345}\big)\,,$$
$$H_{[1234,567]} = \frac{1}{7}\big(3H'_{123,4,567} - 3H'_{12,3,5674} + 3H'_{1,2,56743} - 4H'_{56,7,1234} + 4H'_{5,6,12347}\big)\,,$$

and the simplest non-vanishing instances of the primed superfields in (4.64) are

$$\begin{aligned} H'_{1,2,3} &= H_{1,2,3}\,, \\ H'_{12,3,4} &= H_{12,3,4} + \frac{1}{6}\big[H_{1,2,3}(k_{123}\cdot A_4) - (3\leftrightarrow 4)\big]\,. \end{aligned} \tag{4.67}$$

For example, the explicit expressions for the first two superfields above are given by

$$\begin{aligned} H_{[12,3]} &= -\frac{1}{12}A_1^m A_2^n F_3^{mn} + \frac{1}{6}(W_1\gamma_m W_2)A_3^m + \mathrm{cyc}(1,2,3)\,, \\ H_{[123,4]} &= \frac{1}{4}(H_{12,3,4} + H_{34,1,2}) + \frac{1}{24}\big[H_{1,2,3}(k_{123}\cdot A_4) - H_{1,2,4}(k_{124}\cdot A_3) + H_{3,4,1}(k_{134}\cdot A_2) - H_{3,4,2}(k_{234}\cdot A_1)\big]\,. \end{aligned} \tag{4.68}$$

It is interesting to observe that the expressions for the superfields $H_{[A,B]}$ that lead to the BCJ gauge are not unique. In fact, simpler explicit expressions can be derived using the Bern-Kosower formalism [164].

To complement the definition (4.61), the field strength in the BCJ gauge is defined using the contact-term map (4.21)

$$F^{mn}_{[P,Q]} = k^m_{PQ}A^n_{[P,Q]} - k^n_{PQ}A^m_{[P,Q]} - (A^m\otimes A^n)_{C([P,Q])}\,, \tag{4.69}$$

see the definition (4.31). Concretely, the above superfields can be explicitly checked to satisfy the generalized Jacobi identities. For example, using the notation (4.55) as $A^m_{[1234,5]} = A^m_{12345}$, one can see that

$$A^m_{12345} + A^m_{21435} + A^m_{34125} + A^m_{43215} = 0\,, \tag{4.70}$$

corresponding to the third identity in (4.51) with $C = 5$. In addition, long calculations demonstrate that

$$A^m_{12345} - A^m_{12354} + A^m_{45123} - A^m_{45213} - A^m_{45312} + A^m_{45321} = 0\,, \tag{4.71}$$

in agreement with the expansion (4.56) applied to $A^m_{[123,45]} + A^m_{[45,123]} = 0$.

As an alternative to the method above to obtain multiparticle superfields in the BCJ gauge, one can choose to go directly from the Lorenz gauge to the BCJ gauge. The process is more or less the same, but the explicit formulae make it more evident that the whole process corresponds to a finite gauge transformation of the corresponding perturbiner expansion of Berends–Giele currents to be reviewed shortly. The discussion of these redefinitions is left for the Appendix G.

4.1.7. Equations of motion of multiparticle superfields in the BCJ gauge

Written in terms of the BRST charge $Q = \lambda^\alpha D_\alpha$, the equations of motion for the multiparticle superfields in the BCJ gauge become ($k_\emptyset := 0$) [83]

$$\begin{aligned} QV_P &= \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\, V_{XR}V_{jS}\,, \\ QA^m_P &= (\lambda\gamma^m W_P) + k^m_P V_P + \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\big[V_{XR}A^m_{jS} - V_{jR}A^m_{XS}\big]\,, \\ QW^\beta_P &= \frac{1}{4}(\lambda\gamma^{mn})^\beta F^P_{mn} + \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\big[V_{XR}W^\beta_{jS} - V_{jR}W^\beta_{XS}\big]\,, \\ QF^{mn}_P &= (\lambda\gamma^{[n}W^{m]}_P) + \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\big[V_{XR}F^{mn}_{jS} - V_{jR}F^{mn}_{XS}\big]\,, \end{aligned} \tag{4.72}$$

where $V_P = \lambda^\alpha A^P_\alpha$ is the multiparticle unintegrated vertex operator and the last line involves the superfield $W^{m\alpha}_P$ of higher mass dimension defined in (4.60). In addition, the sum over $P = XjY$ assembles the $|P|-1$ deconcatenations of the word P into a word X, a single letter j, and a word Y. Moreover, $\delta(Y) = R \otimes S$ denotes the deshuffle (C.10) of the word Y into the words R and S. A few examples help to illustrate the above formulae,

$$
\begin{aligned}
QV_1 &= 0\,, \\
QV_{12} &= (k_1\cdot k_2)V_1V_2\,, \\
QV_{123} &= (k_1\cdot k_2)\big[V_1V_{23}+V_{13}V_2\big] + (k_{12}\cdot k_3)V_{12}V_3\,, \\
QV_{1234} &= (k_1\cdot k_2)\big[V_1V_{234}+V_{13}V_{24}+V_{14}V_{23}+V_{134}V_2\big] \\
&\quad + (k_{12}\cdot k_3)\big[V_{12}V_{34}+V_{124}V_3\big] + (k_{123}\cdot k_4)V_{123}V_4\,, \\
QV_{12345} &= (k_1\cdot k_2)\big[V_1V_{2345}+V_{13}V_{245}+V_{134}V_{25}+V_{1345}V_2 \\
&\qquad + V_{135}V_{24}+V_{14}V_{235}+V_{145}V_{23}+V_{15}V_{234}\big] \\
&\quad + (k_{12}\cdot k_3)\big[V_{12}V_{345}+V_{124}V_{35}+V_{1245}V_3+V_{125}V_{34}\big] \\
&\quad + (k_{123}\cdot k_4)\big[V_{123}V_{45}+V_{1235}V_4\big] \\
&\quad + (k_{1234}\cdot k_5)V_{1234}V_5\,.
\end{aligned}
\tag{4.73}
$$

Note that the instances at rank ≤ 4 can be formally obtained from the BRST variations (4.6) and (4.8) of the OPE residues upon promoting $L_{2131\ldots p1} \to V_{123\ldots p}$. In other words, the chain of redefinitions in (4.19) and generalizations to higher rank preserve the form of the covariant BRST algebra.

Recalling the notation (4.55) $K_{\ell(P)} := K_P$ for superfields in the BCJ gauge, the BRST variations in (4.72) for the left-to-right nested commutator $\ell(P)$ can be obtained as the special case $[R,S] := \ell(P)$ of the BRST variations for general commutators

$$
\begin{aligned}
QV_{[R,S]} &= \frac{1}{2}(V\otimes V)_{C([R,S])}\,, \\
QA^m_{[R,S]} &= (\lambda\gamma^m W_{[R,S]}) + k^m_{RS}V_{[R,S]} + (V\otimes A^m)_{C([R,S])}\,, \\
QW^\beta_{[R,S]} &= \frac{1}{4}(\lambda\gamma_{mn})^\beta F^{mn}_{[R,S]} + (V\otimes W^\beta)_{C([R,S])}\,, \\
QF^{mn}_{[R,S]} &= \big(\lambda W^{[m}_{[R,S]}\gamma^{n]}\big) + (V\otimes F^{mn})_{C([R,S])}\,,
\end{aligned}
\tag{4.74}
$$

where we are employing the notation (4.31) for the contact-term map. The fact that

$$
(V\otimes K)_{C(P)} = \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\big[V_{XR}K_{jS} - (X\leftrightarrow j)\big]
\tag{4.75}
$$

for superfields in the BCJ gauge was proven in Lemma 1 of [85][18].

The multiparticle versions of the on-shell constraints $k^m_1(\gamma_m W_1)_\alpha = 0 = k^1_m F^{mn}_1$ take a form similar to (4.72),

$$
\begin{aligned}
k^m_P(\gamma_m W_P)_\alpha &= \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\big[A^m_{XR}(\gamma_m W_{jS})_\alpha - A^m_{jR}(\gamma_m W_{XS})_\alpha\big]\,, \\
k^P_m F^{mn}_P &= \sum_{\substack{P=XjY\\ \delta(Y)=R\otimes S}} (k_X\cdot k_j)\big[\gamma^n_{\alpha\beta}W^\alpha_{XR}W^\beta_{jS} + A^{XR}_m F^{mn}_{jS} - (X\leftrightarrow j)\big]\,,
\end{aligned}
\tag{4.76}
$$

for instance

$$
k^m_{12}(\gamma_m W_{12})_\alpha = (k_1\cdot k_2)\big[A^m_1(\gamma_m W_2)_\alpha - A^m_2(\gamma_m W_1)_\alpha\big]\,,
\tag{4.77}
$$

[18]The replacement $K_{\ell(P)} \to K_P$ was left implicit in the proof of [85].

$$k^{12}_m F^{mn}_{12} = (k_1 \cdot k_2)\big[2(W_1\gamma^n W_2) + A^1_m F^{mn}_2 - A^2_m F^{mn}_1\big] .$$

Using the multiparticle SYM superfields in the BCJ gauge, it is natural to define the multiparticle massless vertices in the pure spinor formalism as

$$V_{[P,Q]} = \lambda^\alpha A^{[P,Q]}_\alpha , \quad U_{[P,Q]} = \partial\theta^\alpha A^{[P,Q]}_\alpha + \Pi^m A^{[P,Q]}_m + d_\alpha W^\alpha_{[P,Q]} + \tfrac{1}{2} N^{mn} F^{[P,Q]}_{mn} , \tag{4.78}$$

where $[P,Q]$ denotes an arbitrary Lie monomial, e.g. $[[1,[2,3]],[4,5]]$. The multiparticle equations of motion (4.74) and the non-linear Dirac equation (4.76) imply the relations

$$QU_{[P,Q]} = \partial V_{[P,Q]} + (V \otimes U)_{C([P,Q])} \tag{4.79}$$

between the two types of vertex operators, in lines with $QU_1 = \partial V_1$ and the rank-two example (4.18).

4.2. Non-local superfields and Berends–Giele currents

In addition to the local multiparticle superfields reviewed above, the pure spinor formalism naturally leads to another class of multiparticle superfields containing kinematic poles in generalized Mandelstam invariants [160, 83]. These non-local SYM superfields are denoted collectively by $\mathcal{K}_P$ and were dubbed *Berends–Giele currents*[19] [160] due to their relation with the standard gluonic currents J^m_P defined by Berends and Giele in the 80s [27]. Specifically, they share the same shuffle symmetries, and the $\theta = 0$ term in the Berends–Giele superfield current $\mathcal{A}^m_P(\theta)$ in a suitable gauge is equal to J^m_P, see section 5 for a review.

Each local superfield representative in $K_P \in \{A^P_\alpha, A^m_P, W^\alpha_P, F^{mn}_P\}$ admits a corresponding Berends–Giele current with multiparticle label $P = 12\ldots p$ denoted by calligraphic letters

$$\mathcal{K}_P \in \{\mathcal{A}^P_\alpha, \mathcal{A}^m_P, \mathcal{W}^\alpha_P, \mathcal{F}^{mn}_P\} , \tag{4.80}$$

starting with $\mathcal{K}_1 := K_1$ and

$$\begin{aligned}
\mathcal{K}_{12} &= \frac{K_{12}}{s_{12}} , \\
\mathcal{K}_{123} &= \frac{K_{123}}{s_{12}s_{123}} + \frac{K_{321}}{s_{23}s_{123}} , \\
\mathcal{K}_{1234} &= \frac{1}{s_{1234}}\left(\frac{K_{1234}}{s_{12}s_{123}} + \frac{K_{3214}}{s_{23}s_{123}} + \frac{K_{3421}}{s_{34}s_{234}} + \frac{K_{3241}}{s_{23}s_{234}} + \frac{K_{[12,34]}}{s_{12}s_{34}}\right) ,
\end{aligned} \tag{4.81}$$

with generalized Mandelstam invariants defined in (1.23), $s_{12\ldots p} = \frac{1}{2}k^2_{12\ldots p}$. In contrast to the bosonic Berends–Giele currents in [27], the supercurrents $\mathcal{K}_P$ also contain fermionic degrees of freedom as required by supersymmetry, and their construction does not include any quartic vertices. Note that for historical reasons the Berends–Giele currents associated to the local multiparticle unintegrated vertex $V_P = \lambda^\alpha A^P_\alpha$ is denoted by M_P rather than $\mathcal{V}_P$,

$$M_P = \lambda^\alpha \mathcal{A}^P_\alpha . \tag{4.82}$$

More explicitly, $M_1 = V_1$ and

$$\begin{aligned}
M_{12} &= \frac{V_{12}}{s_{12}} , \\
M_{123} &= \frac{V_{123}}{s_{12}s_{123}} + \frac{V_{321}}{s_{23}s_{123}} ,
\end{aligned} \tag{4.83}$$

[19] The story is longer than this since in [160] the relation with the standard Berends–Giele current J^m_P was observed from a structural similarity between the appearance of the superfield Berends–Giele current M_P in the the pure spinor cohomology formula for SYM tree-level amplitudes and in the role played by J^m_P in the standard Berends–Giele setup. It was much later in [165, 84] that a rigorous relation between the superfield version of the Berends–Giele currents $\mathcal{K}_P$ and the standard bosonic J^m_P was established via $\mathcal{A}^m(\theta)$. So in fact, the non-local superfields $\mathcal{K}_P$ generalize the Berends–Giele currents in a supersymmetric manner and may also be named Berends–Giele *supercurrents*.

$$M_{1234} = \frac{1}{s_{1234}} \left(\frac{V_{1234}}{s_{12}s_{123}} + \frac{V_{3214}}{s_{23}s_{123}} + \frac{V_{3421}}{s_{34}s_{234}} + \frac{V_{3241}}{s_{23}s_{234}} + \frac{V_{[12,34]}}{s_{12}s_{34}} \right).$$

In the early stages of unraveling the cohomology properties of multiparticle superfields with the pure spinor formalism, the Berends–Giele currents $\mathcal{K}_P$ were defined case by case to encompass all tree subdiagrams compatible with the ordering of the external legs in P in such a way as to transform BRST covariantly [160, 21]. For instance, from the equations of motion (4.73) we get

$$\begin{aligned} QM_1 &= 0\,, \\ QM_{12} &= M_1 M_2\,, \\ QM_{123} &= M_{12}M_3 + M_1 M_{23}\,, \\ QM_{1234} &= M_{123}M_4 + M_{12}M_{34} + M_1 M_{234}\,. \end{aligned} \tag{4.84}$$

In contrast to $QV_{123\ldots p}$ as given by (4.73), there are no explicit Mandelstam variables in (4.84) as the propagators $s^{-1}_{i\ldots j}$ in (4.81) absorb the appearance of explicit momenta in the contact terms of the equations of motion of the local superfields. A rigorous proof of this statement from a combinatorial perspective can be found in (4.156). The generalization of (4.84) to higher rank is given by [160]

$$Q\,M_P = \sum_{XY=P} M_X M_Y \tag{4.85}$$

and it was proven in [84]. The sum involves all the $|P|-1$ deconcatenations $XY{=}P$ of P into non-empty[20] words X, Y, e.g. $X = 12\ldots j$ and $Y = j{+}1\ldots p$ with $j = 1, 2, \ldots, p{-}1$ in case of $P = 12\ldots p$. These deconcatenations will be later on associated with partitions of the p on-shell legs on two Berends–Giele currents while preserving the color ordering. For a combinatorial proof of (4.85) from the perspective of BRST variations of the composing local numerators V_Q, see (4.179).

It is useful to define a BRST-exact superfield E_P as

$$E_P = \sum_{XY=P} M_X M_Y \tag{4.86}$$

which will be used in the pure spinor cohomology formula for SYM tree-level amplitudes in section 5.2. One can show that E_P is conditionally BRST exact

$$E_P = QM_P \qquad \text{if } s_P \neq 0\,, \tag{4.87}$$

provided that $s_P \neq 0$ is true as M_P contains the propagator $1/s_P$. We will see later that this condition of the momentum phase space is of crucial importance.

The connection between Berends–Giele currents subject to the deconcatenation equation (4.85) and the cubic tree subdiagrams compatible with a color ordered amplitude is supported by the following consistency check: the number of terms or kinematic pole channels in $M_{12\ldots p}$ is the Catalan number C_{p-1} (see (5.19) below for a proof) which counts the number of cubic diagrams in a color-ordered $(p{+}1)$-point tree amplitude. As the relation with the Catalan number suggests, the definition of Berends–Giele currents admits a beautiful combinatorial interpretation in terms of planar binary trees and is connected with the KLT matrix in many surprising ways [166, 156, 157]. We will return to this point later in section 4.3.

4.2.1. Non-linear wave equations and Berends–Giele currents

In [84] the definition of Berends–Giele currents was shown to arise from solutions of the non-linear wave equations of ten-dimensional SYM theory in the *Lorenz gauge*

$$[\partial_m, \mathbb{A}^m] = 0\,. \tag{4.88}$$

[20]Defining $M_\emptyset := 0$, the restriction to non-empty words may be lifted and the general definition (1.22) may be applied.

To show this one first needs to derive the non-linear wave equations obeyed by the superfields $\mathbb{K} \in \{\mathbb{A}_\alpha, \mathbb{A}^m, \mathbb{W}^\alpha, \mathbb{F}^{mn}\}$. This can be done starting from

$$\Box\mathbb{K} = [\partial^m, [\partial_m, \mathbb{K}]] \tag{4.89}$$

and using the Jacobi identity together with $\partial^m = \nabla^m + \mathbb{A}^m$. That is,

$$\begin{aligned}\Box\mathbb{K} &= [\nabla^m + \mathbb{A}^m, [\partial_m, \mathbb{K}]] \\ &= [[\nabla^m, \partial_m], \mathbb{K}] + [\mathbb{A}^m, [\partial_m, \mathbb{K}]] + [\mathbb{A}^m, [\nabla_m, \mathbb{K}]] + [\nabla^m, [\nabla_m, \mathbb{K}]]\,.\end{aligned} \tag{4.90}$$

The first term in the second line vanishes in Lorentz gauge (4.88) as $[\partial_m, \nabla^m] = -[\partial_m, \mathbb{A}^m]$. For the simpler choices of superfields $\mathbb{K} \to \{\nabla_\alpha, \nabla_m\}$, the last term of (4.90) can be converted to quadratic expressions in the non-linear fields using the Dirac and super Yang–Mills equations (2.8). In the case of $\mathbb{K} \to \{\mathbb{W}^\alpha, \mathbb{F}^{mn}\}$, the analogous conversion necessitates the equations of motion[21],

$$\begin{aligned}\big[\nabla_m, \mathbb{W}^{m\alpha}\big] &= \frac{1}{2}\big[\mathbb{F}_{mn}, (\gamma^{mn}\mathbb{W})^\alpha\big]\,, \\ \big[\nabla_p, \mathbb{F}^{p|mn}\big] &= 2\big[\mathbb{F}^{mp}, \mathbb{F}_p{}^n\big] + 2\big\{(\mathbb{W}^{[m}\gamma^{n]})_\alpha, \mathbb{W}^\alpha\big\}\end{aligned} \tag{4.91}$$

of the higher-dimension superfields $\mathbb{W}^{m\alpha} := [\nabla^m, \mathbb{W}^\alpha]$ and $\mathbb{F}^{p|mn} := [\nabla^p, \mathbb{F}^{mn}]$ from (2.20). Upon inserting (2.8) and (4.91) into (4.90), one gets [84]:

$$\begin{aligned}\Box\mathbb{A}_\alpha &= \big[\mathbb{A}_m, [\partial^m, \mathbb{A}_\alpha]\big] + \big[(\gamma^m\mathbb{W})_\alpha, \mathbb{A}_m\big]\,, \\ \Box\mathbb{A}^m &= \big[\mathbb{A}_p, [\partial^p, \mathbb{A}^m]\big] + \big[\mathbb{F}^{mp}, \mathbb{A}_p\big] + \gamma^m_{\alpha\beta}\{\mathbb{W}^\alpha, \mathbb{W}^\beta\}\,, \\ \Box\mathbb{W}^\alpha &= \big[\mathbb{A}_m, [\partial^m, \mathbb{W}^\alpha]\big] + \big[\mathbb{A}^m, \mathbb{W}^\alpha_m\big] + \frac{1}{2}\big[\mathbb{F}_{mn}, (\gamma^{mn}\mathbb{W})^\alpha\big]\,, \\ \Box\mathbb{F}^{mn} &= [\mathbb{A}_p, [\partial^p, \mathbb{F}^{mn}]] + [\mathbb{A}_p, \mathbb{F}^{p|mn}] + 2[\mathbb{F}^{mp}, \mathbb{F}_p{}^n] + 2\{(\mathbb{W}^{[m}\gamma^{n]})_\alpha\,, \mathbb{W}^\alpha\}\,,\end{aligned} \tag{4.92}$$

with the convention $A^{[m}B^{n]} = A^mB^n - A^nB^m$. We will see below that these equations are the precursors of supersymmetric Berends–Giele recursion relations. In particular, the bosonic restriction of the equation for $\Box\mathbb{A}^m$ will give rise to a derivation of the standard Berends–Giele currents of [27].

4.2.2. Perturbiner solution

To solve the wave equations (4.92), it is convenient to use the perturbiner method of Rosly and Selivanov [23, 24, 25, 26] by expanding the superfields $\mathbb{K} \in \{\mathbb{A}_\alpha, \mathbb{A}^m, \mathbb{W}^\alpha, \mathbb{F}^{mn}\}$ as a series with respect to the generators t^{i_j} of a Lie algebra, summed over all possible non-empty words $P = p_1p_2\ldots p_{|P|}$ [87] (note $t^P := t^{p_1}t^{p_2}\cdots t^{p_{|P|}}$)

$$\begin{aligned}\mathbb{K} &:= \sum_P \mathcal{K}_P t^P e^{k_P\cdot X} = \sum_{i_1}\mathcal{K}_{i_1}t^{i_1}e^{k_{i_1}\cdot X} + \sum_{i_1,i_2}\mathcal{K}_{i_1i_2}t^{i_1}t^{i_2}e^{k_{i_1i_2}\cdot X} + \cdots \\ &= \sum_{p=1}^\infty\sum_{i_1,i_2,\ldots,i_p}\frac{1}{p}\,\mathcal{K}_{i_1i_2\ldots i_p}[t^{i_1}, [t^{i_2}, \ldots, [t^{i_{p-2}}, [t^{i_{p-1}}, t^{i_p}]]\ldots]]e^{k_{i_1i_2\ldots i_p}\cdot X}\,,\end{aligned} \tag{4.93}$$

with coefficients given by $\mathcal{K}_P$, which will be identified with the Berends–Giele currents shortly. The second line follows from the shuffle symmetry (4.103) obeyed by the Berends–Giele currents and guarantees that $\mathbb{K}$ is Lie-algebra valued, see [167] for a proof. Note that we are implicitly considering the generators t^i to be formally nilpotent[22] $t^i\ldots t^i = 0$ in order to avoid repetition of indices like in $\mathcal{K}_{112}$ or $\mathcal{K}_{121}$.

---

[21] The first line of (4.91) can be derived by applying the Clifford algebra $[\nabla_m, \mathbb{W}^{m\alpha}] = \frac{1}{2}[\nabla_m, [\nabla_n, (\gamma^m\gamma^n\mathbb{W})^\alpha + (\gamma^n\gamma^m\mathbb{W})^\alpha]]$ followed by the Dirac equation, Jacobi relations and the definition $\mathbb{F}_{mn} = -[\nabla_m, \nabla_n]$. The second line of (4.91) in turn follows from $[\nabla_m, \mathbb{F}^{m|pq}] = [\nabla_m, [\nabla^{[p}, [\nabla^{q]}, \nabla^m]]]$ combined with the super Yang–Mills equation and additional Jacobi relations.

[22] In the original perturbiner discussion of [23], repeated indices are avoided by adjoining nilpotent symbols $\mathcal{E}^i$ to each t^i in the expansion (4.93).

In order to derive recursion relations for the expansion coefficients $\mathcal{K}_P \in \{\mathcal{A}^P_\alpha, \mathcal{A}^m_P, \mathcal{W}^\alpha_P, \mathcal{F}^{mn}_P\}$, we insert the series (4.93) into (4.92) and use the action of Box operator $\Box e^{k_P\cdot X} = 2s_P e^{k_P\cdot X}$ on the plane-wave factors of the superfields. By isolating the coefficient of t^P in the wave equations, one readily finds that

$$\mathcal{K}_P = \frac{1}{s_P}\sum_{XY=P}\mathcal{K}_{[X,Y]}\,, \tag{4.94}$$

where the contribution from each deconcatenation of P into non-empty X, Y is a non-local version of (4.33)

$$\begin{aligned}
\mathcal{A}^{[P,Q]}_\alpha &= \frac{1}{2}\big[\mathcal{A}^Q_\alpha(k_Q\cdot\mathcal{A}_P) + \mathcal{A}^m_Q(\gamma_m\mathcal{W}_P)_\alpha - (P\leftrightarrow Q)\big]\,, \\
\mathcal{A}^m_{[P,Q]} &= \frac{1}{2}\big[\mathcal{A}^m_Q(k_Q\cdot\mathcal{A}_P) + \mathcal{A}^P_n\mathcal{F}^{nm}_Q + (\mathcal{W}_P\gamma^m\mathcal{W}_Q) - (P\leftrightarrow Q)\big]\,, \\
\mathcal{W}^\alpha_{[P,Q]} &= \frac{1}{4}\mathcal{F}^{rs}_P(\gamma_{rs}\mathcal{W}_Q)^\alpha + \frac{1}{2}\mathcal{W}^\alpha_Q(k_Q\cdot\mathcal{A}_P) + \frac{1}{2}\mathcal{W}^{m\alpha}_Q\mathcal{A}^m_P - (P\leftrightarrow Q)\big]\,, \\
\mathcal{F}^{mn}_{[P,Q]} &= \frac{1}{2}\big[\mathcal{F}^{mn}_Q(k_Q\cdot\mathcal{A}_P) + \mathcal{F}^{p|mn}_Q\mathcal{A}^P_p + \mathcal{F}^{[m}_Q{}_r\mathcal{F}^{n]r}_P - 2\gamma^{[m}_{\alpha\beta}\mathcal{W}^{n]\alpha}_P\mathcal{W}^\beta_Q - (P\leftrightarrow Q)\big]\,.
\end{aligned} \tag{4.95}$$

The definition $\mathbb{F}^{mn} = -[\nabla^m, \nabla^n]$ and those of higher-mass-dimension superfields lead to the following Berends–Giele currents

$$\begin{aligned}
\mathcal{F}^{mn}_P &= k^m_P\mathcal{A}^n_P - k^n_P\mathcal{A}^m_P - \sum_{XY=P}\big(\mathcal{A}^m_X\mathcal{A}^n_Y - \mathcal{A}^m_Y\mathcal{A}^n_X\big)\,, \\
\mathcal{W}^{m\alpha}_P &= k^m_P\mathcal{W}^\alpha_P + \sum_{XY=P}\big(\mathcal{W}^\alpha_X\mathcal{A}^m_Y - \mathcal{W}^\alpha_Y\mathcal{A}^m_X\big)\,, \\
\mathcal{F}^{m|pq}_P &= k^m_P\mathcal{F}^{pq}_P + \sum_{XY=P}\big(\mathcal{F}^{pq}_X\mathcal{A}^m_Y - \mathcal{F}^{pq}_Y\mathcal{A}^m_X\big)\,,
\end{aligned} \tag{4.96}$$

and the above recursion terminates with the single-particle superfields $\mathcal{K}_i = K_i \in \{A^i_\alpha, A^m_i, W^\alpha_i, F^{mn}_i\}$. By comparing the expressions in (4.81) with the first few explicit expansions from (4.95) it is possible to recognize these expansions as the Berends–Giele currents obtained previously using BRST cohomology arguments.

4.2.3. Equations of motion of Berends–Giele currents

By inserting the perturbiner expansions (4.93) of the SYM superfields in $\mathbb{K}$ into their non-linear equations of motion (2.12), one immediately obtains the equations of motion of the Berends–Giele currents in the form

$$\begin{aligned}
D_\alpha\mathcal{A}^P_\beta + D_\beta\mathcal{A}^P_\alpha &= \gamma^m_{\alpha\beta}\mathcal{A}^P_m + \sum_{XY=P}\big(\mathcal{A}^X_\alpha\mathcal{A}^Y_\beta - \mathcal{A}^Y_\alpha\mathcal{A}^X_\beta\big)\,, \\
D_\alpha\mathcal{A}^P_m &= k^P_m\mathcal{A}^P_\alpha + (\gamma_m\mathcal{W}_P)_\alpha + \sum_{XY=P}\big(\mathcal{A}^X_\alpha\mathcal{A}^Y_m - \mathcal{A}^Y_\alpha\mathcal{A}^X_m\big)\,, \\
D_\alpha\mathcal{W}^\beta_P &= \frac{1}{4}(\gamma^{mn})_\alpha{}^\beta\mathcal{F}^P_{mn} + \sum_{XY=P}\big(\mathcal{A}^X_\alpha\mathcal{W}^\beta_Y - \mathcal{A}^Y_\alpha\mathcal{W}^\beta_X\big)\,, \\
D_\alpha\mathcal{F}^{mn}_P &= (\mathcal{W}^{[m}_P\gamma^{n]})_\alpha + \sum_{XY=P}\big(\mathcal{A}^X_\alpha\mathcal{F}^{mn}_Y - \mathcal{A}^Y_\alpha\mathcal{F}^{mn}_X\big)
\end{aligned} \tag{4.97}$$

by comparing coefficients of the products of gauge generators t^P on both sides. Apart from the deconcatenation sum $\sum_{XY=P}$, these equations of motion have the same form as the linearized ones (2.14). For example the two- and three-particle equations of motion of $\mathcal{A}^{12}_\alpha$ and $\mathcal{A}^{123}_\alpha$ read

$$\begin{aligned}
D_\alpha\mathcal{A}^{12}_\beta + D_\beta\mathcal{A}^{12}_\alpha &= \gamma^m_{\alpha\beta}\mathcal{A}^{12}_m + \mathcal{A}^1_\alpha\mathcal{A}^2_\beta - \mathcal{A}^2_\alpha\mathcal{A}^1_\beta\,, \\
D_\alpha\mathcal{A}^{123}_\beta + D_\beta\mathcal{A}^{123}_\alpha &= \gamma^m_{\alpha\beta}\mathcal{A}^{123}_m + \mathcal{A}^1_\alpha\mathcal{A}^{23}_\beta + \mathcal{A}^{12}_\alpha\mathcal{A}^3_\beta - \mathcal{A}^{23}_\alpha\mathcal{A}^1_\beta - \mathcal{A}^3_\alpha\mathcal{A}^{12}_\beta\,.
\end{aligned} \tag{4.98}$$

These equations lead to a simple proof of the deconcatenation property (4.85), based on the action of the pure spinor BRST charge on superfields via $Q = \lambda^\alpha D_\alpha$. Therefore, multiplying the first equation of (4.97) by $\lambda^\alpha \lambda^\beta$ and using the pure spinor constraint $\lambda^\alpha \lambda^\beta \gamma^m_{\alpha\beta} = 0$ together with anti-commutativity of the superfields, one recovers the variation (4.85).

As we will review later, these simple equations of motion for $\mathcal{K}_P$ play a key role in various proofs of BRST invariance of scattering amplitudes in string and field theory, see [159, 160, 21] for examples at tree level and [163, 168, 169, 170, 171, 172] at loop level. The need for superfields that represent multi-particle contact vertices on a skeleton graph was also observed in the worldline version of the pure spinor formalism [173, 174].

In addition, the Lorenz gauge as well as the Dirac and super Yang–Mills equations [87]

$$\big[\partial_m, \mathbb{A}^m\big] = 0\,, \quad \big[\nabla_m, (\gamma^m \mathbb{W})_\alpha\big] = 0\,, \quad \big[\nabla_m, \mathbb{F}^{mn}\big] = \gamma^n_{\alpha\beta}\{\mathbb{W}^\alpha, \mathbb{W}^\beta\} \tag{4.99}$$

imply, after using $\nabla_m = \partial_m - \mathbb{A}_m$, that the Berends–Giele currents satisfy

$$k^P_m \mathcal{A}^m_P = 0\,, \tag{4.100}$$
$$k^P_m (\gamma^m \mathcal{W}_P)_\alpha = \sum_{XY=P} \big[\mathcal{A}^X_m (\gamma^m \mathcal{W}_Y)_\alpha - \mathcal{A}^Y_m (\gamma^m \mathcal{W}_X)_\alpha\big]\,, \tag{4.101}$$
$$k^P_m \mathcal{F}^{mn}_P = \sum_{XY=P} \big[2(\mathcal{W}_X \gamma^n \mathcal{W}_Y) + \mathcal{A}^X_m \mathcal{F}^{mn}_Y - \mathcal{A}^Y_m \mathcal{F}^{mn}_X\big]\,. \tag{4.102}$$

While (4.101) and (4.102) have local counterparts (4.76) in BCJ gauge, the local multiparticle-superfields A^m_P subject to generalized Jacobi identities depart from Lorenz gauge and generically obey $k^P_m A^m_P \neq 0$.

4.2.4. Symmetry properties of Berends–Giele currents

The symmetry properties of the $\mathcal{K}_P$ can be inferred from their cubic-graph expansion and can be summarized in terms of the shuffle product ⧢

$$\mathcal{K}_{A \shuffle B} = 0\,, \quad \forall\, A, B \neq \emptyset\,, \tag{4.103}$$

see (4.140) below for the proof. For example,

$$\begin{aligned} 0 &= \mathcal{K}_{1\shuffle 2} = \mathcal{K}_{12} + \mathcal{K}_{21}\,, \\ 0 &= \mathcal{K}_{1\shuffle 23} = \mathcal{K}_{123} + \mathcal{K}_{213} + \mathcal{K}_{231}\,, \\ 0 &= \mathcal{K}_{12\shuffle 3} - \mathcal{K}_{1\shuffle 32} = \mathcal{K}_{123} - \mathcal{K}_{321}\,. \end{aligned} \tag{4.104}$$

The shuffle symmetry (4.103) was proved for the gluonic currents J^m_P of Berends and Giele in [175], while a proof of $\mathcal{K}_{A\shuffle B} = 0$ for their supersymmetric counterparts $\mathcal{K} \in \{\mathcal{A}_\alpha, \mathcal{A}^m, \mathcal{W}^\alpha, \mathcal{F}^{mn}\}$ can be found in the appendix of [84]. Since the θ-independent component of the Berends–Giele current of the vector connection reduces to the gluonic Berends–Giele current, $\mathcal{A}^m_P(\theta = 0) = J^m_P$, the supersymmetric proof of [84] is an alternative proof of the shuffle symmetry of the standard Berends–Giele current J^m_P.

The shuffle symmetry (4.103) implies that the Berends–Giele currents admit a $(p-1)!$-element basis of permutations $\mathcal{K}_{i_1 \ldots i_p}$ of $\mathcal{K}_{12\ldots p}$ which can be taken as $\mathcal{K}_{1\sigma(23\ldots p)}$ with $\sigma \in S_{p-1}$ via Schocker's identity [176]

$$\mathcal{K}_{B1A} = (-1)^{|B|} \mathcal{K}_{1(A \shuffle \tilde{B})}\,, \tag{4.105}$$

where $\tilde{B}$ denotes the word reversal of B, see section 1.3. In particular, for $A = \emptyset$ we get the alternating parity under reversal of P,

$$\mathcal{K}_P = (-1)^{|P|+1} \mathcal{K}_{\tilde{P}}\,, \tag{4.106}$$

for example, $\mathcal{K}_{12} = -\mathcal{K}_{21}$ and $\mathcal{K}_{123} = \mathcal{K}_{321}$, as can be seen from (4.104).

4.2.5. Berends–Giele currents and finite gauge transformations

It was shown in [84, 85] that in terms of the perturbiner series of Berends–Giele currents $\mathbb{K}$, the local superfield redefinitions reviewed in the previous section correspond to a finite gauge transformation of the superfields $\mathbb{K}$ satisfying the non-linear field equations (2.11). To see this, one can explicitly check that the poles in the definition of the Berends–Giele current cancel the contact terms present in the local redefinitions from Lorenz to BCJ gauge (this will be proven in (4.156)). More explicitly, we first define a perturbiner series of the redefining superfields as (recall e.g. $t^{123} := t^1t^2t^3$ etc)

$$\mathbb{H} := \sum_P \mathcal{H}_P t^P e^{k_P\cdot X} \,, \tag{4.107}$$

as well as Lorenz $\mathbb{K}^{\mathrm{L}}$ and BCJ $\mathbb{K}^{\mathrm{BCJ}}$ perturbiner series in which the numerators are composed of local superfields in the Lorenz or BCJ gauge, respectively. For example,

$$\mathcal{K}^{\mathrm{BCJ}}_{123} = \frac{K_{[12,3]}}{s_{12}s_{123}} + \frac{K_{[1,23]}}{s_{23}s_{123}}, \qquad \mathcal{K}^{\mathrm{L}}_{123} = \frac{\hat K_{[[1,2],3]}}{s_{12}s_{123}} + \frac{\hat K_{[1,[2,3]]}}{s_{23}s_{123}} \,, \tag{4.108}$$

with $\hat K_{[P,Q]} = -\hat K_{[Q,P]}$ from (4.33). The local redefinitions from Lorenz to BCJ gauge of the vector superpotential

$$A^m_{[12,3]} = \hat A^m_{[[1,2],3]} - k^m_{123}\hat H_{[12,3]} \,, \qquad A^m_{[1,23]} = \hat A^m_{[1,[2,3]]} - k^m_{123}\hat H_{[1,23]} \tag{4.109}$$

(with $\hat H_{[12,3]}$ defined by (G.6)) imply that their perturbiner series are related by

$$\mathcal{A}^{m,\mathrm{BCJ}}_{123} = \mathcal{A}^{m,\mathrm{L}}_{123} - k^m_{123}\mathcal{H}_{123} \,, \qquad \mathcal{H}_{123} = \frac{\hat H_{[12,3]}}{s_{12}s_{123}} + \frac{\hat H_{[1,23]}}{s_{23}s_{123}} \,, \tag{4.110}$$

corresponding to the gauge transformation

$$\mathbb{A}^{\mathrm{BCJ}}_m = \mathbb{A}^{\mathrm{L}}_m - [\partial_m, \mathbb{H}] + [\mathbb{A}^{\mathrm{L}}_m, \mathbb{H}] + \cdots \,. \tag{4.111}$$

The ellipsis indicates additional terms of a finite gauge transformation (see below) that do not contribute to (4.110) since $\mathcal{H}_1 = \mathcal{H}_{12} = 0$ at multiplicities one and two vanish identically. In fact, the calculations of [85] using superfields up to multiplicity nine revealed that the relation between the Lorenz and BCJ gauges is given by a finite gauge transformation

$$\mathbb{A}^{\mathrm{BCJ}}_m = U\mathbb{A}^{\mathrm{L}}_m U^{-1} + \partial_m U U^{-1} \,, \quad U = \exp(-\mathbb{H}) \tag{4.112}$$

whose expansion yields the omitted terms in (4.111) as an infinite series

$$\mathbb{A}^{\mathrm{BCJ}}_m = \mathbb{A}^{\mathrm{L}}_m + [\mathbb{H}, \partial_m] - [\mathbb{H}, \mathbb{A}^{\mathrm{L}}_m] - \frac{1}{2}[\mathbb{H}, [\mathbb{H}, \partial_m]] + \frac{1}{2}[\mathbb{H}, [\mathbb{H}, \mathbb{A}^{\mathrm{L}}_m]] + \frac{1}{3!}[\mathbb{H}, [\mathbb{H}, [\mathbb{H}, \partial_m]]] + \cdots \,. \tag{4.113}$$

Following [177], one can obtain the series (4.113) iteratively. To see this, define [178]

$$\mathbb{L}_j(\mathbb{A}_m) = \mathbb{A}_m - \frac{1}{j}[\partial_m, \mathbb{H}] - \frac{1}{j}[\mathbb{H}, \mathbb{L}_{j+1}(\mathbb{A}_m)] \tag{4.114}$$

and evaluate

$$\mathbb{A}^{\mathrm{BCJ}}_m = \mathbb{L}_1(\mathbb{A}^{\mathrm{L}}_m) \,. \tag{4.115}$$

The fact that it is the gauge transformation (4.113) that relates the superfields $\mathbb{A}^{\mathrm{BCJ}}_m$ and $\mathbb{A}^{\mathrm{L}}_m$ justifies the terminology of their corresponding local superfields as being in the Lorenz ($\hat K_{[P,Q]}$) or BCJ gauge ($K_{[P,Q]}$).

4.2.6. The multiparticle Berends–Giele polarizations

In section 2.2.1 we have seen that the linearized superfields admit a θ-expansion where each component depends on single-particle polarizations e_i^m, χ_i^α and field-strengths, f_i^{mn}. In principle, the recursive construction of multiparticle Berends–Giele currents at the superspace level also determines the coefficients in their θ-expansion in terms of single-particle polarizations. However, the tensor structure of the θ-expansion (2.17) in the single-particle case is not preserved under the Lorenz-gauge recursion (4.94) and (4.95): generic orders in the θ-expansion of multiparticle $\mathcal{K}_P$ in Lorenz gauge will receive multiple contributions from different partitions of the θs over the lower-multiplicity superfields in (4.95).

A notable exception arises at the zeroth order in θ, where the Lorenz-gauge recursions in superspace have an immediate echo at the level of components: the multiparticle polarizations $\mathfrak{e}_P^m, \mathcal{X}_P^\alpha, \mathfrak{f}_P^{mn}$ defined by setting $\theta = 0$ in

$$\mathfrak{e}_P^m := \mathcal{A}_P^m(0) \ , \qquad \mathcal{X}_P^\alpha := \mathcal{W}_P^\alpha(0) \ , \qquad \mathfrak{f}_P^{mn} := \mathcal{F}_P^{mn}(0) \, , \tag{4.116}$$

obey the following recursions as a consequence of (4.94) and (4.95) (with $\mathfrak{e}_i^m := e_i^m$ and $\mathcal{X}_i^\alpha := \chi_i^\alpha$ for single-particle labels),

$$\mathfrak{e}_P^m = \frac{1}{s_P} \sum_{XY=P} \mathfrak{e}_{[X,Y]}^m \ , \quad \mathcal{X}_P^\alpha = \frac{1}{s_P} \sum_{XY=P} \mathcal{X}_{[X,Y]}^\alpha \ , \tag{4.117}$$

where

$$\mathfrak{e}_{[X,Y]}^m := \frac{1}{2}\big[\mathfrak{e}_Y^m(k_Y \cdot \mathfrak{e}_X) + \mathfrak{e}_n^X \mathfrak{f}_Y^{nm} + (\mathcal{X}_X \gamma^m \mathcal{X}_Y) - (X \leftrightarrow Y)\big] \, , \tag{4.118}$$

$$\mathcal{X}_{[X,Y]}^\alpha := \frac{1}{2}(k_X^p + k_Y^p)\gamma_p^{\alpha\beta}\big[\mathfrak{e}_X^m(\gamma_m \mathcal{X}_Y)_\beta - \mathfrak{e}_Y^m(\gamma_m \mathcal{X}_X)_\beta\big] \ . \tag{4.119}$$

Moreover, the non-linear component field-strength is given by

$$\mathfrak{f}_P^{mn} := k_P^m \mathfrak{e}_P^n - k_P^n \mathfrak{e}_P^m - \sum_{XY=P} \big(\mathfrak{e}_X^m \mathfrak{e}_Y^n - \mathfrak{e}_X^n \mathfrak{e}_Y^m\big) \, . \tag{4.120}$$

Note that the transversality $(k_i \cdot e_i) = 0$ of the gluon and the Dirac equation $k_m^i(\gamma^m \chi_i)_\alpha = 0$ of the gluino propagate as follows to the multiparticle level,

$$(k_P \cdot \mathfrak{e}_P) = 0 \ , \qquad k_m^P(\gamma^m \mathcal{X}_P)_\alpha = \sum_{XY=P} \big[\mathfrak{e}_X^m(\gamma_m \mathcal{X}_Y)_\alpha - \mathfrak{e}_Y^m(\gamma_m \mathcal{X}_X)_\alpha\big] \, , \tag{4.121}$$

where transversality of $\mathfrak{e}_P^m$ is a peculiarity of the Lorenz gauge (4.88) chosen in the derivation of the corresponding superspace Berends–Giele current $\mathcal{A}_P^m(\theta)$.

In Lorenz gauge, the above $\mathfrak{e}_P^m, \mathcal{X}_P^\alpha, \mathfrak{f}_P^{mn}$ are insufficient to describe higher orders $\sim \theta^m$ of multiparticle $\mathcal{K}_P$ with $1 \leq m \leq 5$ which complicates the component expansions via (3.80). However, one can streamline these θ-expansions by means of non-linear gauge transformation (2.6) with a perturbiner expansion of both the superfields $\mathbb{K}$ and the gauge parameter Ω. As detailed in Appendix F, the non-linear version $\theta^\alpha \mathbb{A}_\alpha^{\text{HS}} = 0$ of Harnad–Shnider gauge reorganizes the θ-expansion of the $\mathcal{K}_P$ to simple combinations of the $\mathfrak{e}_P^m, \mathcal{X}_P^\alpha, \mathfrak{f}_P^{mn}$. In particular, the orders $\theta^{\leq 3}$ of $\mathcal{A}_\alpha^P$ relevant for n-point tree-level amplitudes take the same form as in the single-particle θ-expansion (2.17), see (5.30) below, which dramatically simplifies the component expansions in section 5.2.2.

One can similarly arrange the θ-expansions of local multiparticle superfields in Lorenz or BCJ gauge such that the components relevant to tree-level amplitudes are built from three types of multiparticle polarizations. In case of BCJ gauge, the construction of the superfields $A_P^m, W_P^\alpha, F_P^{mn}$ in section 4.1.6 determines the local multiparticle polarizations

$$e_P^m := A_P^m(0) \ , \qquad \chi_P^\alpha := W_P^\alpha(0) \ , \qquad f_P^{mn} := F_P^{mn}(0) \tag{4.122}$$

via evaluation at $\theta = 0$, for instance

$$e_{12}^m = e_2^m(e_1 \cdot k_2) - e_1^m(e_2 \cdot k_1) + \frac{1}{2}(k_1^m - k_2^m)(e_1 \cdot e_2) + (\chi_1 \gamma^m \chi_2) \, ,$$

$$\chi_{12}^{\alpha} = \frac{1}{2} k_{12}^{p} \gamma_p^{\alpha\beta} \left[e_1^m (\gamma_m \chi_2)_\alpha - e_2^m (\gamma_m \chi_1)_\alpha \right] , \tag{4.123}$$
$$f_{12}^{mn} = k_{12}^m e_{12}^n - k_{12}^n e_{12}^m - (k_1 \cdot k_2)(e_1^m e_2^n - e_1^n e_2^m) \,.$$

The local multiparticle polarizations (4.122) obey generalized Jacobi identities in P by construction and compactly encode the components of the local BCJ numerators to be reviewed in section 7.1.3. Note that transversality of the multiparticle polarizations e_P^m at $|P| \geq 3$ is violated in BCJ gauge, e.g.

$$k_{123}^m e_{123}^m = s_{123} \left(\frac{1}{6} e_1^m e_2^n f_3^{mn} - \frac{1}{3} (\chi_1 \gamma_m \chi_2) e_3^m + \mathrm{cyc}(1,2,3) \right) . \tag{4.124}$$

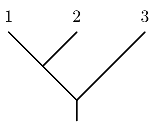
Further details on local multiparticle polarizations can be found in section 4.3 of [165].

4.3. Combinatorial framework of Berends–Giele currents

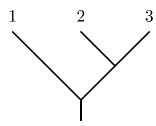
The definition of Berends–Giele currents encompassing all the Catalan number of poles in a color-ordered tree-level amplitude suggests a combinatorial interpretation in terms of planar binary trees. We will see that this point of view provides a rich mathematical framework to prove many assertions related to Berends–Giele currents and associated topics [157].

4.3.1. Planar binary trees

In the appendix of [83] a construction of Berends–Giele currents exploited the fact that nested Lie brackets can be interpreted as planar binary trees and vice versa [179]. A *planar binary tree* is a tree embedded in a plane in which each vertex has three edges: one *root* and two (left and right) *daughters*. An edge is called a *leaf* if it has an end point. In the context of tree-level amplitudes a planar binary tree is also called a *cubic graph* and we map each planar binary tree to a product of inverse Mandelstam invariants (the Feynman propagators) and nested Lie brackets. In addition, each leaf is indexed by a particle label and planarity implies that the labels are in a fixed ordering. For example the two planar binary trees with three leaves labelled $1, 2, 3$ are mapped to



$\frac{[[1,2],3]}{s_{12}s_{123}}$



$\frac{[1,[2,3]]}{s_{23}s_{123}}$

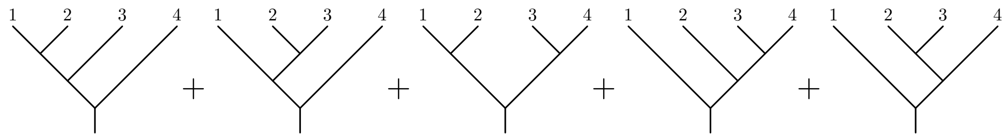
It turns out that the sum over all possible bracketings, or cubic graphs, in a color-ordered tree-level amplitude can be generated from the Lie-polynomial valued recursion proposed in [156] (inspired by [179])

$$b(P) := \frac{1}{s_P} \sum_{XY=P} [b(X), b(Y)] \,, \qquad b(i) := i \,, \qquad b(\emptyset) := 0 \,. \tag{4.125}$$

This recursion constructs combinations $b(P)$ of non-commutative words with inverses of Mandelstam invariants s_P in (3.107) as their coefficients, i.e. the right-hand side of $b(i) = i$ is not understood as $i \in \mathbb{N}$, but as a letter in a non-commutative word. From well-known combinatorial results, the number of terms in the recursion above is given by the Catalan numbers $1, 2, 5, 14, \ldots$[23] and one gets, for example, the following Lie polynomials

$$b(1) = 1, \quad b(12) = \frac{[1,2]}{s_{12}}, \quad b(123) = \frac{[[1,2],3]}{s_{12}s_{123}} + \frac{[1,[2,3]]}{s_{23}s_{123}} \,, \tag{4.126}$$

[23]This can for instance be seen from the recursion $C_{p-1} = \sum_{x+y=p-2} C_x C_y$ for the number of terms in $b(12\ldots p)$ with $C_0 = C_1 = 1$ and $p \geq 3$. As detailed in the discussion around (5.19) below, this coincides with the recursive definition of the Catalan numbers.



$$b(1234) = \frac{[[[1,2],3],4]}{s_{12}s_{123}s_{1234}} + \frac{[[1,[2,3]],4]}{s_{23}s_{123}s_{1234}} + \frac{[[1,2],[3,4]]}{s_{12}s_{34}s_{1234}} + \frac{[1,[2,[3,4]]]}{s_{34}s_{234}s_{1234}} + \frac{[1,[[2,3],4]]}{s_{23}s_{234}s_{1234}}$$

Figure 2: The planar binary trees generated by the recursion of $b(1234)$ from (4.125).

$$b(1234) = \frac{[[[1,2],3],4]}{s_{12}s_{123}s_{1234}} + \frac{[[1,[2,3]],4]}{s_{123}s_{1234}s_{23}} + \frac{[[1,2],[3,4]]}{s_{12}s_{1234}s_{34}} + \frac{[1,[[2,3],4]]}{s_{1234}s_{23}s_{234}} + \frac{[1,[2,[3,4]]]}{s_{1234}s_{234}s_{34}}\,.$$

The nested commutators in the numerators can be expanded in terms of formal words in letters $12\ldots$, and the diagrammatic representation of $b(1234)$ can be found in figure 2.

Lemma 4. *The b map (4.125) is self adjoint,*

$$\langle b(P), Q\rangle = \langle P, b(Q)\rangle\,, \tag{4.127}$$

where $\langle A, B\rangle = \delta_{A,B}$ is the scalar product of words (C.11).

Proof. This is easy to see when $P = i$ is a single letter with $b(i) = i$, so we will use induction over the length the word $|P| := k$ assuming that (4.127) is true for words of length up to $k-1$. Then, from the definition (4.125), the left-hand side of (4.127) becomes

$$\langle b(P), Q\rangle = \frac{1}{s_P}\sum_{XY=P}\langle b(X)b(Y), Q\rangle - (X \leftrightarrow Y)\,. \tag{4.128}$$

Using the elementary property (see (1.5.12) in [143])

$$\langle AB, RS\rangle = \langle A, R\rangle\langle B, S\rangle\,, \quad |A|{=}|R|\,, \quad |B|{=}|S|\,, \tag{4.129}$$

and noting that $|b(X)|{=}|X|$ and $|P|{=}|Q|$ we get

$$\begin{aligned}\langle b(X)b(Y), Q\rangle &= \langle b(X), q_1q_2\ldots q_{|X|}\rangle\langle b(Y), q_{|X|+1}q_{|X|+2}\cdots q_{|Q|}\rangle \\ &= \langle X, b(q_1q_2\ldots q_{|X|})\rangle\langle Y, b(q_{|X|+1}q_{|X|+2}\cdots q_{|Q|})\rangle \\ &= \langle XY, b(q_1q_2\ldots q_{|X|})b(q_{|X|+1}q_{|X|+2}\cdots q_{|Q|})\rangle\,,\end{aligned} \tag{4.130}$$

where in the second line we used the induction hypothesis since $|X| \le k-1$ as the deconcatenation (4.128) vanishes if one of the words is empty due to the definition $b(\emptyset) := 0$. Therefore,

$$\sum_{XY=P}\langle b(X)b(Y), Q\rangle = \sum_{XY=P}\langle P, b(q_1q_2\ldots q_{|X|})b(q_{|X|+1}q_{|X|+2}\cdots q_{|Q|})\rangle = \sum_{XY=Q}\langle P, b(X)b(Y)\rangle\,, \tag{4.131}$$

leading to the conclusion that $\langle b(P), Q\rangle{=}\langle P, b(Q)\rangle$, finishing the proof. ☐

Assuming linearity $b(A + B) := b(A) + b(B)$, the expansion of $b(P)$ satisfies the shuffle symmetry $b(A \sqcup\!\sqcup B) = 0$ for $A, B \neq \emptyset$. We will prove this in two different ways.

Proposition 9. *The planar binary tree expansion $b(P)$ in (4.125) satisfies the shuffle symmetry*

$$b(A \sqcup\!\sqcup B) = 0\,, \qquad \forall\ A, B \neq \emptyset\,. \tag{4.132}$$

Proof 1. We will show this by induction on the length of the word in $b(P)$ starting from $b(1 \sqcup\!\sqcup 2) = b(12) + b(21)$, which is easy to verify. Assume that $b(A \sqcup\!\sqcup B) = 0$ for $|A|+|B| = k$, and consider $b(R \sqcup\!\sqcup S)$ for nonempty words such that $|R|+|S| = k+1$. The result will follow from the word identity (C.13), the antisymmetric nature of the deconcatenation in the definition of the b map (4.125), and the induction hypothesis. That is,

$$
\begin{aligned}
s_{RS} b(R \sqcup\!\sqcup S) &= \sum_{XY=R\sqcup\!\sqcup S} [b(X), b(Y)] \\
&= [b(\emptyset), b(R \sqcup\!\sqcup S)] + [b(R \sqcup\!\sqcup S), b(\emptyset)] + [b(R), b(S)] + [b(S), b(R)] \\
&\quad + \sum_{XY=R}{}' \sum_{ZW=S}{}' [b(X \sqcup\!\sqcup Y), b(Z \sqcup\!\sqcup W)]\,,
\end{aligned}
\tag{4.133}
$$

where we used the identity (C.13) to expand the deconcatenation sum in the second line. The second line vanishes by the antisymmetry of the Lie bracket, while the third line vanishes since $|X|+|Y|=|R| \leq k$ with nonempty X, Y implies, by the induction hypothesis, $b(X \sqcup\!\sqcup Y)=0$. Therefore $b(P\sqcup\!\sqcup R)=0$ for $P, R \neq \emptyset$. ☐

Proof 2. Recall Ree's theorem [167] that a Lie polynomial Γ is orthogonal to shuffles with non-empty words (see Theorem 3.1 (iv) in [143])

$$
\langle \Gamma, R \sqcup\!\sqcup S \rangle = 0\,, \quad R, S \neq \emptyset\,. \tag{4.134}
$$

Since $b(P)$ is a Lie polynomial by the definition (4.125) and b is self-adjoint by (4.127), we have

$$
0 = \langle b(P), R \sqcup\!\sqcup S \rangle = \langle P, b(R \sqcup\!\sqcup S) \rangle\,, \quad R, S \neq \emptyset\,, \quad \forall P\,, \tag{4.135}
$$

and the result follows. ☐

4.3.2. Berends–Giele currents from planar binary trees

Having the planar binary tree recursion (4.125), one can define Berends–Giele currents in BCJ gauge $\mathcal{K}_P$ or Lorenz gauge $\hat{\mathcal{K}}_P$ as

$$
\mathcal{K}_P = K_{b(P)}\,, \quad \hat{\mathcal{K}}_P = \hat{K}_{b(P)}\,, \tag{4.136}
$$

where $K_{b(P)}$ and $\hat{K}_{b(P)}$ are defined by linearity. We are here departing from the notation in section 4.2, where Berends–Giele current in Lorenz gauge were denoted by $\mathcal{K}_P$ or $\mathcal{K}^{\rm L}_P$ and those in BCJ gauge by $\mathcal{K}^{\rm BCJ}_P$. For example, with $K_{[P,Q]} = A^m_{[P,Q]}$ we get $\mathcal{A}^m_1 = A^m_1$ and

$$
\begin{aligned}
\mathcal{A}^m_{12} &= \frac{A^m_{[1,2]}}{s_{12}}\,, \\
\mathcal{A}^m_{123} &= \frac{A^m_{[[1,2],3]}}{s_{12}s_{123}} + \frac{A^m_{[1,[2,3]]}}{s_{123}s_{23}}\,, \\
\mathcal{A}^m_{1234} &= \frac{A^m_{[[[1,2],3],4]}}{s_{12}s_{123}s_{1234}} + \frac{A^m_{[[1,[2,3]],4]}}{s_{123}s_{1234}s_{23}} + \frac{A^m_{[[1,2],[3,4]]}}{s_{12}s_{1234}s_{34}} + \frac{A^m_{[1,[[2,3],4]]}}{s_{1234}s_{23}s_{234}} + \frac{A^m_{[1,[2,[3,4]]]}}{s_{1234}s_{234}s_{34}}\,, \\
\mathcal{A}^m_{12345} &= \frac{A^m_{[[[[1,2],3],4],5]}}{s_{12}s_{123}s_{1234}s_{12345}} + \frac{A^m_{[[[1,[2,3]],4],5]}}{s_{123}s_{1234}s_{12345}s_{23}} + \frac{A^m_{[[[1,2],[3,4]],5]}}{s_{12}s_{1234}s_{12345}s_{34}} + \frac{A^m_{[[[1,2],3],[4,5]]}}{s_{12}s_{123}s_{12345}s_{45}} \\
&\quad + \frac{A^m_{[[1,[[2,3],4]],5]}}{s_{1234}s_{12345}s_{23}s_{234}} + \frac{A^m_{[[1,[2,[3,4]]],5]}}{s_{1234}s_{12345}s_{234}s_{34}} + \frac{A^m_{[[1,[2,3]],[4,5]]}}{s_{123}s_{12345}s_{23}s_{45}} + \frac{A^m_{[[1,2],[[3,4],5]]}}{s_{12}s_{12345}s_{34}s_{345}} \\
&\quad + \frac{A^m_{[[1,2],[3,[4,5]]]}}{s_{12}s_{12345}s_{345}s_{45}} + \frac{A^m_{[1,[[[2,3],4],5]]}}{s_{12345}s_{23}s_{234}s_{2345}} + \frac{A^m_{[1,[[2,[3,4]],5]]}}{s_{12345}s_{234}s_{2345}s_{34}} + \frac{A^m_{[1,[[2,3],[4,5]]]}}{s_{12345}s_{23}s_{2345}s_{45}} \\
&\quad + \frac{A^m_{[1,[2,[[3,4],5]]]}}{s_{12345}s_{2345}s_{34}s_{345}} + \frac{A^m_{[1,[2,[3,[4,5]]]]}}{s_{12345}s_{2345}s_{345}s_{45}}\,.
\end{aligned}
\tag{4.137}
$$

Since these expansions will be frequently used later we also write the expansions of the Berends–Giele currents

$$
M_P := V_{b(P)} \tag{4.138}
$$

using this algorithm to get (4.126)

$$
\begin{aligned}
M_1 &= V_1\,, \quad M_{12} = \frac{V_{[1,2]}}{s_{12}}\,, \quad M_{123} = \frac{V_{[[1,2],3]}}{s_{12}s_{123}} + \frac{V_{[1,[2,3]]}}{s_{123}s_{23}}\,, \\
M_{1234} &= \frac{V_{[[[1,2],3],4]}}{s_{12}s_{123}s_{1234}} + \frac{V_{[[1,[2,3]],4]}}{s_{123}s_{1234}s_{23}} + \frac{V_{[[1,2],[3,4]]}}{s_{12}s_{1234}s_{34}} + \frac{V_{[1,[[2,3],4]]}}{s_{1234}s_{23}s_{234}} + \frac{V_{[1,[2,[3,4]]]}}{s_{1234}s_{234}s_{34}}\,, \\
M_{12345} &= \frac{V_{[[[[1,2],3],4],5]}}{s_{12}s_{123}s_{1234}s_{12345}} + \frac{V_{[[[1,[2,3]],4],5]}}{s_{123}s_{1234}s_{12345}s_{23}} + \frac{V_{[[[1,2],[3,4]],5]}}{s_{12}s_{1234}s_{12345}s_{34}} + \frac{V_{[[[1,2],3],[4,5]]}}{s_{12}s_{123}s_{12345}s_{45}} \\
&+ \frac{V_{[[1,[[2,3],4]],5]}}{s_{1234}s_{12345}s_{23}s_{234}} + \frac{V_{[[1,[2,[3,4]]],5]}}{s_{1234}s_{12345}s_{234}s_{34}} + \frac{V_{[[1,[2,3]],[4,5]]}}{s_{123}s_{12345}s_{23}s_{45}} + \frac{V_{[[1,2],[[3,4],5]]}}{s_{12}s_{12345}s_{34}s_{345}} \\
&+ \frac{V_{[[1,2],[3,[4,5]]]}}{s_{12}s_{12345}s_{345}s_{45}} + \frac{V_{[1,[[[2,3],4],5]]}}{s_{12345}s_{23}s_{234}s_{2345}} + \frac{V_{[1,[[2,[3,4]],5]]}}{s_{12345}s_{234}s_{2345}s_{34}} + \frac{V_{[1,[[2,3],[4,5]]]}}{s_{12345}s_{23}s_{2345}s_{45}} \\
&+ \frac{V_{[1,[2,[[3,4],5]]]}}{s_{12345}s_{2345}s_{34}s_{345}} + \frac{V_{[1,[2,[3,[4,5]]]]}}{s_{12345}s_{2345}s_{345}s_{45}}\,.
\end{aligned}
\tag{4.139}
$$

After using Baker's identity (4.56) to expand the nested brackets in the basis of $\ell(1Q)$ and adopting the notation (4.55), e.g., $V_{[[[1,2],3],4]} = V_{1234}$, these examples reproduce the Berends–Giele expansions of $\mathcal{K}_P \to \mathcal{A}^m_P, M_P$ given before in (4.81) and (4.83).

The proof of (4.132) shows that any antisymmetric deconcatenation will satisfy the shuffle symmetry, as this is a property obeyed by the underlying words. Hence, the Berends–Giele supercurrents, defined by their antisymmetric recursion (4.94), and the BRST-closed superfield E_P, defined by (4.86), both satisfy the shuffle symmetry. We therefore obtain the following corollary:

Corollary 2. *The Berends–Giele supercurrents $\mathcal{K}_P$ (4.80) and the BRST-closed superfield E_P (4.86) satisfy*

$$
\mathcal{K}_{R\sqcup\!\sqcup S} = E_{R\sqcup\!\sqcup S} = 0\,, \quad \forall\, R, S \neq \emptyset\,. \tag{4.140}
$$

4.3.3. The S bracket

BCJ relations for SYM amplitudes were expressed in [83, 165] using the so-called *S map* defined in [83] by its action on Berends–Giele currents. The properties of this map provided the motivation for a more general definition in [156, 157] as a bracket $\{\cdot,\cdot\}$, dubbed the *S bracket*, acting on words in the dual space of Lie polynomials $\mathcal{L}^*$ and producing words in the dual space $\mathcal{L}^*$, i.e. $\{\cdot,\cdot\} : \mathcal{L}^* \otimes \mathcal{L}^* \to \mathcal{L}^*$. For our purposes, this space is defined by the equivalence classes of words differing by proper shuffles, i.e.,

$$
A \sim B \ \text{ if } \ A = B + \textstyle\sum R \sqcup\!\sqcup S \ \text{ with } \ R, S \neq \emptyset\,. \tag{4.141}
$$

For instance $\{1,2\} \sim -\{2,1\}$ because $\{1,2\} = s_{12}12 = -s_{12}21 + s_{12}1 \sqcup\!\sqcup 2 \sim -\{2,1\}$. See the Appendix C.1 for more information.

There are several equivalent definitions of the S bracket [83, 156, 157]. A recursive definition for letters i, j and words A, B was given in [157] as

$$
\begin{aligned}
\{iAj, B\} &= i\{Aj, B\} - j\{iA, B\}\,, \\
\{B, iAj\} &= \{B, iA\}j - \{B, Aj\}i\,, \\
\{i, j\} &= s_{ij}\, ij\,.
\end{aligned}
\tag{4.142}
$$

Example applications are given by

$$
\begin{aligned}
\{1,2\} &= s_{12}12\,, \\
\{1,23\} &= s_{12}123 - s_{13}132\,, \\
\{12,3\} &= s_{23}123 - s_{13}213\,, \\
\{1,234\} &= s_{12}1234 - s_{13}1324 - s_{13}1342 + s_{14}1432\,,
\end{aligned}
\tag{4.143}
$$

$$\begin{aligned}\{123,4\} &= s_{34}1234 - s_{24}1324 - s_{24}3124 + s_{14}3214\,,\\ \{12,34\} &= s_{23}1234 - s_{24}1243 - s_{13}2134 + s_{14}2143\,.\end{aligned}$$

We note that the original definition of the S bracket in [83] is given in terms of a closed formula

$$\{P,Q\} = \sum_{\substack{XiY=P\\ RjS=Q}} k_i \cdot k_j (X \sqcup\!\sqcup \tilde{Y}) ij (\tilde{R} \sqcup\!\sqcup S)(-1)^{|Y|+|R|}\,, \tag{4.144}$$

which, in particular, yields the following for one-letter words $P \to i$

$$\{i,Q\} = \sum_{RjS=Q} k_i \cdot k_j ij (\tilde{R} \sqcup\!\sqcup S)(-1)^{|R|}\,. \tag{4.145}$$

Several properties of the S bracket were proven in [157]:

Proposition 10. *The S bracket satisfies:*

i. $\{A \sqcup\!\sqcup B, C\} = 0$ *for* $A, B \neq \emptyset$

ii. $\{\cdot,\cdot\}$ *is a Lie bracket in the space of dual Lie polynomials (4.141)*

iii. The binary tree map b of (4.125) acting on the S bracket satisfies [156]

$$b(\{P,Q\}) = [b(P), b(Q)]\,. \tag{4.146}$$

iv. $\sum_{XY=P}\{X,Y\} \sim s_P P$

The proofs of these statements can be found in [157], we will restrict ourselves to showcasing some examples. As a simple illustration of the shuffle property, we consider:

$$\{1 \sqcup\!\sqcup 2, 3\} = \{12,3\} + \{21,3\} = (s_{23}123 - s_{13}213) + (s_{13}213 - s_{23}123) = 0\,, \tag{4.147}$$

this is consistent with the fact that the S bracket operates on dual words (4.141), where $A \sqcup\!\sqcup B \sim 0$. For the Lie-bracket property we explicitly verify the simplest cases, antisymmetry for two letters and the Jacobi identity:

$$\{1,2\} + \{2,1\} = s_{12}12 + s_{21}21 = s_{12}(1 \sqcup\!\sqcup 2) \sim 0\,, \tag{4.148}$$

$$\begin{aligned}\{\{1,2\},3\} + \{\{2,3\},1\} + \{\{3,1\},2\} &= s_{12}s_{13}(3 \sqcup\!\sqcup 12 - 2 \sqcup\!\sqcup 13) + s_{12}s_{23}(23 \sqcup\!\sqcup 1 - 3 \sqcup\!\sqcup 21)\\ &\quad + s_{13}s_{23}(2 \sqcup\!\sqcup 31 - 32 \sqcup\!\sqcup 1) \sim 0\,.\end{aligned}$$

An example for the third property is given by

$$\begin{aligned}b(\{12,3\}) &= s_{23}b(123) - s_{13}b(213)\\ &= s_{23}\left(\frac{[[1,2],3]}{s_{12}s_{123}} + \frac{[1,[2,3]]}{s_{23}s_{123}}\right) - s_{13}\left(\frac{[[2,1],3]}{s_{12}s_{123}} + \frac{[2,[1,3]]}{s_{13}s_{123}}\right)\\ &= \frac{[[1,2],3]}{s_{12}} = [b(12), b(3)]\,,\end{aligned} \tag{4.149}$$

where we used the third example from (4.143), $s_{12}+s_{13}+s_{23} = s_{123}$ and the Jacobi identity. Note that there is no pole $1/s_{123}$ in the right-hand side of (4.149). An illustration of the fourth property is the following

$$\begin{aligned}\sum_{XY=123}\{X,Y\} &= \{1,23\} + \{12,3\} = s_{12}123 - s_{13}132 + s_{23}123 - s_{13}213\\ &= (s_{12}+s_{13}+s_{23})123 - s_{13}(2 \sqcup\!\sqcup 13) \sim s_{123}123\,.\end{aligned} \tag{4.150}$$

From the property (4.146) it is straightforward to conclude:

Corollary 3. *There is no $1/s_{PQ}$ propagator in $b(\{P,Q\})$, that is [157]*

$$\lim_{s_{PQ}\to 0} s_{PQ}b(\{P,Q\}) = 0\,. \tag{4.151}$$

BCJ amplitude relations. We see that the S bracket in $\{P, Q\}$ cancels the overall propagator $1/s_{PQ}$ from the linear combinations in $b(\{P, Q\})$. From what we have seen in (4.87), this condition implies that the superfield $E_{\{P,Q\}}$ is a BRST exact expression, $E_{\{P,Q\}} = QM_{\{P,Q\}}$ as the divergent propagator $1/s_{PQ}$ (in the context of an amplitude of $|P|+|Q|+1$ massless particles) is absent from $M_{\{P,Q\}}$. This only happens when the numerators satisfy the Jacobi identity, and this fact plays a key role in the proof of the BCJ amplitude relations using the cohomology of pure spinor superspace, see the discussion in section 5.2.5. A result we will need later is the following:

Lemma 5. *The S bracket in the special case when one of the words is a letter admits the form*

$$\{i, Q\} = \sum_{RS=Q} k_i \cdot k_S RiS\,. \tag{4.152}$$

Proof. We will show that the right-hand side of (4.152) is shuffle equivalent to the expression (4.145). Using the shuffle equivalence proven in (5.46) we get

$$\begin{aligned}\sum_{RS=Q} k_i \cdot k_S RiS &\sim \sum_{RS=Q} k_i \cdot k_S i(\tilde{R} \sqcup\!\sqcup S)(-1)^{|R|} \\ &\sim \sum_{RjkS=Q} k_i \cdot k_{kS} i(\widetilde{Rj} \sqcup\!\sqcup kS)(-1)^{|R|+1} \\ &\sim \sum_{RjkS=Q} \big[k_i \cdot k_{kS} ij(\tilde{R} \sqcup\!\sqcup kS)(-1)^{|R|+1} + k_i \cdot k_{kS} ik(\widetilde{Rj} \sqcup\!\sqcup S)(-1)^{|R|+1}\big]\,,\end{aligned} \tag{4.153}$$

where to arrive at the second line we relabeled the summation variables as $R \to Rj$ and $S \to kS$, and in the third line we used $(j\tilde{R}) \sqcup\!\sqcup (kS) = j(\tilde{R} \sqcup\!\sqcup kS) + k(j\tilde{R} \sqcup\!\sqcup S)$ from the definition of the shuffle product (C.5). Now we relabel $kS \to S$ in the first sum in the right-hand side of (4.153) and $Rj \to R$, $k \to j$ in the second one (so that $(-1)^{|R|+1} \to (-1)^{|R|}$). This implies that

$$\begin{aligned}\sum_{RS=Q} k_i \cdot k_S RiS &\sim \sum_{RjS=Q} -k_i \cdot k_S ij(\tilde{R} \sqcup\!\sqcup S)(-1)^{|R|} + \sum_{RjS=Q} k_i \cdot k_{jS} ij(\tilde{R} \sqcup\!\sqcup S)(-1)^{|R|} \\ &\sim \sum_{RjS=Q} k_i \cdot k_j ij(\tilde{R} \sqcup\!\sqcup S)(-1)^{|R|}\,,\end{aligned} \tag{4.154}$$

which is the same expression as (4.145), finishing the proof. □

4.3.4. The contact-term map and the S bracket

The S bracket is intimately related to the contact-term map discussed in section 4.1.1. In fact, the recursive definition of the contact-term map (4.21) admits an equivalent representation in terms of the S bracket. If Γ is a Lie monomial, then [157]

$$\langle P \otimes Q, C(\Gamma)\rangle = \langle \{P, Q\}, \Gamma\rangle\,, \tag{4.155}$$

where the scalar product of words $\langle A, B\rangle$ takes values 1 for $A = B$ and 0 for $A \neq B$, see (C.11). This means that the adjoint C^* of the contact-term map is the S bracket; $C^*(A \otimes B) = \{A, B\}$. Exploiting this interpretation allows one to prove that $C^2 = 0$ as stated earlier in (4.27), see Appendix H.

Important properties of the contact-term map relevant to the of description of SYM in terms of local and non-local multiparticle superfields were proven in [85, 157]. For instance, the contact-term map deconcatenates the planar binary tree map involving pole cancellations in a highly non-trivial manner:

Lemma 6. *The contact-term map (4.155) satisfies*

$$C\big(b(P)\big) = \sum_{XY=P} \big(b(X) \otimes b(Y) - b(Y) \otimes b(X)\big) := \sum_{XY=P} b(X) \wedge b(Y)\,, \tag{4.156}$$

where $A \wedge B = A \otimes B - B \otimes A$.

Proof. From the characterization (4.155) as the adjoint of the S bracket we obtain

$$\begin{aligned}
\langle R\otimes S, C(b(P))\rangle &= \langle\{R,S\}, b(P)\rangle = \langle b(\{R,S\}), P\rangle \\
&= \langle [b(R), b(S)], P\rangle = \langle b(R)b(S), P\rangle - (R\leftrightarrow S) \\
&= \sum_{XY=P}\langle b(R), X\rangle\langle b(S), Y\rangle - (R\leftrightarrow S) \\
&= \sum_{XY=P}\langle R, b(X)\rangle\langle S, b(Y)\rangle - (X\leftrightarrow Y) \\
&= \sum_{XY=P}\langle R\otimes S, \big(b(X)\otimes b(Y) - (X\leftrightarrow Y)\big)\rangle\,,
\end{aligned}\tag{4.157}$$

where in the first line we used that b is self adjoint (4.127), in the second we used the property (4.146), and (4.129) in the third. In the fourth line we used the self-adjoint property of the b map again and finally the definition $\langle A\otimes B, R\otimes S\rangle = \langle A, R\rangle\langle B, S\rangle$ in the last line. Since R, S are arbitrary, the result follows. $\square$

For example, the simple identity $C([1,2]) = s_{12} 1\wedge 2$ leads to the deconcatenation of $b(12) = \frac{[1,2]}{s_{12}}$ (note $b(\emptyset) := 0$)

$$C\big(b(12)\big) = b(1)\wedge b(2) = \sum_{XY=12} b(X)\wedge b(Y)\,. \tag{4.158}$$

However, it is already non-trivial to explicitly check using $C([[1,2],3])$ and $C([1,[2,3]])$ given in (4.23) that $C(b(123)) = b(12)\wedge b(3) + b(1)\wedge b(23)$ with $b(123)$ given in (4.126).

In [157, 180], the definition (4.155) was used to show that the S bracket is in fact a *Lie cobracket* as defined in the context of Lie coalgebras [181].

4.3.5. The KLT map

The KLT relation was derived in [182] as a way to express the closed-string tree-level amplitude as a sum over products of color-ordered open-string tree amplitudes, see section 7.2.1 for a brief review. In the field-theory limit $\alpha'\to 0$, it readily implies the same type of squaring relations between n-point supergravity amplitudes $M_n^{\rm grav}$ and color-ordered SYM amplitudes $A(\ldots)$. In a modern language, the field-theory KLT relation can be written as

$$M_n^{\rm grav} = -\sum_{P,Q} A(1,P,n,n{-}1)S(P|Q)_1\tilde A(1,Q,n{-}1,n)\,, \tag{4.159}$$

where the $\tilde A(\ldots)$ feature polarizations $\tilde e_i, \tilde\chi_i$ independent from the e_i, χ_i in $A(\ldots)$ and $S(P|Q)_1$ is the *KLT matrix* or the *momentum kernel* indexed by permutations $P, Q\in S_{n-3}$ of legs $2,3,\ldots,n{-}2$. In a series of papers [183, 184, 185, 186] the algorithm to obtain the precise form of the KLT matrix was sequentially simplified to a recursive definition [187]

$$S(Aj|BjC)_i = k_j\cdot k_{iB}\, S(A|BC)_i\,, \qquad S(\emptyset|\emptyset)_i := 1\,, \tag{4.160}$$

where i is some fixed leg, conventionally chosen $i=1$. For example,

$$\begin{aligned}
S(2|2)_1 &= k_1\cdot k_2\,, & S(23|23)_1 &= (k_3\cdot k_{12})(k_1\cdot k_2)\,,\\
S(23|32)_1 = S(32|23)_1 &= (k_1\cdot k_2)(k_1\cdot k_3)\,, & S(32|32)_1 &= (k_2\cdot k_{13})(k_1\cdot k_3)\,.
\end{aligned}\tag{4.161}$$

In the framework of twisted deRham theory, the entries of the inverse KLT matrix have been interpreted as intersection numbers [188, 189, 190], see section 7.3.2 for further details.

In recent years, the KLT matrix has been found in various relations involving the computation of string scattering amplitudes. These relations can often be understood from a combinatorial/free-Lie-algebra perspective, usually intimately related to planar binary trees. For instance, we will see that the expression relating local multiparticle superfields V_P in the BCJ gauge and Berends–Giele currents in (4.182) descends

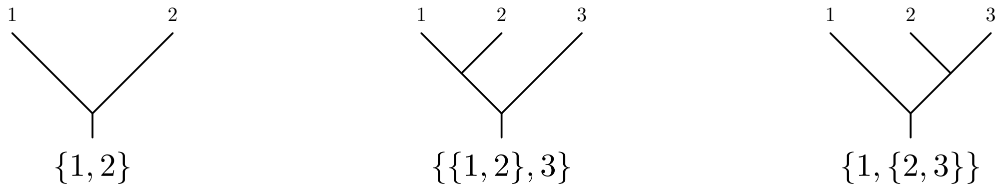


Figure 3: Examples of the KLT map (4.164), where $\{\cdot,\cdot\}$ is the S bracket (4.142). Each planar binary tree is mapped to the expressions given in (4.163).

from the free-Lie-algebra relation (4.180) below. In addition, the KLT matrix also plays a major role in the integration-by-parts identities used in the derivation of a closed formula for the massless n-point superstring disk amplitude in section 6.3.1. Moreover, the KLT matrix is the inverse of the Berends–Giele double currents from which the tree-level amplitudes of the bi-adjoint scalar theory are calculated. These in turn are related to the field-theory limit of the superstring disk integrals which, as we will see in section 6.4.4, admit a combinatorial interpretation. In summary, the KLT matrix is indeed a central player connecting various objects participating in the calculation of string scattering amplitudes.

Generalized KLT matrix. In the pursuit of a combinatorial framework for understanding the standard KLT matrix and its relations to multiparticle superfields, a *generalized KLT matrix* has been proposed in [156] and analyzed further in [157, 180] (see also [191]). To see this more precisely, one defines a map that converts every Lie bracket $[\cdot,\cdot]$ in an arbitrary Lie monomial Γ to a S bracket $\{\cdot,\cdot\}$. This conversion is denoted by $\{\Gamma\}$

$$\{\Gamma\} : [\cdot,,\cdot] \to \{\cdot,\cdot\}\,, \tag{4.162}$$

and acts recursively in commutator depth, transforming a Lie polynomial to a dual Lie polynomial (see Appendix C for the definitions). For example,

$$\begin{aligned} \{[1,2]\} &= \{1,2\} = s_{12}12\,, \\ \{[[1,2],3]\} &= \{\{1,2\},3\} = s_{12}\big(s_{23}123 - s_{13}213\big)\,, \\ \{[1,[2,3]]\} &= \{1,\{2,3\}\} = s_{23}\big(s_{12}123 - s_{13}132\big)\,, \\ \{[[1,2],[3,4]]\} &= \{\{1,2\},\{3,4\}\} = s_{12}s_{34}\big(s_{23}1234 - s_{24}1243 - s_{13}2134 + s_{14}2143\big)\,. \end{aligned} \tag{4.163}$$

The KLT map is defined as a map between planar binary trees Γ and its S bracket version

$$S : \Gamma \to \{\Gamma\}\,. \tag{4.164}$$

A graphical illustration of the KLT map is given in the figure 3.

The matrix elements of the KLT map with respect to a basis of Lie monomials Γ,Σ are given by $S(\Gamma,\Sigma) = \langle\{\Gamma\},\Sigma\rangle$ which motivate the definition of the *generalized KLT matrix* for words P and Q [156]

$$S^\ell(P|Q) = \langle \ell\{P\}, \ell(Q)\rangle\,, \tag{4.165}$$

where the dual Dynkin bracket $\ell\{P\}$ is defined as the conversion (4.162) of the Dynkin bracket (4.49),

$$\ell\{P\} := \{\ell(P)\} \tag{4.166}$$

with $\ell(P)$ defined in (4.49) and $\{\Gamma\}$ defined in (4.163). Alternatively, a recursive definition is given by

$$\ell\{123\ldots n\} = \{\ell(123\ldots n{-}1), n\}\,, \qquad \ell\{i\} = i\,, \qquad \ell\{\emptyset\} = 0\,. \tag{4.167}$$

The simplest examples of (4.165) include $S^\ell(12|12) = s_{12}$, and $S^\ell(12|21) = -s_{12}$, as well as

$$S^\ell(123|123) = s_{12}(s_{13}+s_{23})\,, \qquad S^\ell(1234|1234) = s_{12}(s_{13}+s_{23})(s_{14}+s_{24}+s_{34})\,,$$

$$
\begin{aligned}
&S^\ell(132|123) = s_{12}s_{13}\,, && S^\ell(1243|1234) = s_{12}(s_{13}+s_{23})(s_{14}+s_{24})\,,\\
&S^\ell(312|123) = -s_{12}s_{13}\,, && S^\ell(3412|1234) = -s_{12}s_{13}s_{34}\,,\\
&S^\ell(231|123) = -s_{12}s_{23}\,, && S^\ell(3421|1234) = s_{12}s_{23}s_{34}\,,
\end{aligned}
\tag{4.168}
$$

and it is shown in [157] that $S^\ell(P|Q)$ is a symmetric matrix, $S^\ell(P|Q) = S^\ell(Q|P)$, see the reference for further details. One of the main attributes of the above generalized KLT matrix $S^\ell(P|Q)$ is that it satisfies generalized Jacobi identities in both P and Q,

$$
S^\ell(A\ell(B)|Q) + S^\ell(B\ell(A)|Q) = 0\,, \quad \forall\, Q\,. \tag{4.169}
$$

For example, with $A = 12$ and $B = 34$

$$
S^\ell(1234|Q) - S^\ell(1243|Q) + S^\ell(3412|Q) - S^\ell(3421|Q) = 0\,, \quad \forall\, Q\,. \tag{4.170}
$$

Moreover, the standard KLT matrix recursion (4.160) is obtained as the special case when the first letters in both words coincide [156, 157],

$$
S^\ell(iP|iQ) = S(P|Q)_i\,. \tag{4.171}
$$

The unrestricted nature of the first letters in the permutations of the matrix (4.165) is the motivation for the qualifier *generalized*.

An important interplay between the Lie-bracket conversion (4.162) from the space of Lie polynomials to the space of dual Lie polynomials is summarized in:

Lemma 7. *The planar binary tree map (4.125) satisfies*

$$
b(\ell\{P\}) = \ell(P)\,, \tag{4.172}
$$

where the dual Dynkin bracket is given by (4.167) and the Dynkin bracket by (4.49). In fact,

$$
b(\{\Gamma\}) = \Gamma \tag{4.173}
$$

for any Lie polynomial Γ.

Proof. Using the property (4.146) together with the recursive definition (4.167) yields

$$
\begin{aligned}
b(\ell\{123\dots n\}) &= b(\{\ell\{123\dots n{-}1\}, b(n)\})\\
&= [b(\ell\{123\dots n{-}1\}), n] = \dots\\
&= [[\dots[[1,2],3],\dots], n] = \ell(123\dots n)\,,
\end{aligned}
\tag{4.174}
$$

where we used that $b(j) = j$ for a letter j. Since $\ell(P)$ is a basis of Lie polynomials, the result (4.173) follows by a basis expansion. $\square$

In effect, as $\ell\{P\} = \{\ell(P)\}$ implies $\ell\{P\} = S(\ell(P))$ in terms of the KLT map (4.164), the result (4.173) is the $b \circ S = Id$ part of a more general statement proven in [157]:

Proposition 11. *The planar binary tree map* $b : \mathcal{L}^* \to \mathcal{L}$ *defined in (4.125) and the KLT map* $S : \mathcal{L} \to \mathcal{L}^*$ *defined in (4.164) are inverses to each other,*

$$
\begin{aligned}
b \circ S : \mathcal{L} &\to \mathcal{L}\,, & S \circ b : \mathcal{L}^* &\to \mathcal{L}^*\,,\\
\Gamma &\mapsto \Gamma\,, & P^* &\mapsto P^*\,.
\end{aligned}
\tag{4.175}
$$

Correspondence between multiparticle superfields and free Lie algebra. Over the years, it became clear that relations governing multiparticle superfields discovered in pursuit of expressions for string amplitudes had a combinatorial flavor of the type commonly studied within the free-Lie-algebra framework. This is particularly true in the context of the color-kinematics duality [29], where the generalized Jacobi identities played a major role in the simple form of the general massless disk amplitude of [21].

The correspondence suggested above can be made precise with the following mapping between free-Lie-algebra structures on one side and multiparticle superfields in the pure spinor formalism on the other[24]:

$$C \leftrightarrow Q_{\rm BRST}\,, \quad \ell(P) \leftrightarrow V_P\,, \quad b(P) \leftrightarrow M_P\,, \tag{4.176}$$

where the Dynkin bracket $\ell(P)$ is defined in (C.1). That is, the contact-term map C is identified with the BRST charge Q as already hinted in (4.28), the Lie monomials encoded in the Dynkin bracket $\ell(P)$ correspond to the multiparticle unintegrated vertex V_P in the BCJ gauge, and the planar binary tree expansion $b(P)$ corresponds to the Berends–Giele current M_P as in (4.138). As an immediate consistency check, note that both C and Q are nilpotent, see (4.27) and (3.25). In addition, the symmetries on both sides agree: generalized Jacobi identities (4.48) for both $\ell(P)$ and V_P as well as shuffle symmetries for both the planar binary tree expansion $b(P)$ in (4.132) and the Berends–Giele currents M_P in (4.140) (or more generally $\mathcal{K}_P$).

For a more precise relation, the identities (4.29) and (4.72) illustrate the correspondence between V_P and $\ell(P)$ as well as the contact-term map C and the BRST charge Q. For example, note the parallels of these equations at multiplicity three:

$$\begin{aligned} C(\ell(123)) &= (k_1\cdot k_2)\big(\ell(1)\wedge\ell(23)+\ell(13)\wedge\ell(2)\big)+(k_{12}\cdot k_3)\ell(12)\wedge\ell(3)\,, \\ QV_{123} &= (k_1\cdot k_2)\big(V_1V_{23}+V_{13}V_2\big)+(k_{12}\cdot k_3)V_{12}V_3\,, \end{aligned} \tag{4.177}$$

where the fermionic property $V_PV_Q=-V_QV_P$ is mapped to the antisymmetric wedge product $\wedge$ of (4.24). Moreover, using the notation defined in (4.31), the BRST variation of the unintegrated vertex operator V_Γ for an arbitrary Lie monomial Γ can be written as

$$QV_\Gamma = (V\wedge V)_{C(\Gamma)}\,, \tag{4.178}$$

which is extended to arbitrary Lie polynomials by linearity.

The precise cancellations between the contact terms in the equations of motion for local superfields and the Mandelstam propagators featured in the definition of Berends–Giele currents constituted an early indication of a beautiful and rigorous underlying mathematical framework. See, for example, the discussions in [156] and several proofs in [157]. For instance, the proof that the Berends–Giele current M_P deconcatenates under the action of the BRST charge over its local numerators V_Q easily follows in a free-Lie-algebra setting,

Lemma 8. *The Berends–Giele current M_P in the BCJ gauge satisfies $QM_P=\sum_{XY=P}M_XM_Y$.*

Proof. The Berends–Giele current M_P is given by an expansion of local multiparticle vertices V_R encoded in terms of planar binary trees as $M_P=V_{b(P)}$ given in (4.138). The deconcatenation property of $C(b(P))$ in (4.156) implies

$$QM_P = QV_{b(P)} = (V\wedge V)_{C(b(P))} = \sum_{XY=P}(V\wedge V)_{b(X)\wedge b(Y)} = \sum_{XY=P}V_{b(X)}V_{b(Y)} = \sum_{XY=P}M_XM_Y\,, \tag{4.179}$$

where we used the notation (4.31). □

This proof sheds light on the deconcatenation property (4.85) of the Berends–Giele current from a different perspective compared with the equations of motion (4.97) derived from the perturbiner expansion. The result arises from the use of the equations of motion of the local multiparticle superfields yielding

[24] The generalizations $\ell(P)\leftrightarrow K_P$ and $b(P)\leftrightarrow\mathcal{K}_P$ are immediate, where K_P and $\mathcal{K}_P$ are defined in (4.1) and (4.80), respectively. We chose the representatives V_P and M_P for pedagogical reason.

contact terms that cancel the propagators present in the planar-binary-tree expansion, demonstrating that the patterns observed in (4.84) hold to all orders. The other equations of motion for the currents in $\mathcal{K}_P$ can be derived from their local counterparts in a similar fashion [85].

Finally, the relation between the local multiparticle superfields V_P satisfying generalized Jacobi identities and the non-local Berends–Giele supercurrents M_P in BCJ gauge satisfying shuffle symmetries follows from the identity proven in [157]:

Lemma 9. *The Dynkin bracket (4.49) and the planar-binary-tree expansion (4.125) are related by*

$$\ell(R) = \sum_Q S^\ell(R|iQ) b(iQ) \,, \tag{4.180}$$

where $S^\ell(R|iQ)$ is the generalized KLT matrix (4.165).

Proof. Consider the dual Dynkin bracket $\ell\{R\}$. Since it is a dual Lie polynomial it can be expanded in the Lyndon basis iQ of the dual Lie polynomials using the formula (C.16)

$$\ell\{R\} = \sum_Q \langle \ell\{R\}, \ell(iQ)\rangle iQ \,. \tag{4.181}$$

Acting with the b map (4.125) on both sides gives $b(\ell\{R\}) = \sum_Q \langle \ell\{R\}, \ell(iQ)\rangle b(iQ)$. The left-hand side can be rewritten using (4.172), while the scalar product on the right-hand side is the definition of the generalized KLT matrix. Thus, $\ell(R) = \sum_Q S^\ell(R|iQ) b(iQ)$. □

After setting $R = iP$ and using the definition (4.171) the generalized KLT matrix reduces to the usual matrix $S(P|Q)_i$ in (4.160). Replacing $\ell(iP) \to V_{iP}$ and $b(iQ) \to M_{iQ}$ as suggested by the correspondence (4.176) leads to the relation between Berends–Giele currents and multiparticle unintegrated vertex operators:

$$V_{iP} = \sum_Q S(P|Q)_i M_{iQ} \,. \tag{4.182}$$

This identity was first explicitly mentioned in [166], but it had already played an implicit role in the derivation of the closed formula of the massless n-point open-superstring amplitude in [21], see section 6.3.1. The inverse of (4.182) expressing M_{iP} as a linear combination of V_{iQ} will be given in (6.103).

The relation (4.182) can be straightforwardly adapted to reproduce the local multiparticle superfields (see section 4.1.6) from their respective Berends–Giele currents in BCJ gauge,

$$K_{iP} = \sum_Q S(P|Q)_i \mathcal{K}_{iQ} \,. \tag{4.183}$$

However, plugging the Berends–Giele currents in Lorenz gauge into the right-hand side of (4.182) and (4.183) will lead to a non-local outcome of the $S(P|Q)_i$ multiplication.[25]

5. SYM tree amplitudes from the cohomology of pure spinor superspace

In this section we are going to review how to obtain supersymmetric expressions for SYM tree-level amplitudes. One could in principle start by computing the n-point superstring disk amplitudes and then take its $\alpha' \to 0$ limit [147, 148, 149, 150]. However, the construction in this section relies entirely on pure spinor cohomology considerations [160], following the ideas of [159], and predates the calculation of the

[25]This can be anticipated from the alternative form $\mathcal{H}_{123} = \frac{H_{1,2,3}}{3s_{123}}(\frac{1}{s_{12}} - \frac{1}{s_{23}})$ of the gauge transformation (4.110) due to total antisymmetry of $\hat{H}_{[12,3]} = \frac{1}{3}H_{1,2,3}$. The KLT matrix renders the difference between the Lorenz- and BCJ-gauge variants of (4.183) for $\mathcal{A}^{123}_\alpha$ and $\mathcal{A}^m_{123}$ proportional to $S(23|23)_1\mathcal{H}_{123} + S(23|32)_1\mathcal{H}_{132} = \frac{H_{1,2,3}}{3s_{123}}(s_{13}+s_{23}-2s_{12})$ and therefore non-local.

n-point superstring disk amplitude in [21, 192]. The alternative derivation of SYM amplitudes from the $\alpha' \to 0$ limit of string amplitude was done later, see section 6.5.

Inspired by progress in organizing string amplitudes, it was realized in the 1980's that gauge-theory amplitude calculations simplify tremendously by considering ordered gauge invariants depending only upon kinematics – called color-ordered or color-stripped partial amplitudes [193, 194]. The full color-dressed S-matrix elements could be obtained by summing over a product of these color-ordered amplitudes with appropriate color-weights, either somewhat redundantly in a trace basis, or more efficiently in the Del Duca–Dixon–Maltoni basis of [195]. The advantages in considering stripped or ordered partial amplitudes are enormous; they grow exponentially rather than factorially in local diagram contributions. Thus here we only consider color-ordered amplitudes.

The pure spinor cohomology formula for n-point SYM tree-level amplitudes turns out to be the supersymmetrization of the standard Berends–Giele recursion relations [27], as one might have correctly suspected from the discussion of the Berends–Giele currents in the previous section. Therefore we will first review the recursive method proposed by Berends and Giele to compute Yang–Mills amplitudes.

5.1. Berends–Giele recursion relations

In the 80s, Berends and Giele proposed a recursive method to compute color-ordered gluon amplitudes at tree level with the formula [27]

$$A^{\mathrm{YM}}(1,2,\ldots,p,p{+}1) = s_{12\ldots p} J^m_{12\ldots p} J^m_{p+1} \,. \tag{5.1}$$

The *Berends–Giele currents* J^m_P are defined recursively in the number of external particles starting with the polarization vector e^m_i of a single-particle gluon, by (note $J^m_\emptyset := 0$)

$$\begin{aligned} J^m_i := e^m_i \,, \qquad s_P J^m_P &:= \sum_{XY=P} [J_X, J_Y]^m + \sum_{XYZ=P} \{J_X, J_Y, J_Z\}^m \,, \\ [J_X, J_Y]^m &:= (k_Y\cdot J_X) J^m_Y + \frac{1}{2} k^m_X (J_X \cdot J_Y) - (X \leftrightarrow Y) \,, \\ \{J_X, J_Y, J_Z\}^m &:= (J_X\cdot J_Z) J^m_Y - \frac{1}{2}(J_X\cdot J_Y) J^m_Z - \frac{1}{2}(J_Y\cdot J_Z) J^m_X \,, \end{aligned} \tag{5.2}$$

where the brackets $[\cdot,\cdot]^m$ and $\{\cdot,\cdot,\cdot\}^m$ are given by stripping off one gluon field (with vector index m) from the cubic and quartic vertices of the Yang–Mills Lagrangian. The deconcatenation of the word P into non-empty words X and Y is denoted by $\sum_{XY=P}$, with obvious generalization to $\sum_{XYZ=P}$, and $P = 12\ldots p$ encompasses several external particles, see section 1.3 for more details on the notation. In addition, the Mandelstam invariants s_P and multiparticle momenta k^m_P are defined as in (3.107) and (4.9).

In [27] the Berends–Giele currents J^m_P were shown to be conserved

$$k^m_P J^m_P = 0 \,, \tag{5.3}$$

which can alternatively be understood as the statement that the currents are in the Lorenz gauge [84].

As the simplest example of the recursion (5.2), the Berends–Giele current of multiplicity two is,

$$s_{12} J^m_{12} = e^m_2 (e_1\cdot k_2) - e^m_1 (e_2 \cdot k_1) + \frac{1}{2}(k^m_1 - k^m_2)(e_1\cdot e_2) \,, \tag{5.4}$$

and leads to the well-known three-point amplitude,

$$A^{\mathrm{YM}}(1,2,3) = s_{12} J^m_{12} J^m_3 = (e_1\cdot e_2)(k_1\cdot e_3) + \mathrm{cyc}(123) \,, \tag{5.5}$$

whose manifestly cyclic form is attained after using momentum conservation $k_1{+}k_2{+}k_3 = 0$ and transversality $e_j \cdot k_j = 0$. Higher-point amplitudes are generated by a straightforward application of the recursion (5.2). The multiplicity-three current

$$s_{123} J^m_{123} = [J_{12}, J_3]^m + [J_1, J_{23}]^m + \{J_1, J_2, J_3\}^m \tag{5.6}$$

gives rise to the four-point amplitude $A^{\rm YM}(1,2,3,4) = s_{123}J^m_{123}e^m_4$ and so forth. These recursion relations are a very efficient method to calculate tree amplitudes numerically, see e.g. [196, 155]. While the Berends–Giele formula (5.1) is not supersymmetric – it computes purely gluonic amplitudes – its supersymmetrization via uplift to pure spinor superspace will be given below.

5.1.1. Kleiss–Kuijf amplitude relations

A crucial identity satisfied by the currents was also demonstrated by Berends and Giele in [175], the shuffle symmetries:

$$J^m_{A\shuffle B} = 0, \quad \forall A, B \neq \emptyset\,. \tag{5.7}$$

Together with the amplitude formula (5.1), the shuffle identity (5.7) can be used to demonstrate that the Yang–Mills tree amplitudes satisfy the Kleiss–Kuijf (KK) relations [28]

$$A^{\rm YM}(P,1,Q,n) = (-1)^{|P|}A^{\rm YM}(1,\tilde P\shuffle Q,n)\,. \tag{5.8}$$

To see this one exploits the mathematical literature of free Lie algebras [167, 176]. More precisely, Ree's theorem [167] shows that a necessary and sufficient condition for a series of the form $\sum_{n>0} J_{i_1i_2\ldots i_n}X^{i_1}X^{i_2}\ldots X^{i_n}$ with non-commutative indeterminates X^i to be a Lie polynomial is the shuffle symmetry (5.7) of its coefficients. In the context of Yang–Mills tree-level amplitudes, the X^{i_j} are gauge-group generators which have to conspire to Lie polynomials and therefore contracted structure constants f^{abc} by Yang–Mills Feynman rules. Corollary 2.4 of [176] in turn states that $\sum_{n>0} J_{i_1i_2\ldots i_n}X^{i_1}X^{i_2}\ldots X^{i_n}$ is a Lie polynomial if and only if

$$J_{PiQ} = (-1)^{|P|}J_{i\tilde P\shuffle Q}\,. \tag{5.9}$$

Since both results (5.7) and (5.9) are "if and only if" statements, they must be equivalent (for a proof of this, see (5.45)). As a corollary of this equivalence together with the Berends–Giele amplitude formula (5.1) and the shuffle relation (5.7), it follows that the KK relation (5.8) must be satisfied. An alternative proof of the KK relations appears in [195].

We are now going to recover the Berends–Giele recursion (5.2) and the amplitude formula (5.1) from the bosonic components of a supersymmetric formula for SYM tree amplitudes derived from pure spinor cohomology considerations.

5.2. The pure spinor superspace formula for SYM tree amplitudes

In this subsection we will review the derivation of the recursive method in pure spinor superspace for the computation of supersymmetric tree amplitudes of ten-dimensional SYM theory. The method first appeared in [160] and it relies on the simple cohomology properties of multiparticle superfields in pure spinor superspace as suggested earlier in [159]. The end result is a method based on the recursive nature of the BRST variations of the supersymmetric Berends–Giele currents (4.84). When truncated to its bosonic components, the pure spinor formula was later shown in [165] to reproduce the standard Berends–Giele gluonic formula of [27].

However, there are certain beneficial novelties in the pure spinor approach worth highlighting:

- the direct derivation of multiple currents for each type of superfield $\mathbb{K} \in \{\mathbb{A}_\alpha, \mathbb{A}_m, \mathbb{W}^\alpha, \mathbb{F}^{mn}\}$ using their non-linear equations of motion,
- the usage of only cubic interactions as a result of identifying the natural superfields in the recursion[26] – the quartic interactions appear due to the quadratic terms in the field strength,
- the structural relation to planar binary trees and the construction using local numerators,
- the derivation of the shuffle symmetry of the currents using free-Lie-algebra methods,
- the identification of the different gauges associated to these local numerators and the subsequent derivation of local BCJ-satisfying numerators.

[26]See [197] for a reformulation of the purely gluonic Berends–Giele currents using cubic vertices.

Reasons for the cohomology method. In intermediate states of the calculations, the prescription to compute disk amplitudes in the pure spinor formalism yields a superspace expression containing three pure spinors and the superfields of ten-dimensional SYM as well as their covariant derivatives; in other words they constitute pure spinor superspace expressions. Superstring theory dictates that the field-theory SYM tree-level amplitudes must be recovered in the $\alpha' \to 0$ limit of the disk amplitudes, which are obtained from multiple OPEs among the vertex operators of schematic form $V_1 U_2 \ldots U_{n-2} V_{n-1} V_n$, see (3.75). Given that SYM tree-level amplitudes are supersymmetric and gauge invariant, we know from [1] that their corresponding pure spinor expressions must be in the BRST cohomology. It is important that the amplitudes are left written in *pure spinor superspace* since integrating out the pure spinors and the fermionic theta variables via (3.80) would lead to a multitude of terms in polarizations and momenta where all the simple superspace patterns are no longer present [89].

Let us first review the explicit superstring calculations for low multiplicities that led to the general method for arbitrary multiplicities, using the latest conventions for the notations.

Explicit results and the birth of the cohomology method. At low multiplicities the SYM amplitudes written in pure spinor superspace were obtained from the field-theory limit of the corresponding superstring disk amplitudes: the three-point case was known from the very start [1] while the four- and five-point amplitudes were computed in pure spinor superspace in [158, 109]:

$$
\begin{aligned}
A(1,2,3) &= \langle V_1 V_2 V_3 \rangle \,, \\
A(1,2,3,4) &= \frac{1}{s_{12}} \langle V_{12} V_3 V_4 \rangle + \frac{1}{s_{23}} \langle V_1 V_{23} V_4 \rangle \,, \\
A(1,2,3,4,5) &= \frac{\langle V_{123} V_4 V_5 \rangle}{s_{12} s_{45}} + \frac{\langle V_{321} V_4 V_5 \rangle}{s_{23} s_{45}} + \frac{\langle V_{12} V_{34} V_5 \rangle}{s_{12} s_{34}} + \frac{\langle V_1 V_{234} V_5 \rangle}{s_{23} s_{51}} + \frac{\langle V_1 V_{432} V_5 \rangle}{s_{34} s_{51}} \,.
\end{aligned}
\tag{5.10}
$$

Furthermore, using pure spinor cohomology arguments, expressions for the six- and seven-point SYM tree amplitudes in pure spinor superspace were proposed in [159]. The six-point amplitude was later reproduced from the field-theory limit of the pure spinor superstring disk amplitude in [161][27]. In possession of these results a general pattern was discovered in [160] in terms of the Berends–Giele currents written with multiparticle versions of the unintegrated vertices V_P in the BCJ gauge[28].

The clue was to notice the composing factors in the above amplitudes satisfied a regular pattern under BRST variation (note $s_{45} = s_{123}$ at five points)

$$
QV_1 = 0 \,, \qquad Q\frac{V_{12}}{s_{12}} = V_1 V_2 \,, \qquad Q\Big(\frac{V_{123}}{s_{12}s_{123}} + \frac{V_{321}}{s_{23}s_{123}}\Big) = V_1 \frac{V_{23}}{s_{23}} + \frac{V_{12}}{s_{12}} V_3 \,. \tag{5.11}
$$

These early computations together with six-points examples not shown led to the pattern of the Berends–Giele currents in (4.84). In addition, the the Catalan numbers govern both the number of cubic graphs in a color-ordered tree amplitude and the number of kinematic poles in the Berends–Giele currents M_P as derived in (5.19), so the assumption was that tree amplitudes would be composed of M_P.

Given that the SYM tree amplitudes are the $\alpha' \to 0$ field-theory limit of the superstring [147, 148, 149, 150], whose correlator in the pure spinor formalism is in the cohomology of the BRST charge, the proposal of [160] was based on finding a superfield expression in the cohomology of the BRST charge that was constructed using the Berends–Giele currents M_P. Rewriting the low-multiplicity examples (5.10) as

$$
\begin{aligned}
A(1,2,3) &= \langle M_1 M_2 M_3 \rangle \,, \\
A(1,2,3,4) &= \langle M_{12} M_3 M_4 \rangle + \langle M_1 M_{23} M_4 \rangle \,, \\
A(1,2,3,4,5) &= \langle M_{123} M_4 M_5 \rangle + \langle M_{12} M_{34} M_5 \rangle + \langle M_1 M_{234} M_5 \rangle
\end{aligned}
\tag{5.12}
$$

[27] The prior computations for five and six external bosons in the RNS formalism were performed in [198] and [199], respectively, with considerably longer expressions in their final results.

[28] It was later understood in [84] that the construction of Berends–Giele currents does not require any particular gauge of the associated local superfields, so the requirement of BCJ gauge in [160] was stronger than necessary.

not only simplifies their presentation but also suggests the n-point generalization [160]

$$A(1,2,\ldots,n) = \sum_{j=1}^{n-2} \langle M_{12\ldots j} M_{j+1\ldots n-1}\, M_n \rangle \,. \tag{5.13}$$

Given that the superspace current $M_{12\ldots j}$ is associated with the cubic tree-level subdiagrams in a color-ordered $(j{+}1)$-point amplitude, the sum in (5.13) gathers cubic n-point diagrams as shown in figure 4.

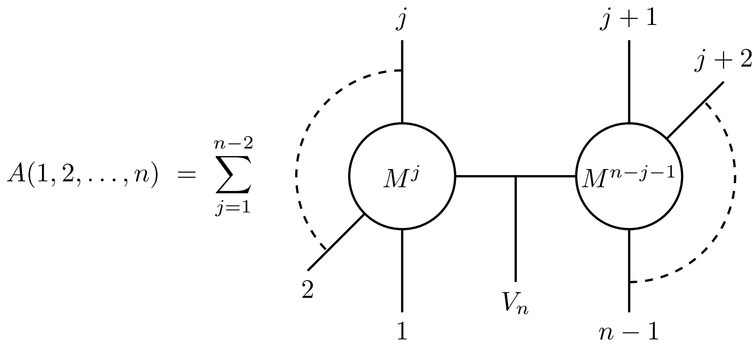


Figure 4: Berends–Giele decomposition of the color ordered SYM amplitude according to (5.13).

More generally, the color-ordered tree amplitudes are given by

$$A(P,n) = \langle E_P\, V_n \rangle \tag{5.14}$$

in terms of the BRST-closed superfield E_P in (4.86). This formula was later rigorously shown to match the $\alpha' \to 0$ limit of the superstring amplitude in [21, 192] and it was also shown in [165] to reduce to the standard Berends–Giele gluonic formula [27] reviewed in (5.1). In spite of these validations, let us now give a separate proof that the expression (5.14) satisfies all the requirements of a color-ordered SYM tree amplitude.

Proposition 12. *In the momentum phase of n massless states where $s_{12\ldots n-1} = 0$ the superfield*

$$E_{12\ldots n-1} V_n \tag{5.15}$$

is in the cohomology of the BRST charge.

Proof. Since $QV_n = 0$, to show that $E_{12\ldots n-1}V_n$ is BRST closed it is enough to show that $QE_P = 0$,

$$\begin{aligned} QE_P = \sum_{XY=P} Q(M_X M_Y) &= \sum_{XY=P} \sum_{RS=X} M_R M_S M_Y - \sum_{XY=P} \sum_{RS=Y} M_X M_R M_S \\ &= \sum_{RSY=P} \big(M_R M_S M_Y - M_R M_S M_Y \big) = 0 \,, \end{aligned} \tag{5.16}$$

where in the last line we consolidated the sums and renamed the dummy words in the second term.

To show that (5.15) is not BRST exact we note that the relation (4.86) depends crucially on the momentum phase space,

$$E_P = QM_P \,, \text{ if } s_P \neq 0 \tag{5.17}$$

$$E_P \neq QM_P \,, \text{ if } s_P = 0 \,. \tag{5.18}$$

This is because $M_P = \frac{1}{s_P}\big(\ldots\big)$ contains a propagator $1/s_P$ which makes the left-hand side of (5.18) ill defined in case of $s_P = 0$. Hence, in the momentum phase space of n massless particles where $s_{12\ldots n-1} = 0$, the superfield $E_{12\ldots n-1}$ is not exact and the expression $E_{12\ldots n-1}V_n$ is in the cohomology of the BRST charge. □

Proposition 13. *The number of kinematic poles from cubic graphs in the color-ordered n-point tree amplitude (5.14) with $n \geq 4$ is given by the Catalan number C_{n-2}.*

Proof. The number of kinematic pole configurations in E_P with P of length $p \geq 3$ and M_X of length $x \geq 2$ are the Catalan numbers C_{p-1} and C_{x-1}, respectively[29]. To see this note that all poles on the right-hand side of (4.86) are distinct, so it implies the recursion relation

$$C_{p-1} = \sum_{x+y=p-2} C_x C_y \,, \quad C_0 = C_1 = 1 \,, \quad p \geq 3 \,. \tag{5.19}$$

The recursion (5.19) coincides with the recursive definition of the Catalan numbers with explicit solution $C_n = \frac{1}{n+1}\binom{2n}{n}$. Therefore the number of poles in the n-point amplitude formula $\langle E_P V_n \rangle$ where $|P| = n-1$ is C_{n-2}. This is the same number of cubic diagrams as in the color ordered n-point SYM amplitude, see e.g. [29]. With $n = 4, 5, 6, 7$, for instance, we get $C_{n-2} = 2, 5, 14, 42$. And since the deconcatenation in the expression (4.86) for E_P is ordered, the kinematic poles in $E_{12\ldots n-1}$ are the same as in the color-ordered n-point tree amplitude. □

Proposition 14. *The color-ordered n-point tree amplitude (5.14) is cyclically symmetric,*

$$A(1, 2, \ldots, n-1, n) = A(2, 3, \ldots, n, 1) \,. \tag{5.20}$$

Proof. By conveniently regrouping terms of $QM_{12\ldots j} = \sum_{i=1}^{j-1} M_{12\ldots i} M_{i+1\ldots j}$, we can recover the difference of cyclic images $E_{23\ldots n} V_1$ and $E_{12\ldots n-1} V_n$ from

$$\begin{aligned} Q \sum_{j=2}^{n-2} M_{12\ldots j} M_{j+1\ldots n} &= M_1 \sum_{j=2}^{n-2} M_{2\ldots j} M_{j+1\ldots n} + \sum_{2 \leq i < j}^{n-2} M_{12\ldots i} M_{i+1\ldots j} M_{j+1\ldots n} \\ &\quad - \sum_{2 \leq j < k}^{n-2} M_{12\ldots j} M_{j+1\ldots k} M_{k+1\ldots n} - \sum_{j=2}^{n-2} M_{12\ldots j} M_{j+1\ldots n-1} M_n \\ &= M_1 \big(E_{23\ldots n} - M_{23\ldots n-1} M_n \big) - \big(E_{12\ldots n-1} - M_1 M_{23\ldots n-1} \big) M_n \\ &= V_1 E_{23\ldots n} - E_{12\ldots n-1} V_n \,. \end{aligned} \tag{5.21}$$

The double sums in the first two lines are easily seen to cancel upon renaming the summation variables i, j, k, and the contributions of $M_1 M_{23\ldots n-1} M_n$ in the third line drop out in the last step. Given that all the $M_{12\ldots j} M_{j+1\ldots n}$ in (5.21) are perfectly valid in the momentum phase space of n massless particles (the highest-multiplicity currents contains non-singular $s^{-1}_{12\ldots n-2}$ and $s^{-1}_{23\ldots n-1}$), we conclude that

$$\langle E_{12\ldots n-1} V_n \rangle - \langle E_{23\ldots n} V_1 \rangle = -\langle Q \big(M_{12} M_{34\ldots n} + M_{123} M_{4\ldots n} + \cdots + M_{12\ldots n-2} M_{n-1 n} \big) \rangle = 0 \,, \tag{5.22}$$

using the vanishing of BRST-exact expressions under the pure spinor bracket. We have thus shown equivalence of the amplitude (5.20) to its cyclic image $i \to i-1 \bmod n$

$$\langle E_{12\ldots n-1} V_n \rangle = \langle E_{23\ldots n} V_1 \rangle \tag{5.23}$$

which concludes the proof. □

For example, from the formula (5.14) and the definition (4.86), one can also verify directly in the momentum phase space of the corresponding n-point amplitude that

$$\begin{aligned} A(1,2,3,4,5) - A(2,3,4,5,1) &= \langle M_{12} M_{34} M_5 - M_1 M_2 M_{345} + M_{123} M_4 M_5 - M_1 M_{23} M_{45} \rangle \\ &= -\langle Q(M_{12} M_{345} + M_{123} M_{45}) \rangle = 0 \,. \end{aligned} \tag{5.24}$$

[29] The difference in the initial lengths for p and x is related to the absence of the overall multiplicative pole $1/s_P$ present in M_P but not in E_P, as it is easy to verify.

5.2.1. Manifesting cyclic symmetry via BRST integration by parts

In the above discussion we have proven that the pure spinor cohomology formula (5.14) is cyclically symmetric, but this is not manifest. We will now show how to exploit the cohomological properties of the pure spinor formula to derive alternative expressions with manifest cyclic symmetry. In doing so, the multiplicity of the Berends–Giele currents featured in the formulae is reduced which renders computations including component evaluations more efficient.

In order to manifest cyclic symmetry in the pure spinor cohomology formula (5.14), we exploit the decoupling of BRST-exact terms $\langle Q(\ldots)\rangle = 0$ from the cohomology. Let us start with a simple example to understand the mechanism. Consider the five point amplitude

$$A(1,2,3,4,5) = \langle M_1 M_{234} M_5 + M_{12} M_{34} M_5 + M_{123} M_4 M_5\rangle \tag{5.25}$$

and note that there are BRST-exact factors of the form $M_i M_j = QM_{ij} = E_{ij}$. So it can be rewritten as

$$\begin{aligned} A(1,2,3,4,5) &= \langle E_{51} M_{234} + M_{12} M_{34} M_5 + M_{123} E_{45}\rangle \\ &= \langle M_{51} E_{234} + M_{12} M_{34} M_5 + E_{123} M_{45}\rangle \\ &= \langle M_{51}(M_2 M_{34} + M_{23} M_4) + M_{12} M_{34} M_5 + (M_1 M_{23} + M_{12} M_3) M_{45}\rangle \\ &= \langle M_{12} M_3 M_{45}\rangle + \mathrm{cyc}(12345) \end{aligned} \tag{5.26}$$

with manifest cyclic symmetry in the last line. Note that in the second line we integrated the BRST charge by parts; by (5.17) this amounts to

$$\langle E_P M_Q\rangle = \langle M_P E_Q\rangle\,, \tag{5.27}$$

for instance $\langle M_{123} E_{45}\rangle = \langle E_{123} M_{45}\rangle = \langle (M_1 M_{23} + M_{12} M_3) M_{45}\rangle$. Notice the reduction of the highest-rank Berends–Giele currents (M_{123}, M_{234}) on the left-hand side of (5.26) to rank-two M_{ij} on the right-hand side.

Naively, in the pure spinor cohomology formula (5.13) for n-point SYM tree amplitudes one needs to know all Berends–Giele currents M_P with multiplicities up to $n{-}2$. For example, in the five-point amplitude (5.26) the first line contains Berends–Giele currents of all multiplicities up to $n{-}2 = 3$. However, after BRST integration by parts, the maximum multiplicity in the last line of (5.26) is two, and its cyclic symmetry is manifest.

In fact, using BRST integrations by parts it was shown in [160] that the highest multiplicity of currents can be lowered to at most $\lfloor \frac{n}{2} \rfloor$ while at the same time yielding superspace formulae for n-point trees with manifest cyclic symmetry

$$\begin{aligned} A(1,2,\ldots,4) &= \frac{1}{2}\langle M_{12} E_{34}\rangle + \mathrm{cyc}(1234)\,, \\ A(1,2,\ldots,5) &= \langle M_{12} M_3 M_{45}\rangle + \mathrm{cyc}(12345)\,, \\ A(1,2,\ldots,6) &= \frac{1}{2}\langle M_{123} E_{456}\rangle + \frac{1}{3}\langle M_{12} M_{34} M_{56}\rangle + \mathrm{cyc}(123456)\,, \\ A(1,2,\ldots,7) &= \langle M_{123} M_{45} M_{67}\rangle + \langle M_1 M_{234} M_{567}\rangle + \mathrm{cyc}(12\ldots7)\,, \\ A(1,2,\ldots,8) &= \frac{1}{2}\langle M_{1234} E_{5678}\rangle + \langle M_{123} M_{456} M_{78}\rangle + \mathrm{cyc}(12\ldots8)\,, \\ A(1,2,\ldots,9) &= \frac{1}{3}\langle M_{123} M_{456} M_{789}\rangle + \langle M_{1234}(M_{567} M_{89} + M_{56} M_{789} + M_{5678} M_9)\rangle + \mathrm{cyc}(12\ldots9)\ , \\ A(1,2,\ldots,10) &= \frac{1}{2}\langle M_{12345} E_{6789;10}\rangle + \langle M_{1234}(M_{567} M_{89;10} + M_{5678} M_{9;10})\rangle + \mathrm{cyc}(12\ldots10)\,, \end{aligned} \tag{5.28}$$

where the fractional coefficients are introduced to avoid overcounting. For example, the four- and six-point instances of (5.28) can be rewritten as

$$\begin{aligned} A(1,2,3,4) &= \langle M_{12} M_3 M_4\rangle + \langle M_{23} M_4 M_1\rangle\,, \\ A(1,2,\ldots,6) &= \langle M_{12} M_{34} M_{56}\rangle + \langle M_{23} M_{45} M_{61}\rangle + \langle M_{123}(M_{45} M_6 + M_4 M_{56})\rangle \end{aligned} \tag{5.29}$$

$$+\langle M_{234}(M_{56}M_1+M_5M_{61})\rangle+\langle M_{345}(M_{61}M_2+M_6M_{12})\rangle\,,$$

via BRST integration by parts and/or the fermionic nature of M_P, so the factors of $\frac{1}{2}$ and $\frac{1}{3}$ in the cyclic sums in (5.28) are essential to not overcount these terms. The manifestly cyclic form of the n-point amplitudes (5.28) is free of fractional coefficients when n is not divisible by 2 or 3.

5.2.2. Component expansion of the pure spinor SYM tree amplitude

BRST invariance of the superfields implies gauge-invariant and supersymmetric components, see the discussion in section 3.4.2. The gauge invariance of the SYM tree-level amplitudes allows one to choose the Harnad–Shnider gauge at a non-linear level to perform the θ-expansion of the multiparticle supersymmetric Berends–Giele currents, see Appendix F. After stripping off the plane-wave factor $e^{k_P\cdot X}$ as in (4.93), this leads to an expansion of $\mathcal{A}^P_\alpha$

$$\mathcal{A}^P_\alpha(\theta)=\frac{1}{2}(\theta\gamma_m)_\alpha\mathfrak{e}^m_P+\frac{1}{3}(\theta\gamma^m)_\alpha(\theta\gamma_m\mathcal{X}_P)-\frac{1}{32}(\theta\gamma^p)_\alpha(\theta\gamma_{mnp}\theta)\mathfrak{f}^{mn}_P+\ldots \tag{5.30}$$

that takes the same form as the θ-expansion (2.17) of the linearized superfield A^i_α in the Harnad–Shnider gauge. The Berends–Giele *polarization* currents $\mathfrak{e}^m_P, \mathcal{X}^\alpha_P$ and $\mathfrak{f}^{mn}_P$ in the component formulation of $D=10$ SYM are given by the recursions (4.117) to (4.120). In this way, the simple $\lambda^3\theta^5$ correlators (3.100) and (3.96) that govern the single-particle correlator $\langle M_1M_2M_3\rangle$ in the three-point amplitude [1] (see section 3.4.5)

$$A(1,2,3)=\langle M_1M_2M_3\rangle=\frac{1}{2}\mathfrak{e}^m_1\mathfrak{f}^{mn}_2\mathfrak{e}^n_3+(\mathcal{X}_1\gamma_m\mathcal{X}_2)\mathfrak{e}^m_3+\mathrm{cyc}(123) \tag{5.31}$$

also determine the multiparticle constituents $\langle M_XM_YM_Z\rangle$ of the n-point amplitudes (5.14),

$$\langle M_XM_YM_Z\rangle=\frac{1}{2}\mathfrak{e}^m_X\,\mathfrak{f}^{mn}_Y\mathfrak{e}^n_Z+(\mathcal{X}_X\gamma_m\mathcal{X}_Y)\mathfrak{e}^m_Z+\mathrm{cyc}(XYZ)=:\mathfrak{M}_{X,Y,Z}\,, \tag{5.32}$$

which defines the shorthand $\mathfrak{M}_{X,Y,Z}$. The component expressions for the above cohomology formulae follow easily by reducing any $\langle M_XM_YM_Z\rangle$ to the combinations $\mathfrak{M}_{X,Y,Z}$ of $\mathfrak{e}^m_P$, $\mathfrak{f}^{mn}_P$ and $\mathcal{X}^\alpha_P$ in (5.32). For instance, the earlier representation (5.13) yields components

$$A(1,2,\ldots,n{-}1,n)=\sum_{XY=12\ldots n-1}\mathfrak{M}_{X,Y,n}\,, \tag{5.33}$$

while the manifestly cyclic representation in (5.28) gives rise to

$$\begin{aligned}
A(1,2,\ldots,4)&=\frac{1}{2}\mathfrak{M}_{12,3,4}+\mathrm{cyc}(1234)\,,\\
A(1,2,\ldots,5)&=\mathfrak{M}_{12,3,45}+\mathrm{cyc}(12345)\,,\\
A(1,2,\ldots,6)&=\frac{1}{3}\mathfrak{M}_{12,34,56}+\frac{1}{2}(\mathfrak{M}_{123,45,6}+\mathfrak{M}_{123,4,56})+\mathrm{cyc}(12\ldots6)\,,\\
A(1,2,\ldots,7)&=\mathfrak{M}_{123,45,67}+\mathfrak{M}_{1,234,567}+\mathrm{cyc}(12\ldots7)\,,\\
A(1,2,\ldots,8)&=\frac{1}{2}(\mathfrak{M}_{1234,567,8}+\mathfrak{M}_{1234,56,78}+\mathfrak{M}_{1234,5,678})+\mathfrak{M}_{123,456,78}+\mathrm{cyc}(12\ldots8)\,.
\end{aligned} \tag{5.34}$$

Given the recursive nature of the definitions of the component Berends–Giele currents $\mathfrak{e}^m_P$, $\mathfrak{f}^{mn}_P$ and $\mathcal{X}^\alpha_P$, the full component expansion of the above amplitudes is readily available and reproduce the results for SYM tree amplitudes available on the website [200]. Furthermore, as discussed in [196, 155], the Berends–Giele currents lead to fast numerical evaluation of the amplitudes.

Note that one can obtain matrix elements of the effective operators $\alpha'\mathbb{F}^3$ and $\alpha'^2\mathbb{F}^4$ of the open bosonic string from (5.34) by introducing α'-corrections of the gluonic components of $\mathfrak{M}_{X,Y,Z}$ as detailed in [201].

5.2.3. Equivalence with the gluonic Berends–Giele recursion

Using the component field-strength (4.120), it follows that the gluonic three-point amplitudes of the Berends–Giele and pure spinor formulae match.

Proposition 15. *When restricted to its gluon components the pure spinor cohomology formula for SYM tree amplitudes (5.13) is equivalent to the standard Berends–Giele formula (5.1).*

Proof. The proof that the general n-point amplitudes agree was written in [165]. The outline of the proof is as follows: first one shows that the lowest gluonic components of $\mathfrak{e}^m_P$ in the superfield (5.30) reproduce the standard Berends–Giele current

$$\mathfrak{e}^m_P\big|_{\chi_j=0} = J^m_P \ . \tag{5.35}$$

Then using transversality (4.121) and momentum conservation in the form of $k^m_X + k^m_Y + k^m_Z = 0$ one rewrites (5.32) as

$$\begin{aligned}\langle M_X M_Y M_Z\rangle &= (\mathfrak{e}_{[X,Y]}\cdot\mathfrak{e}_Z) + \mathfrak{e}^m_X(\mathcal{X}_Y\gamma_m\mathcal{X}_Z) - \mathfrak{e}^m_Y(\mathcal{X}_X\gamma_m\mathcal{X}_Z)\\ &\quad+\frac{1}{2}\sum_{RS=Z}\big[(\mathfrak{e}_R\cdot\mathfrak{e}_X)(\mathfrak{e}_S\cdot\mathfrak{e}_Y) - (\mathfrak{e}_R\cdot\mathfrak{e}_Y)(\mathfrak{e}_S\cdot\mathfrak{e}_X)\big] \ ,\end{aligned} \tag{5.36}$$

which simplifies as follows when Z is a single-particle label $p{+}1$:

$$\langle M_X M_Y M_{p+1}\rangle = (\mathfrak{e}_{[X,Y]}\cdot\mathfrak{e}_{p+1}) + \mathfrak{e}^m_X(\mathcal{X}_Y\gamma_m\mathcal{X}_{p+1}) - \mathfrak{e}^m_Y(\mathcal{X}_X\gamma_m\mathcal{X}_{p+1})\,. \tag{5.37}$$

The pure spinor superspace formula for tree-level SYM amplitudes (5.13) is given by the deconcatenation sum of the correlator (5.37) over $XY = 12\ldots p$ and yields

$$\begin{aligned}A(1,2,\ldots p,p{+}1) &= \sum_{XY=12\ldots p}\Big[(\mathfrak{e}_{[X,Y]}\cdot\mathfrak{e}_{p+1}) + \mathfrak{e}^m_X(\mathcal{X}_Y\gamma_m\mathcal{X}_{p+1}) - \mathfrak{e}^m_Y(\mathcal{X}_X\gamma_m\mathcal{X}_{p+1})\Big]\\ &= s_{12\ldots p}(\mathfrak{e}_{12\ldots p}\cdot\mathfrak{e}_{p+1}) + k^m_{12\ldots p}(\mathcal{X}_{12\ldots p}\gamma_m\mathcal{X}_{p+1})\,,\end{aligned} \tag{5.38}$$

where in the second line the recursions (4.117) and (4.121) were used to identify $\mathfrak{e}^m_{12\ldots p}$ and $\mathcal{X}^\alpha_{12\ldots p}$. Setting the fermions to zero and using (5.35) yields the gluonic Berends–Giele formula [27] and finishes the proof that the pure spinor cohomology formula and the Berends–Giele formula (5.1) are equivalent. ☐

Short representations in the standard Berends–Giele formula. Despite missing the powerful BRST cohomology manipulations, a reduction in the multiplicities of the currents was derived in the standard gluonic Berends–Giele method in [202] to obtain "short" and manifestly cyclic representations of bosonic amplitudes up to eight points. For example, the six-point amplitude was found to be

$$\begin{aligned}A^{\rm YM}(1,2\ldots,6) &= \frac{1}{2}s_{123}J^m_{123}J^m_{456} + \frac{1}{3}[J_{12},J_{34}]^m J^m_{56} + \frac{1}{2}\{J_1,J_{23},J_4\}^m J^m_{56}\\ &\quad+\{J_1,J_2,J_{34}\}^m J^m_{56} + {\rm cyc}(123456)\,,\end{aligned} \tag{5.39}$$

see (5.2) for the brackets $[\ldots]^m$ and $\{\ldots\}^m$, and similar expressions were written for the seven- and eight-point amplitudes [202].

5.2.4. Kleiss–Kuijf amplitude relations

In [28] the color-ordered tree amplitudes were observed to obey the KK relations

$$A(P,1,Q,n) = (-1)^{|P|}A(1,\tilde P \sqcup Q,n)\,, \tag{5.40}$$

which singles out legs 1 and n leading to $(n{-}2)!$ linearly independent amplitudes (w.r.t. constant rather than s_{ij}-dependent coefficients). As a simple example with $P = 2,3$ and $Q = 4$, we have

$$A(2,3,1,4,5) = A(1,3,2,4,5) + A(1,3,4,2,5) + A(1,4,3,2,5)\,. \tag{5.41}$$

In [28] a proof of (5.40) was argued based on the shuffle symmetry (5.7) of the Berends–Giele currents derived by Berends and Giele in [175], see section 5.1.1. The proof that the pure spinor cohomology formula (5.14) satisfies the KK relations is analogous, it follows as a corollary to the equivalence in the proofs of [167] and [176]. However, in this section we wish to see this equivalence more explicitly and use it to prove the KK relations not as an indirect corollary but as a direct statement. The explicit equivalence between [167] and [176] is given by the following lemma, first stated in [165] and proven in [203] (see also equation (41) of [204]):

Lemma 10. *Let P, Q be arbitrary words and j a letter, then*

$$\sum_{XY=P} (-1)^{|X|} \tilde{X} \sqcup\!\sqcup (YjQ) = (-1)^{|P|} j(\tilde{P} \sqcup\!\sqcup Q)\,, \tag{5.42}$$

where X and Y are allowed to be empty in the sums and $\tilde{P}$ denotes the reversal of P.

Proof. The proof follows from an induction on the length $|P|$ of the word P [203]. For the base case of length zero the formula is true as it reduces to $jQ = jQ$. Assume that the formula holds for $|P| = p$, then for $|P| = p{+}1$ set $P = Ci$ for a letter i and a word C of length p. Using the elementary property of summations

$$\sum_{XY=Ci} f(X, Y) = f(Ci, \emptyset) + \sum_{XY=C} f(X, Yi) \tag{5.43}$$

we get

$$\begin{aligned}
\sum_{XY=Ci} (-1)^{|X|} \tilde{X} \sqcup\!\sqcup (YjQ) &= (-1)^{|Ci|} (i\tilde{C}) \sqcup\!\sqcup (jQ) + \sum_{XY=C} (-1)^{|X|} \tilde{X} \sqcup\!\sqcup (YijQ) \\
&= (-1)^{|P|} (i\tilde{C}) \sqcup\!\sqcup Q' + (-1)^{|C|} i(\tilde{C} \sqcup\!\sqcup Q') \\
&= (-1)^{|P|} \big((i\tilde{C}) \sqcup\!\sqcup (jQ) - i(\tilde{C} \sqcup\!\sqcup jQ)\big) \\
&= (-1)^{|P|} j(\tilde{P} \sqcup\!\sqcup Q)\,,
\end{aligned} \tag{5.44}$$

where we defined $Q' = jQ$ to use the induction hypothesis in the second line, and we used the recursive definition of the shuffle product $(i\tilde{C}) \sqcup\!\sqcup (jQ) = i\big(\tilde{C} \sqcup\!\sqcup (jQ)\big) + j\big((i\tilde{C}) \sqcup\!\sqcup Q\big)$ in the fourth line. So the induction holds true when $|P| = p{+}1$ and (5.42) follows. □

It is convenient to rewrite the identity (5.42) such that there are no empty words in the sum on its left-hand side, $\sum_{RS=P} (-1)^{|R|} \tilde{R} \sqcup\!\sqcup (SjQ) + (-1)^{|P|} \tilde{P} \sqcup\!\sqcup (jQ) + PjQ = (-1)^{|P|} j(\tilde{P} \sqcup\!\sqcup Q)$ where $R, S \neq \emptyset$. That is,

$$PjQ = (-1)^{|P|} j(\tilde{P} \sqcup\!\sqcup Q) - \sum_{RS=P} (-1)^{|R|} \tilde{R} \sqcup\!\sqcup (SjQ) - (-1)^{|P|} \tilde{P} \sqcup\!\sqcup (jQ)\,, \quad R, S \neq \emptyset\,, \tag{5.45}$$

or

$$PjQ = (-1)^{|P|} j(\tilde{P} \sqcup\!\sqcup Q) + \text{shuffles}\,, \tag{5.46}$$

giving rise to an equivalence relation in the dual of Lie polynomials [157].

Proposition 16. *The pure spinor cohomology formula for SYM tree amplitudes (5.14) satisfies the KK relations (5.40)*

$$A(P, 1, Q, n) = (-1)^{|P|} A(1, \tilde{P} \sqcup\!\sqcup Q, n)\,. \tag{5.47}$$

Proof. The Corollary 2 in (4.140) shows that the superfield E_P in (4.86) also satisfies the shuffle symmetry

$$E_{R \sqcup\!\sqcup S} = 0\,, \quad \forall\, R, S \neq \emptyset \tag{5.48}$$

since the definition of $E_P = \sum_{XY=P} M_X M_Y$ is also over an antisymmetric deconcatenation in view of the fermionic nature of M_X and M_Y. Since the words R and S in (5.48) must be non-empty, the identity (5.45) can be used to yield

$$E_{PjQ} = (-1)^{|P|} E_{j(\tilde{P} \sqcup\!\sqcup Q)}\,. \tag{5.49}$$

Consequently, from the pure spinor cohomology formula (5.14) we obtain $\langle E_{PjQ} V_n \rangle = (-1)^{|P|} \langle E_{j(\tilde{P} \sqcup\!\sqcup Q)} V_n \rangle$ and therefore the KK relation (5.47) is satisfied. □

5.2.5. Bern-Carrasco-Johansson amplitude relations

The SYM amplitudes from the pure spinor cohomology formula (5.14) are almost trivially zero due to the fact that $E_P = QM_P$ for generic values of all the $s_{i\ldots j}$. The only reason why the superspace expression for the n-point amplitude is not BRST exact is because M_P with $P = 12\ldots n{-}1$ contains a propagator $1/s_P$ which is ill-defined in a momentum phase space of $n = |P|{+}1$ massless particles. The BCJ amplitude relations arise when certain linear combinations of $s_{ij}E_P$ become BRST-exact expressions. Let us consider one simple example to understand the mechanism.

Four point BCJ relation. Let us review the argument of [165]. Consider the Berends–Giele current (4.139)

$$M_{123} = \frac{V_{[[1,2],3]}}{s_{12}s_{123}} + \frac{V_{[1,[2,3]]}}{s_{23}s_{123}} \tag{5.50}$$

in the BCJ gauge, where the local superfields $V_{[P,Q]}$ satisfy the generalized Jacobi identities (4.48). Now consider the linear combination $s_{23}M_{123} - s_{13}M_{213}$

$$s_{23}M_{123} - s_{13}M_{213} = s_{23}\Big(\frac{V_{[[1,2],3]}}{s_{12}s_{123}} + \frac{V_{[1,[2,3]]}}{s_{23}s_{123}}\Big) - s_{13}\Big(\frac{V_{[[2,1],3]}}{s_{12}s_{123}} + \frac{V_{[2,[1,3]]}}{s_{13}s_{123}}\Big) = \frac{V_{[[1,2],3]}}{s_{12}}\,, \tag{5.51}$$

where we used the generalized Jacobi identities of $V_{[P,Q]}$ and $s_{12} + s_{13} + s_{23} = s_{123}$. Note the crucial cancellation of the overall pole in $1/s_{123}$, for which the Jacobi identity is an essential requirement. The identity (5.51) is the Berends–Giele counterpart of the planar binary tree relation (4.149) in terms of Lie polynomials.

Therefore, while M_{123} and M_{213} are ill-defined objects for a four-point amplitude where $s_{123} = 0$, the linear combination $s_{23}M_{123} - s_{13}M_{213}$ is not! This means that it is a valid object to use as a "BRST ancestor" to derive Q-exact expressions with vanishing components

$$E_{\{12,3\}} = s_{23}E_{123} - s_{13}E_{213} = Q\big(s_{23}M_{123} - s_{13}M_{213}\big) = Q\bigg(\frac{V_{[[1,2],3]}}{s_{12}}\bigg)\,. \tag{5.52}$$

Multiplying by the BRST-closed V_4 on the right and using the pure spinor SYM formula (5.14) we get

$$s_{23}A(1,2,3,4) - s_{13}A(2,1,3,4) = \bigg\langle Q\bigg(\frac{V_{[[1,2],3]}}{s_{12}}V_4\bigg)\bigg\rangle = 0\,. \tag{5.53}$$

To summarize, the four-point BCJ amplitude relation [29]

$$s_{23}A(1,2,3,4) - s_{13}A(2,1,3,4) = 0 \tag{5.54}$$

holds because the superspace expression underlying the left-hand side is BRST exact. This is a consequence of the cancellation of the propagator $1/s_{123}$ in the linear combination (5.51) which is only true if $V_{[P,Q]}$ satisfies the generalized Jacobi identities. In other words, in the BCJ gauge of multiparticle superfields the four-point BCJ amplitude relation is obtained due to the vanishing of BRST-exact expressions. Note, however, that the BCJ amplitude relations are valid independently of the precise details of the numerators by non-linear gauge invariance of the cohomology formula (5.13) for SYM amplitudes.[30] Hence, there is no loss of generality in employing BCJ-gauge numerators in the discussions above to identify BRST-exact combinations of the superfields E_P in the cohomology approach.

[30] The perturbiner components of the non-linear gauge variation (2.6) with Berends–Giele currents Ω_P of the gauge scalar lead to the variation $\delta_\Omega M_P = Q\Omega_P + \sum_{P=XY}(\Omega_X M_Y - \Omega_Y M_Y)$. The resulting non-linear gauge variation of the SYM amplitudes $\sum_{12\ldots n-1=XY}\langle M_X M_Y M_n\rangle$ then conspires to a BRST-exact expression after assembling the contributions from $\delta_\Omega M_X$, $\delta_\Omega M_Y$ and $\delta_\Omega M_n$.

BCJ amplitude relations in general. The strategy to derive n-point BCJ amplitude relations from the pure spinor cohomology method hinges on linear combinations of Berends–Giele currents of multiplicity $n-1$ such that the leading propagator $1/s_{12\ldots n-1}$ is absent, as illustrated by the example (5.51) at $n=4$. These combinations can be found in BCJ gauge, and their combinatorial structure is most conveniently encoded in the S bracket of section 4.3.3 which was used in [157] to rigorously prove the n-point statements of [165]. From the pure spinor cohomology discussion above, the property (4.146) demonstrates that certain linear combinations of SYM trees vanish. More precisely,

Proposition 17. ***(BCJ relations)*** *The pure spinor cohomology formula for SYM tree amplitudes (5.14) satisfies BCJ relations*

$$A(\{P,Q\},n)=0\,, \tag{5.55}$$

for any possible distribution of the labels $\{1,2,\ldots,n-1\}$ between P and Q. Moreover, the fundamental BCJ relations in the terminology of [205] are obtained in the special case when $P=1$, $Q=23\ldots n-1$ as

$$-A(\{1,23\ldots n-1\},n)=\sum_{XY=23\ldots n-1} k_1\cdot k_X A(X,1,Y,n)=0\,. \tag{5.56}$$

Proof. Note from the prescription (4.138) and the corollary (4.151) that there is no propagator $1/s_{PQ}$ in

$$M_{\{P,Q\}}=V_{b(\{P,Q\})}\,. \tag{5.57}$$

This implies, by a similar reasoning as in (4.87), that the superfield $E_{\{P,Q\}}$ for $|P|+|Q|=n-1$ is BRST exact in the momentum phase space of n massless particles

$$E_{\{P,Q\}}=QM_{\{P,Q\}}\,, \tag{5.58}$$

even though $s_{PQ}=0$. Since $QV_n=0$, this means that the expression $E_{\{P,Q\}}V_n$ is BRST exact; $Q(M_{\{P,Q\}}V_n)$ and therefore $\langle E_{\{P,Q\}}V_n\rangle=0$ vanishes in the cohomology of the pure spinor bracket. The pure spinor cohomology formula (5.14) implies that $A(\{P,Q\},n)=0$, proving the first claim in (5.55).

To prove that the fundamental BCJ relation of [205] is recovered as in (5.56) we use the lemma (4.152)

$$\{i,Q\}\sim\sum_{XY=Q} k_i\cdot k_Y XiY\,. \tag{5.59}$$

Therefore the BCJ relation (5.55) implies

$$\begin{aligned}0=-A(\{1,Q\},n)&=-\sum_{XY=Q} k_1\cdot k_Y A(X,1,Y,n)\\&=\sum_{XY=Q} k_1\cdot k_X A(X,1,Y,n)+\sum_{XY=Q} k_1\cdot k_n A(X,1,Y,n)\\&=\sum_{XY=Q} k_1\cdot k_X A(X,1,Y,n)\,,\end{aligned} \tag{5.60}$$

where we used that momentum conservation $k_Y=-(k_X+k_1+k_n)$ implies that $k_1\cdot k_Y=-k_1\cdot k_X-k_1\cdot k_n$ to obtain the second line and $\sum_{XY=Q} X1Y=1\sqcup Q$ together with $A((R\sqcup S),n)=0$ to obtain the third line. The last equality follows from $E_{R\sqcup S}=0$ for non-empty R,S as in (4.140). Choosing $Q=23\ldots n-1$ finishes the derivation of the fundamental BCJ relation (5.56) and the proposition is proven. ☐

For examples of the BCJ relations (5.55) generated by the S bracket, we consider

$$\begin{aligned}A(\{12,3\},4)&=s_{23}A(1,2,3,4)-s_{13}A(2,1,3,4)=0\,,\\A(\{1,23\},4)&=s_{12}A(1,2,3,4)-s_{13}A(1,3,2,4)=0\,,\\A(\{123,4\},5)&=s_{34}A(1,2,3,4,5)-s_{24}A(1,3,2,4,5)-s_{24}A(3,1,2,4,5)+s_{14}A(3,2,1,4,5)=0\,,\end{aligned} \tag{5.61}$$

$$A(\{12,34\},5) = s_{23}A(1,2,3,4,5) - s_{24}A(1,2,4,3,5) - s_{13}A(2,1,3,4,5) + s_{14}A(2,1,4,3,5) = 0\,.$$

More generally, the distribution in $-A(\{1,23\ldots n{-}1\},n)=0$ is equivalent, via (5.60), to the fundamental BCJ relations

$$0 = k_1\cdot k_2 A(2,1,3,\ldots,n) + k_1\cdot k_{23}A(2,3,1,4,\ldots,n) + \cdots + k_1\cdot k_{23\ldots n-1}A(2,3,\ldots,n{-}1,1,n)\,, \tag{5.62}$$

whose permutations are known to leave $(n{-}3)!$ independent partial amplitudes [29, 134, 135, 205, 206]. As will be briefly reviewed in section 7.1.2, the BCJ relations were derived in [29] from the color-kinematics duality, see [30, 31] for reviews. The emergence of (local) BCJ-satisfying numerators from the pure spinor superstring will be discussed in sections 7.1.3 and 7.1.6.

5.3. The generating series of tree-level amplitudes

The SYM tree-level amplitudes from the pure spinor superspace expression (5.14) can be compactly described by a generating function [165]. To see this one uses the perturbiner series (4.93) of the unintegrated vertex operator expanded in terms of the Berends–Giele currents (4.82)

$$\mathbb{V} := \lambda^\alpha \mathbb{A}_\alpha = \sum_{i_1} M_{i_1} t^{i_1} e^{k_{i_1}\cdot X} + \sum_{i_1,i_2} M_{i_1 i_2} t^{i_1} t^{i_2} e^{k_{i_1 i_2}\cdot X} + \sum_{i_1,i_2,i_3} M_{i_1 i_2 i_3} t^{i_1} t^{i_2} t^{i_3} e^{k_{i_1 i_2 i_3}\cdot X} + \cdots\,. \tag{5.63}$$

One can show that the generating function of color-dressed SYM amplitudes is given by the natural generalization of the three-point amplitude $\langle V_1 V_2 V_3\rangle$ as

$$\frac{1}{3}\mathrm{Tr}\langle \mathbb{V}\mathbb{V}\mathbb{V}\rangle = \sum_{n=3}^{\infty}\frac{n-2}{n}\sum_{i_1,i_2,\ldots,i_n}\mathrm{Tr}(t^{i_1}t^{i_2}\ldots t^{i_n})A(i_1,i_2,\ldots,i_n)\,. \tag{5.64}$$

It is reassuring to note that the generating function of tree-level amplitudes (5.64) reproduces the interaction term of the ten-dimensional SYM Lagrangian of [207] evaluated on the generating series of (non-local) Berends–Giele currents in superspace: $\mathbb{F}^{mn}(X,0)$ and $\mathbb{W}^\alpha(X,0)$. To see this note from (4.116) that $\mathfrak{e}^m_P, \mathcal{X}^\alpha_P$ and $\mathfrak{f}^{mn}_P$ are the $\theta=0$ components of the generating series $\mathbb{A}^m$, $\mathbb{W}^\alpha$ and $\mathbb{F}^{mn}$. Therefore (5.32) implies that

$$\begin{aligned}\frac{1}{3}\mathrm{Tr}\langle \mathbb{V}\mathbb{V}\mathbb{V}\rangle &= \frac{1}{4}\mathrm{Tr}([\mathbb{A}_m,\mathbb{A}_n]\mathbb{F}^{mn}) + \mathrm{Tr}(\mathbb{W}\gamma^m\mathbb{A}_m\mathbb{W})\Big|_{\theta=0}\\ &= \mathrm{Tr}\Big(\frac{1}{4}\mathbb{F}_{mn}\mathbb{F}^{mn} + (\mathbb{W}\gamma^m\nabla_m\mathbb{W})\Big)\Big|_{\theta=0}\,,\end{aligned} \tag{5.65}$$

where we have used the massless Dirac equation $\nabla_m\gamma^m_{\alpha\beta}\mathbb{W}^\beta = 0$ as well as the field equation $\partial_m\mathbb{F}^{mn} = [\mathbb{A}_m,\mathbb{F}^{mn}] + \gamma^n_{\alpha\beta}\{\mathbb{W}^\alpha,\mathbb{W}^\beta\}$ and discarded a total derivative to rewrite $(\partial_m\mathbb{A}_n)\mathbb{F}^{mn} = -\mathbb{A}_n\big([\mathbb{A}_m,\mathbb{F}^{mn}] + \gamma^n_{\alpha\beta}\{\mathbb{W}^\alpha,\mathbb{W}^\beta\}\big)$. The matching of the Lagrangian with the resummation of all tree-level amplitudes is of course a strong consistency check for the manipulations with the perturbiner series, see e.g. [208].

6. Superstring disk amplitudes with the pure spinor formalism

In the previous section, we have derived the elegant expression (5.13) for n-point SYM tree-level amplitudes in pure spinor superspace solely from locality and cohomology consideration. In this section, we will review how this representation of SYM amplitudes emerges from the CFT prescription (3.75) for superstring disk amplitudes through the field-theory limit $\alpha'\to 0$. This will in fact be a corollary of the key result of this review – a minimal and manifestly BRST invariant form of n-point open-superstring amplitudes: as will become clear from the final and exact-in-α' result (6.49), the entire polarization dependence of the string amplitude is carried by field-theory building blocks. The CFT computation is guided by the local multiparticle superfields in section 4.1 and their generalized Jacobi identities. The manifestly local representation (6.8) of the n-point disk amplitude encountered in intermediate steps will be later on identified as the origin of the color-kinematics duality, see section 7.1.

Both the local and manifestly BRST invariant representations of the string amplitude resonate with field-theory structures when expressed in a Parke–Taylor basis of disk integrals with characteristic cyclic denominators $z_{i_1 i_2} z_{i_2 i_3} \ldots z_{i_{n-1} i_n} z_{i_n i_1}$. On the one hand, the full α'-dependent open-superstring amplitude lines up with the field-theory KLT formula for supergravity tree amplitudes, see (6.69). On the other hand, the field-theory limit $\alpha' \to 0$ of the Parke–Taylor integrals is reviewed to reproduce tree amplitudes of bi-adjoint scalars which play a central role for the color-kinematics duality of gauge theories and different formulations of the gravitational double copy. In fact, as will be argued in the present section and the following, the α'-corrections to Parke–Taylor integrals admit an effective-field-theory interpretation in terms of bicolored scalars with higher-derivative interactions.

6.1. CFT analysis

Tree-level scattering amplitudes of open-string states are determined by iterated integrals on the boundary of a disk worldsheet, as can be seen in the pure spinor prescription (3.75). Using the prescription to compute n-point disk amplitudes in the pure spinor formalism requires the evaluation of a CFT correlator

$$\langle\!\langle V_1(z_1) \prod_{j=2}^{n-2} U_j(z_j) V_{n-1}(z_{n-1}) V_n(\infty) \rangle\!\rangle =: \langle \mathcal{K}_n \rangle \prod_{i<j}^{n} |z_{ij}|^{-2\alpha' s_{ij}} \,, \tag{6.1}$$

with the massless vertex operators (3.59) and (3.62).

The definition of the tree-level *correlators*[31] $\mathcal{K}_n$ on the right-hand side is such that it strips off the Koba–Nielsen factor $\prod_{i<j}^{n} |z_{ij}|^{-2\alpha' s_{ij}}$ from the path integral. We have re-instated the α'-dependence adapted to the correlation functions on a disk worldsheet which differs from the Koba–Nielsen exponents on the sphere in (4.4) and section 7.2.1.

6.1.1. Double poles versus logarithmic singularities

The computation of the correlators $\mathcal{K}_n$ boils down to using the CFT rules of the pure spinor formalism to perform OPE contractions among the vertex operators in (6.1). Since the conformal $h = 1$ primaries $[\partial\theta^\alpha, \Pi^m, d_\alpha, N^{mn}]$ within the integrated vertex (3.59) do not have any zero modes at tree level, the correlator (6.1) can be computed by summing all their OPE singularities summarized in section 3.3.4. As shown in (4.11), these OPEs generically give rise to both single- and double-poles. However, as alluded to in (4.12) and observed in explicit calculations for five [109] and six points [161], the role of the double-pole integrals is to correct the numerators of the single-pole integrals such that any OPE residue $L_{jiki\ldots}$ as defined in (4.7) is transformed to the associated multiparticle vertex operator in the BCJ gauge $V_{ijk\ldots}$. This is the consequence of a subtle interplay between integration-by-parts identities among the disk integrals and the explicit form of the local superfields multiplying these integrals. In particular, the double-pole residues feature factors of $(1+2\alpha' s_{ij})$, as for instance seen in the last line of (4.11), which cancel the tachyon poles $(1+2\alpha' s_{ij})^{-1}$ that would arise from disk integrals over $|z_{ij}|^{-2\alpha' s_{ij}-2}$.

More generally, numerators $(1 + 2\alpha' s_{ijk\ldots})$ from nested OPEs cancel tachyonic poles of multiparticle channels in a highly nontrivial way, see for instance appendix B.3 of [161]. In this way, all the singularities of the correlator $\langle \mathcal{K}_n \rangle$ in (6.1) become *logarithmic*: when visualizing all factors of z_{ij}^{-1} by an edge between vertices i and j, logarithmic singularities are characterized by obtaining a tree graph in the frame $z_n \to \infty$. Integration by parts removes loop subgraphs associated for instance with z_{ij}^{-2} or $(z_{ij} z_{jk} z_{ki})^{-1}$, and the accompanying numerators of $(1+2\alpha' s_{ij})$ or $(1+2\alpha' s_{ijk})$ due to the superfields ensure that no tachyon poles are generated in the coefficients of the logarithmic singularities. It follows from Aomoto's work [209] that non-logarithmic singularities can always be removed via integration by parts, but it is a peculiarity of the superstring (as compared to bosonic or heterotic strings [210, 211]) that the coefficients of the logarithmic integrands become free of tachyon poles and in fact homogeneous in α'.

[31] The context and the subscript clearly differentiates the n-point correlator $\mathcal{K}_n$ in (6.1) and the Berends–Giele current $\mathcal{K}_n$ in (4.80) for the single-particle n of a generic superfield in K_n of (4.1).

6.1.2. Lie-polynomial structure of the correlator

After discarding total worldsheet derivatives and BRST-exact terms (signaled by the $\cong$ symbol below), the calculation of the correlator (6.1) can be summarized by an elegant pattern relating the symmetries of the kinematic factors and logarithmic integrands. By the superfield contributions from the double poles and more general non-logarithmic singularities, all kinematic factors can be written in terms of the multiparticle vertex V_P (4.78) subject to the generalized Jacobi identities (4.48). The logarithmic singularities in turn are carried by the following worldsheet functions $\mathcal{Z}_P$ satisfying shuffle symmetries

$$\mathcal{Z}_{123\ldots p} := \frac{1}{z_{12}z_{23}\ldots z_{p-1,p}}\,, \qquad \mathcal{Z}_{A\sqcup B} = 0\,, \quad \forall\, A, B \neq \emptyset\,. \tag{6.2}$$

At rank $p = 2, 3$, shuffle symmetry is a consequence of antisymmetry $\mathcal{Z}_{12} = z_{12}^{-1} = -z_{21}^{-1} = -\mathcal{Z}_{21}$ and the partial fraction $\mathcal{Z}_{123} + \mathcal{Z}_{213} + \mathcal{Z}_{231} = (z_{12}z_{23})^{-1} + \mathrm{cyc}(1, 2, 3) = 0$, and a general proof can be found in Lemma 5.4 of [180].

At low multiplicities, the sum of OPEs involving one unintegrated and any number of integrated vertices is given by

$$\begin{aligned} V_1(z_1)U_2(z_2) &\cong V_{12}\mathcal{Z}_{12}\,, \\ V_1(z_1)U_2(z_2)U_3(z_3) &\cong V_{123}\mathcal{Z}_{123} + V_{132}\mathcal{Z}_{132}\,, \\ V_1(z_1)U_2(z_2)U_3(z_3)U_4(z_4) &\cong V_{1234}\mathcal{Z}_{1234} + \mathrm{perm}(2, 3, 4)\,, \end{aligned} \tag{6.3}$$

where already at rank two, we have discarded the BRST-exact term in (4.19) to convert the OPE residue L_{21} defined by (4.3) into the two-particle vertex V_{12}.

Given the generalized Jacobi symmetry of V_P and the shuffle symmetry of $\mathcal{Z}_P$, the sums of terms on the right-hand side of (6.3) furnish Lie polynomials [167]. They are in fact permutation symmetric in the labels of V_1 and all the U_{a_i}, say $V_{123}\mathcal{Z}_{123} + V_{132}\mathcal{Z}_{132} = V_{213}\mathcal{Z}_{213} + V_{231}\mathcal{Z}_{231}$, and generalize to

$$V_1(z_1)\prod_{i=1}^{n} U_{a_i}(z_{a_i}) \cong \sum_{|A|=n} V_{1A}\mathcal{Z}_{1A}\,, \tag{6.4}$$

where the summation range $|A| = n$ refers to the $n!$ words A formed by permutations of $a_1a_2\ldots a_{|A|}$. The Lie-polynomial structure implies that the right-hand side of (6.4) is permutation symmetric in $1, a_1, a_2, \ldots, a_{|A|}$ even though only the weaker symmetry in $a_1, a_2, \ldots, a_{|A|}$ is manifest.[32]

Following this reasoning the correlator $\mathcal{K}_n$ can be assembled from two factors of (6.4) corresponding to sequences of OPEs terminating on one of the unintegrated vertex operators $V_1(0)$ or $V_{n-1}(1)$. OPE contributions involving the third unintegrated vertex operator $V_n(z_n)$ are suppressed in our choice of $\mathrm{SL}_2(\mathbb{R})$ frame with $z_n \to \infty$. These selection rules for OPEs lead to $n-2$ deconcatenations $AB = 23\ldots n-2$ (including the ones with $A = \emptyset$ or $B = \emptyset$) and an overall permutation over $(n-3)!$ labels for a total of $(n-2)!$ terms [21]:

$$\mathcal{K}_n = \sum_{AB=23\ldots n-2} \big(V_{1A}\mathcal{Z}_{1A}\big)\big(V_{n-1,\tilde{B}}\mathcal{Z}_{n-1,\tilde{B}}\big)V_n + \mathrm{perm}(2, 3, \ldots, n-2)\,, \tag{6.5}$$

where $\tilde{B}$ denotes the reversal of the word B. The first few expansions of (6.5) read ($\mathcal{Z}_i := 1$),

$$\begin{aligned} \mathcal{K}_3 &= V_1V_2V_3\,, \\ \mathcal{K}_4 &= V_{12}\mathcal{Z}_{12}V_3V_4 + V_1V_{32}\mathcal{Z}_{32}V_4\,, \\ \mathcal{K}_5 &= \big(V_{123}\mathcal{Z}_{123} + V_{132}\mathcal{Z}_{132}\big)V_4V_5 + V_1\big(V_{423}\mathcal{Z}_{423} + V_{432}\mathcal{Z}_{432}\big)V_5 \\ &\quad + \big(V_{12}\mathcal{Z}_{12}\big)\big(V_{43}\mathcal{Z}_{43}\big)V_5 + \big(V_{13}\mathcal{Z}_{13}\big)\big(V_{42}\mathcal{Z}_{42}\big)V_5\,, \\ \mathcal{K}_6 &= V_{1234}\mathcal{Z}_{1234}V_5V_6 + V_{123}\mathcal{Z}_{123}V_{54}\mathcal{Z}_{54}V_6 + V_{12}\mathcal{Z}_{12}V_{543}\mathcal{Z}_{543}V_6 + V_1V_{5432}\mathcal{Z}_{5432}V_6 + \mathrm{perm}(2, 3, 4)\,, \end{aligned} \tag{6.6}$$

[32] This follows from the identity $\sum_A \frac{1}{|A|}\mathcal{Z}_AV_A = \sum_B \mathcal{Z}_{iB}V_{iB}$.

and we reiterate that, by the Lie-polynomial structure of the correlator, $V_{123}\mathcal{Z}_{123} + V_{132}\mathcal{Z}_{132}$ is symmetric in $1, 2, 3$ even though only two out of 3! permutations are spelled out.

One can verify that (6.5) can be obtained using the following two effective rules for multiparticle OPEs

$$V_A(z_a)U_B(z_b) \to \frac{V_{[A,B]}(z_a)}{z_{ab}} , \qquad U_A(z_a)U_B(z_b) \to \frac{U_{[A,B]}(z_a)}{z_{ab}} , \tag{6.7}$$

where z_a and z_b are the worldsheet positions corresponding to the first letters of the words A and B, and the nested brackets in $V_{[A,B]}$ or $U_{[A,B]}$ are expanded as in (4.56) by virtue of the generalized Jacobi identities satisfied by V_P and U_Q.

6.2. Local form of the disk correlator

Using the above results the n-point superstring disk amplitude computed with the pure spinor formalism becomes a sum over $(n-2)!$ superfield numerators along with different worldsheet functions (6.2) [21],

$$\begin{aligned}
\mathcal{A}(\mathbb{1}_n) &= \mathcal{A}(1,2,\ldots,n) \\
&= (2\alpha')^{n-3} \int\limits_{D(\mathbb{1}_n)} \prod_{j=2}^{n-2} dz_j \prod_{1\le i<j}^{n-1} |z_{ij}|^{-2\alpha' s_{ij}} \sum_{AB=23\ldots n-2} \big\langle \big(V_{1A}\mathcal{Z}_{1A}\big)\big(V_{n-1,\tilde{B}}\mathcal{Z}_{n-1,\tilde{B}}\big)V_n\big\rangle + \mathrm{perm}(23\ldots n{-}2) \\
&= (2\alpha')^{n-3} \int\limits_{D(\mathbb{1}_n)} \prod_{j=2}^{n-2} dz_j \prod_{1\le i<j}^{n-1} |z_{ij}|^{-2\alpha' s_{ij}} \\
&\quad \times \Big\{ \sum_{p=1}^{n-2} \frac{\langle V_{12\ldots p}V_{n-1,n-2,\ldots,p+1}V_n\rangle}{(z_{12}z_{23}\cdots z_{p-1,p})(z_{n-1,n-2}\cdots z_{p+2,p+1})} + \mathrm{perm}(23\ldots n{-}2)\Big\} .
\end{aligned} \tag{6.8}$$

Recall that the integration domain $D(\mathbb{1}_n) = D(1,2,\ldots,n)$ defined in (3.77) in the present $\mathrm{SL}_2(\mathbb{R})$-frame with $(z_1, z_{n-1}, z_n) = (0, 1, \infty)$ amounts to the disk ordering $0 < z_2 < z_3 < \ldots < z_{n-2} < 1$.

In order to avoid cluttering, we adopt the notation

$$\int d\mu_P^n := \int_{D(P)} dz_2\, dz_3 \cdots dz_{n-2} \prod_{1\le i<j}^{n-1} |z_{ij}|^{-2\alpha' s_{ij}} , \tag{6.9}$$

where the superscript of the measure tracks the number $|P|{=}n$ of external labels. This shorthand is suited for the choice of $\mathrm{SL}_2(\mathbb{R})$-frame where the worldsheet positions (z_1, z_{n-1}, z_n) are fixed to $(0, 1, \infty)$ (or to $(1, 0, \infty)$ to accommodate all the $(n{-}1)!$ cyclically inequivalent choices of P). The local form of the superstring amplitude then becomes

$$\mathcal{A}_n(P) = (2\alpha')^{n-3} \int d\mu_P^n \sum_{AB=23\ldots n-2} \big\langle \big(V_{1A}\mathcal{Z}_{1A}\big)\big(V_{n-1,\tilde{B}}\mathcal{Z}_{n-1,\tilde{B}}\big)V_n\big\rangle + \mathrm{perm}(23\ldots n{-}2) , \tag{6.10}$$

where we write $\mathcal{A}_n(P) = \mathcal{A}(P)\big|_{|P|=n}$ whenever the multiplicity is not obvious from the shorthand P for the color-ordering.

6.2.1. Four-point example

While the three-point amplitude (3.97) is completely determined by zero modes, the simplest instance of OPE contributions occurs at four points. According to (6.1), the four-point correlator is defined by

$$\langle\!\langle V_1(z_1)U_2(z_2)V_3(z_3)V_4(\infty)\rangle\!\rangle = \langle \mathcal{K}_4\rangle\, |z_{12}|^{-2\alpha' s_{12}} |z_{23}|^{-2\alpha' s_{23}} \tag{6.11}$$

and computed by integrating out the $h = 1$ fields in $U_2(z_2)$:

$$\mathcal{K}_4 \cong \frac{V_{[1,2]}(z_1)}{z_{12}} V_3(z_3)V_4(\infty) + V_1(z_1)\frac{V_{[2,3]}(z_3)}{z_{23}} V_4(\infty)$$

$$\cong \frac{V_{12}V_3V_4}{z_{12}} + \frac{V_1V_{32}V_4}{z_{32}} \,. \tag{6.12}$$

The first line illustrates the origin of the four-point correlator from the OPE effective rules (6.7). The second line (which is equivalent by $V_{ij} = V_{[i,j]}$) reproduces (6.6) and results from permutations of (4.3) and (4.19) while dropping BRST-exact terms and OPEs involving the vertex V_4 at infinity. The $\langle \ldots \rangle$ bracket only refers to the zero modes of $\lambda^\alpha, \theta^\alpha$, see (3.80), that is why the positions of the V_i, V_{ij} are no longer displayed. The integrals in the resulting amplitude

$$\begin{aligned}
\mathcal{A}(\mathbb{1}_4) = \mathcal{A}(1,2,3,4) &= 2\alpha' \int_0^1 dz_2 \left(\frac{\langle V_{12}V_3V_4\rangle}{z_{12}} + \frac{\langle V_1V_{32}V_4\rangle}{z_{32}} \right) |z_{12}|^{-2\alpha' s_{12}} |z_{23}|^{-2\alpha' s_{23}} \\
&= \left(\frac{\langle V_{12}V_3V_4\rangle}{s_{12}} + \frac{\langle V_1V_{23}V_4\rangle}{s_{23}} \right) \frac{\Gamma(1-2\alpha' s_{12})\Gamma(1-2\alpha' s_{23})}{\Gamma(1-2\alpha' s_{12} - 2\alpha' s_{23})} \\
&= A(1,2,3,4) \frac{\Gamma(1-2\alpha' s_{12})\Gamma(1-2\alpha' s_{23})}{\Gamma(1-2\alpha' s_{12} - 2\alpha' s_{23})}
\end{aligned} \tag{6.13}$$

can be straightforwardly identified with the Euler beta function $\int_0^1 dx\, x^{A-1}(1-x)^{B-1} = \frac{\Gamma(A)\Gamma(B)}{\Gamma(A+B)}$ after fixing $(z_1, z_3) = (0, 1)$ which is the backbone of the famous Veneziano amplitude [212]. In passing to the second line, we have used the functional identity $\Gamma(A+1) = A\Gamma(A)$ to make the ratio $-\frac{s_{12}}{s_{23}}$ of the integrals over z_{12}^{-1} and z_{32}^{-1} manifest (which is in fact a special case of the integration-by-parts identities of section 6.3.1). As a result, the four-point SYM amplitude in (5.10) has been factored out in the last line of (6.13), and the remainder of this section is dedicated to the appearance of SYM amplitudes from n-point correlators of the open superstring. Historically, explicit four-point tree-level computations in the pure spinor formalism date back to 2006 and 2008 [89, 158].

6.2.2. Five-point example

To illustrate the multiparticle techniques leading to the result (6.8) above, it is useful to consider the evaluation of the five-point disk correlator via multiparticle vertex operators and the effective OPE calculations (6.7). That is, consider

$$\langle\!\langle V_1(z_1)U_2(z_2)U_3(z_3)V_4(z_4)V_5(z_5)\rangle\!\rangle = \langle \mathcal{K}_5 \rangle\, |z_{12}|^{-2\alpha' s_{12}} |z_{13}|^{-2\alpha' s_{13}} |z_{23}|^{-2\alpha' s_{23}} |z_{24}|^{-2\alpha' s_{24}} |z_{34}|^{-2\alpha' s_{34}} \,, \tag{6.14}$$

where we set $(z_1, z_4, z_5) = (0, 1, \infty)$ at the end (this means that V_5 does not participate in OPEs). First we eliminate z_2 using the OPEs of $U_2(z_2)$ to get

$$\mathcal{K}_5 = \frac{V_{[1,2]}(z_1)}{z_{12}} U_3(z_3)V_4(z_4)V_5(\infty) + V_1(z_1)\frac{U_{[3,2]}(z_3)}{z_{32}} V_4V_5(\infty) + V_1(z_1)U_3(z_3)\frac{V_{[4,2]}(z_4)}{z_{42}} V_5(\infty) \,, \tag{6.15}$$

followed by elimination of z_3 via effective OPEs (6.7) of $U_3(z_3)$

$$\begin{aligned}
\mathcal{K}_5 = &\frac{V_{[[1,2],3]}}{z_{12}z_{13}} V_4V_5 + \frac{V_{[1,2]}}{z_{12}} \frac{V_{[4,3]}}{z_{43}} V_5 \\
&+ \frac{V_{[1,[3,2]]}}{z_{32}z_{13}} V_4V_5 + V_1 \frac{V_{[4,[3,2]]}}{z_{32}z_{43}} V_5 \\
&+ \frac{V_{[1,3]}}{z_{13}} \frac{V_{[4,2]}}{z_{42}} V_5 + V_1 \frac{V_{[[4,2],3]}}{z_{42}z_{43}} V_5 \,.
\end{aligned} \tag{6.16}$$

The contributions from the first, second and third term of (6.15) are organized into separate lines, and we are no longer tracking the locations of the multiparticle vertex operators since only their zero modes remain at this point.

Using the generalized Jacobi identity (4.56) followed by the shuffle symmetry (6.2) we get

$$\mathcal{K}_5 = \frac{V_{123}}{z_{12}z_{23}} V_4V_5 + \frac{V_{132}}{z_{13}z_{32}} V_4V_5 + \frac{V_{12}}{z_{12}} \frac{V_{43}}{z_{43}} V_5 + \frac{V_{13}}{z_{13}} \frac{V_{42}}{z_{42}} V_5 + V_1 \frac{V_{432}}{z_{43}z_{32}} V_5 + V_1 \frac{V_{423}}{z_{42}z_{23}} V_5 \,, \tag{6.17}$$

which reproduces (6.6) and leads to

$$\mathcal{A}(\mathbb{1}_5) = (2\alpha')^2 \int_0^1 dz_3 \int_0^{z_3} dz_2\, |z_{12}|^{-2\alpha' s_{12}} |z_{13}|^{-2\alpha' s_{13}} |z_{23}|^{-2\alpha' s_{23}} |z_{24}|^{-2\alpha' s_{24}} |z_{34}|^{-2\alpha' s_{34}}$$
$$\times \left[\frac{\langle V_{123} V_4 V_5 \rangle}{z_{12} z_{23}} + \frac{\langle V_{12} V_{43} V_5 \rangle}{z_{12} z_{43}} + \frac{\langle V_1 V_{423} V_5 \rangle}{z_{42} z_{23}} + (2 \leftrightarrow 3) \right] \tag{6.18}$$

with $(z_1, z_4) = (0, 1)$.

In contrast to the single integral in the four-point amplitude (6.13), the double integrals in (6.18) cannot be expressed in terms of Gamma functions but instead involve a hypergeometric ${}_3F_2$ function [213] with s_{ij}-dependent parameters at $z = 1$. Five-point tree-level computations in the RNS formalism with external bosons include [198, 214] from the perspective of low-energy effective actions and [215, 216, 217] in the spinor-helicity formalism upon dimensional reduction to $D = 4$. The simplified five-point results in pure spinor superspace [109, 159] address the entire gauge multiplet and furnish key steps towards the representation in (6.18).

6.2.3. Six-point example

The six-point instance of (6.8) is given by

$$\mathcal{A}(\mathbb{1}_6) = (2\alpha')^3 \int_0^1 dz_4 \int_0^{z_4} dz_3 \int_0^{z_3} dz_2 \prod_{1 \le i < j}^{5} |z_{ij}|^{-2\alpha' s_{ij}} \tag{6.19}$$
$$\times \left[\frac{\langle V_{1234} V_5 V_6 \rangle}{z_{12} z_{23} z_{34}} + \frac{\langle V_{123} V_{54} V_6 \rangle}{z_{12} z_{23} z_{54}} + \frac{\langle V_{12} V_{543} V_6 \rangle}{z_{12} z_{54} z_{43}} + \frac{\langle V_1 V_{5432} V_6 \rangle}{z_{54} z_{43} z_{32}} + \mathrm{perm}(2,3,4) \right]$$

with $(z_1, z_5) = (0, 1)$ and builds upon the pure spinor computation in [161]. Earlier six-point tree-level computations in the RNS formalism have been performed for D-dimensional external gluons in [199] and in the spinor-helicity formalism upon dimensional reduction to $D = 4$ [215, 216, 217, 218].

6.3. Non-local form of the disk correlator

The expression (6.8) for the massless n-point open-superstring amplitude is characterized by its total number of $(n-2)!$ terms, written in terms of local superfields V_P in the BCJ gauge and $(n-2)!$ worldsheet integrals. The integrands are given in terms of combinations of $\mathcal{Z}_P \mathcal{Z}_Q$ functions (6.2) with logarithmic singularities and with a distinctive pattern of label distributions among the words P and Q. We will now see how this form can be streamlined and rewritten using only $(n-3)!$ terms.

Rearranging worldsheet functions. The driving force in this rewriting is the judicious use of worldsheet integration by parts in the presence of the Koba–Nielsen factor [21]. To do this, we will first introduce a new set of worldsheet functions X_P, indexed by a word P, whose integration-by-parts relations involve *constant* rather than s_{ij}-dependent coefficients. For reasons to become clear below, it is convenient to define X_{iP} for a fixed label i as

$$X_{iP} = \sum_Q S(P|Q)_i \mathcal{Z}_{iQ} \,, \tag{6.20}$$

where $S(P|Q)_i$ is the KLT matrix (4.160) and $\mathcal{Z}_{iQ}$ is the shuffle-symmetric worldsheet function (6.2). For example, $X_{12} = \frac{s_{12}}{z_{12}}$ and

$$\begin{aligned} X_{123} &= S(23|23)_1 \mathcal{Z}_{123} + S(23|32)_1 \mathcal{Z}_{132} = s_{12}(s_{13} + s_{23}) \frac{1}{z_{12} z_{23}} + s_{12} s_{13} \frac{1}{z_{13} z_{32}} \\ &= s_{12} s_{13} \left(\frac{1}{z_{12} z_{23}} + \frac{1}{z_{13} z_{32}} \right) + \frac{s_{12} s_{23}}{z_{12} z_{23}} = \frac{s_{12} s_{13}}{z_{12} z_{13}} + \frac{s_{12} s_{23}}{z_{12} z_{23}} \\ &= \frac{s_{12}}{z_{12}} \left(\frac{s_{13}}{z_{13}} + \frac{s_{23}}{z_{23}} \right), \end{aligned} \tag{6.21}$$

where we used partial fractions in the second line. In general, one can show that after using partial-fraction identities the X_P functions can be written recursively as

$$X_{Pi} = X_P(X_{p_1 i} + X_{p_2 i} + \cdots + X_{p_k i})\,, \qquad X_i = 1\,,\ k = |P|-1\,, \tag{6.22}$$

where the base case for a letter i is set to one for later convenience. Solving the recursion leads to the simplest instances

$$X_1 = 1\,, \quad X_{12} = \frac{s_{12}}{z_{12}}\,, \quad X_{123} = \frac{s_{12}}{z_{12}}\Big(\frac{s_{13}}{z_{13}} + \frac{s_{23}}{z_{23}}\Big)\,, \quad X_{1234} = \frac{s_{12}}{z_{12}}\Big(\frac{s_{13}}{z_{13}} + \frac{s_{23}}{z_{23}}\Big)\Big(\frac{s_{14}}{z_{14}} + \frac{s_{24}}{z_{24}} + \frac{s_{34}}{z_{34}}\Big) \tag{6.23}$$

and more generally to

$$X_P = \prod_{j=2}^{|P|}\sum_{i=1}^{j-1}\frac{s_{p_i p_j}}{z_{p_i p_j}}\,. \tag{6.24}$$

One can also describe the worldsheet functions (6.20) in terms of the *generalized KLT matrix* (4.165) [156] satisfying generalized Jacobi identities in P and Q such that $S^\ell(iA|iB) = S(A|B)_i$. The definition (6.20) then generalizes to arbitrary words (not necessarily starting with i) as

$$X_P := \frac{1}{|P|}\sum_Q S^\ell(P|Q)\mathcal{Z}_Q \tag{6.25}$$

such that the factor $\frac{1}{|P|}$ compensates the higher number of permutations being summed over objects that satisfy shuffle symmetry and generalized Jacobi identities.

The following property of (6.20) has been first experimentally observed in [163] and later proved in [156, 157] from the properties of $S^\ell(P|Q)$ in (6.25):

Lemma 11. *The worldsheet functions X_P satisfy the generalized Jacobi identities*

$$X_{P\ell(Q)} + X_{Q\ell(P)} = 0\,, \tag{6.26}$$

for instance

$$X_{12} = -X_{21}\,, \quad X_{123} + X_{231} + X_{312} = 0\,, \quad X_{1234} - X_{1243} + X_{3412} + X_{3421} = 0\,. \tag{6.27}$$

6.3.1. Integration by parts

As observed in [21], the chain of $\frac{s_{ij}}{z_{ij}}$ factors that appear in X_P is ideally suited for integration by parts (IBP) when multiplied by the Koba–Nielsen factor of the disk. The key idea is to exploit the vanishing of boundary terms in the total worldsheet derivatives

$$\int_{z_a}^{z_b} dz_k\,\frac{\partial}{\partial z_k}\frac{\prod_{1\le i<j}^{n-1}|z_{ij}|^{-2\alpha' s_{ij}}}{z_{i_1 j_1}\cdots z_{i_{n-4}j_{n-4}}} = 0 \tag{6.28}$$

relevant to arbitrary orderings $D(\ldots,a,k,b,\ldots)$ of n-point disk amplitudes. The absence of boundary terms follows from the contributions $|z_k - z_b|^{-2\alpha' s_{bk}}$ and $|z_k - z_a|^{-2\alpha' s_{ak}}$ to the Koba–Nielsen factor which evidently vanish as $z_k \to z_b$ and $z_k \to z_a$ if $\mathrm{Re}(s_{bk}), \mathrm{Re}(s_{ak}) < 0$. Analytic continuations in the s_{ij} then imply the validity of (6.28) for generic complex kinematics which has already been used in the context of the canceled-propagator argument in (3.87).

Particularly simple instances of (6.28) arise if z_k does not appear in the denominator, i.e. if $k \notin \{i_l, j_l\}$. In these cases, the derivative $\frac{\partial}{\partial z_k}$ only acts on the Koba–Nielsen factor and the resulting IBP identity is homogeneously linear in Mandelstam invariants,

$$\int_{D(P)} dz_2\, dz_3 \cdots dz_{n-2}\frac{\prod_{1\le i<j}^{n-1}|z_{ij}|^{-2\alpha' s_{ij}}}{z_{i_1 j_1}\cdots z_{i_{n-4}j_{n-4}}}\sum_{\substack{m=1\\ m\neq k}}^{n-1}\frac{s_{mk}}{z_{mk}} = 0\,, \tag{6.29}$$

with an arbitrary permutation $P = p_1p_2 \ldots p_n$ of $12 \ldots n$ characterizing the integration domain $D(P)$ in (3.77). The simplest examples at $n = 4$

$$\int_{D(P)} dz_2 \, |z_{12}|^{-2\alpha' s_{12}} |z_{23}|^{-2\alpha' s_{23}} \frac{s_{12}}{z_{12}} = \int_{D(P)} dz_2 \, |z_{12}|^{-2\alpha' s_{12}} |z_{23}|^{-2\alpha' s_{23}} \frac{s_{23}}{z_{23}} \tag{6.30}$$

reproduces the ratio $-\frac{s_{12}}{s_{23}}$ between the four-point integrals in the first line of (6.13) without invoking any Gamma-function identity. At five points, IBP implies

$$\int_{D(P)} dz_2 \, dz_3 \prod_{1 \le i < j}^{4} |z_{ij}|^{-2\alpha' s_{ij}} \frac{s_{12}}{z_{12}} \left(\frac{s_{13}}{z_{13}} + \frac{s_{23}}{z_{23}} \right) = \int_{D(P)} dz_2 \, dz_3 \prod_{1 \le i < j}^{4} |z_{ij}|^{-2\alpha' s_{ij}} \frac{s_{12}}{z_{12}} \frac{s_{34}}{z_{34}} \, . \tag{6.31}$$

Using the notation (6.9), the four- and five-point instances of IBP identities (6.29) relevant to the local correlators (6.13) and (6.18) are obtained by

$$\begin{aligned} \int d\mu_P^4 \, X_{32} &= -\int d\mu_P^4 \, X_{12} \, , \\ \int d\mu_P^5 \, X_{12} X_{43} &= -\int d\mu_P^5 \, X_{123} \, , \\ \int d\mu_P^5 \, X_{432} &= \int d\mu_P^5 \, X_{123} \end{aligned} \tag{6.32}$$

and relabelings $2 \leftrightarrow 3$ of the five-point cases. At six points, the set of master IBPs for the correlator in (6.19) is given by

$$\begin{aligned} \int d\mu_P^6 \, X_{123} X_{54} &= -\int d\mu_P^6 \, X_{1234} \, , \\ \int d\mu_P^6 \, X_{12} X_{543} &= \int d\mu_P^6 \, X_{1234} \, , \\ \int d\mu_P^6 \, X_{5432} &= -\int d\mu_P^6 \, X_{1234} \end{aligned} \tag{6.33}$$

and permutations in $2, 3, 4$, while at seven points we get

$$\begin{aligned} \int d\mu_P^7 \, X_{1234} X_{65} &= -\int d\mu_P^7 \, X_{12345} \, , & \int d\mu_P^7 \, X_{12} X_{6543} &= -\int d\mu_P^7 \, X_{12345} \, , \\ \int d\mu_P^7 \, X_{123} X_{654} &= \int d\mu_P^7 \, X_{12345} \, , & \int d\mu_P^7 \, X_{65432} &= \int d\mu_P^7 \, X_{12345} \, , \end{aligned} \tag{6.34}$$

and permutations in $2, 3, 4, 5$. In general, these IBP identities can be written as

$$\int d\mu_P^n \, X_{1A} X_{(n-1)\tilde{B}} = (-1)^{|B|} \int d\mu_P^n \, X_{1AB} \, , \tag{6.35}$$

where $\tilde{B}$ denotes the reversal of the word B, see the notation in section 1.3. Recalling that we defined $X_i = 1$, the general form (6.35) is valid even when one of A and B is empty. Note again that the labels $i = 1$ and $i = n-1$ are singled out, reflecting the $\mathrm{SL}_2(\mathbb{R})$-frame implicit in the shorthand notation (6.9). Still, the IBP identity (6.35) holds universally for any disk ordering P since any cyclically inequivalent domain $D(P)$ in (3.77) is compatible with the $\mathrm{SL}_2(\mathbb{R})$ frames $(z_1, z_{n-1}, z_n) = (0, 1, \infty)$ or $(1, 0, \infty)$ employed in (6.9).

Moreover, the $|B|$-dependent minus signs cancel out when the identity (6.35) is used together with the reflection property (4.106) of the Berends–Giele supercurrents. This leads to the important identity,

$$\int d\mu_P^n \, (M_{1A} X_{1A})(M_{n-1\tilde{B}} X_{(n-1)\tilde{B}}) = \int d\mu_P^n \, X_{1AB} \, M_{1A} M_{B(n-1)} \, , \tag{6.36}$$

which will be used in the derivation of the non-local and manifestly BRST invariant form of the superstring n-point scattering amplitude on the disk in the next section.

6.3.2. The trading identity

The IBP identity (6.36) will ultimately allow us to derive the non-local $(n-3)!$-term representation of the massless n-point superstring disk amplitude in an elegant manner. Before we do this there is one final important identity to prove, the so-called *trading identity* [21].

Proposition 18. *The local superfields V_P satisfying generalized Jacobi identities and the worldsheet functions $\mathcal{Z}_P$ satisfying shuffle symmetries are related by*

$$\sum_A V_{iA}\mathcal{Z}_{iA} = \sum_A M_{iA}X_{iA} \tag{6.37}$$

to the Berends–Giele supercurrents M_P satisfying shuffle symmetries and the worldsheet functions X_P satisfying generalized Jacobi identities[33].

Proof. Starting from $V_{iA} = \sum_B S(A|B)_i M_{iB}$ (see (4.182) and [166, 157]), we have

$$\sum_A V_{iA}\mathcal{Z}_{iA} = \sum_{A,B} M_{iB}S(A|B)_i\mathcal{Z}_{iA} = \sum_B M_{iB}X_{iB}\,, \tag{6.38}$$

where the second step is based on the symmetry $S(A|B)_i = S(B|A)_i$ of the KLT matrix and the definition of X_{iB} in (6.20). Similarly, we could have used the relation $\mathcal{Z}_{iA} = \sum_B \langle b(iA), iB\rangle X_{iB}$, where the binary-tree map introduced in (4.125) inverts the KLT matrix in the sense of (4.180), to obtain the same conclusion. □

6.3.3. The n-point disk amplitude

The identities derived above allow us to cast the massless n-point disk amplitude into a manifestly gauge invariant form that contains $(n-3)!$ terms [21]. Let us first write down explicit examples at low multiplicities before stating the final result.

Four points. Starting from the rewriting

$$\mathcal{A}_4(P) = 2\alpha' \int d\mu_P^4\, \langle V_{12}\mathcal{Z}_{12}V_3V_4 + V_1V_{32}\mathcal{Z}_{32}V_4\rangle \tag{6.39}$$

of the local form (6.12) of the four-point disk correlator, the trading identity (6.37) regroups the Mandelstam factors to

$$\mathcal{A}_4(P) = 2\alpha' \int d\mu_P^4\, \langle X_{12}M_{12}M_3M_4 + X_{32}M_1M_{32}M_4\rangle\,. \tag{6.40}$$

Then, IBP using (6.32) and the shuffle symmetry of the Berends–Giele current $M_{32} = -M_{23}$ yield

$$\begin{aligned}\mathcal{A}_4(P) &= 2\alpha' \int d\mu_P^4\, X_{12}\langle M_{12}M_3M_4 + M_1M_{23}M_4\rangle\\ &= 2\alpha' \int d\mu_P^4\, X_{12}\langle E_{123}M_4\rangle\\ &= 2\alpha' \int d\mu_P^4\, X_{12}A(1,2,3,4)\,,\end{aligned} \tag{6.41}$$

where we identified the four-point SYM tree amplitude (5.14). After unfolding the notation (6.9) we get the equivalent of (6.13),

$$\mathcal{A}_4(P) = 2\alpha' \int_{D(P)} dz_2\, |z_{12}|^{-2\alpha' s_{12}}|z_{23}|^{-2\alpha' s_{23}}\, \frac{s_{12}}{z_{12}}A(1,2,3,4)\,, \tag{6.42}$$

for the massless four-point superstring amplitude on the disk.

[33] As a side note, the interplay of the generalized Jacobi identity and shuffle symmetry in the trading identity (6.37) gives rise to a Lie-series interpretation of the string disk correlator. This same structural behavior was argued to be present in the genus-one string correlator and exploited to derive the genus-one correlators up to seven points (with partial results at eight points) in [171]. In fact, similar Lie-polynomial structures are expected for correlators at all genera.

Five points. Similarly, starting from the local form (6.18) of the five-point disk amplitude, the trading identity (6.37) together with the IBP identities (6.32) yields

$$\begin{aligned}
\mathcal{A}_5(P) &= (2\alpha')^2 \int d\mu_P^5 \left[\langle V_{123}\mathcal{Z}_{123}V_4V_5\rangle + \langle V_{12}\mathcal{Z}_{12}V_{43}\mathcal{Z}_{43}V_5\rangle + \langle V_1 V_{423}\mathcal{Z}_{423}V_5\rangle + (2\leftrightarrow 3)\right] \qquad (6.43)\\
&= (2\alpha')^2 \int d\mu_P^5 \left[\langle X_{123}M_{123}M_4M_5 + X_{12}X_{43}M_{12}M_{43}M_5 + X_{432}M_1M_{432}M_5\rangle + (2\leftrightarrow 3)\right]\\
&= (2\alpha')^2 \int d\mu_P^5 \left[X_{123}\langle M_{123}M_4M_5 + M_{12}M_{34}M_5 + M_1M_{234}M_5\rangle + (2\leftrightarrow 3)\right]\\
&= (2\alpha')^2 \int d\mu_P^5 \left[X_{123}\langle E_{1234}M_5\rangle + (2\leftrightarrow 3)\right]\\
&= (2\alpha')^2 \int d\mu_P^5 \left[X_{123}A(1,2,3,4,5) + (2\leftrightarrow 3)\right].
\end{aligned}$$

In passing to the third line, we have used the instances $M_{43} = -M_{34}$ and $M_{432} = M_{234}$ of the reflection identity (4.106). We then identified the BRST-closed superfields E_P using (4.86) and the five-point SYM amplitude from the pure spinor cohomology formula (5.14). Finally, restoring the integrals from the shorthand notation (6.9), we obtain the massless superstring five-point amplitude on the disk:

$$\begin{aligned}
\mathcal{A}_5(P) = (2\alpha')^2 &\int_{D(P)} dz_2\, dz_3 \prod_{1\le i<j}^{4} |z_{ij}|^{-2\alpha' s_{ij}} \qquad (6.44)\\
&\times \left[\frac{s_{12}}{z_{12}}\left(\frac{s_{13}}{z_{13}} + \frac{s_{23}}{z_{23}}\right)A(1,2,3,4,5) + \frac{s_{13}}{z_{13}}\left(\frac{s_{12}}{z_{12}} + \frac{s_{32}}{z_{32}}\right)A(1,3,2,4,5)\right].
\end{aligned}$$

Six points. The non-local form of the massless six-point disk amplitude can be derived in similar fashion. Starting from the local form (6.19),

$$\mathcal{A}_6(P) = (2\alpha')^3 \int d\mu_P^6 \left[\sum_{AB=234} \big\langle \big(V_{1A}\mathcal{Z}_{1A}\big)\big(V_{5\tilde{B}}\mathcal{Z}_{5\tilde{B}}\big)V_6\big\rangle + \mathrm{perm}(2,3,4)\right], \qquad (6.45)$$

we use the trading identity (6.37) to obtain

$$\begin{aligned}
\mathcal{A}_6(P) = (2\alpha')^3 \int d\mu_P^6 \Big[\langle\big(&M_{1234}X_{1234}M_5 + M_{123}X_{123}M_{54}X_{54} \qquad (6.46)\\
&+ M_{12}X_{12}M_{543}X_{543} + M_1M_{5432}X_{5432}\big)M_6\rangle + \mathrm{perm}(2,3,4)\Big].
\end{aligned}$$

The IBP identities (6.33) of the worldsheet functions multiplied by the Berends–Giele currents yield

$$\begin{aligned}
\mathcal{A}_6(P) &= (2\alpha')^3 \int d\mu_P^6 \left[X_{1234}\langle M_{1234}M_5M_5 + M_{123}M_{45}M_6 + M_{12}M_{345}M_6 + M_1M_{2345}M_6\rangle + \mathrm{perm}(2,3,4)\right]\\
&= (2\alpha')^3 \int d\mu_P^6 \left[X_{1234}A(1,2,3,4,5,6) + \mathrm{perm}(2,3,4)\right], \qquad (6.47)
\end{aligned}$$

where we easily recognize the expansion of $E_{12345}M_6$ from (4.86) in the first line, and consequently of the tree-level six-point SYM amplitude (5.14) in the last line. So finally [21],

$$\begin{aligned}
\mathcal{A}_6(P) = (2\alpha')^3 &\int_{D(P)} dz_2\, dz_3\, dz_4 \prod_{1\le i<j}^{5} |z_{ij}|^{-2\alpha' s_{ij}} \qquad (6.48)\\
&\times \left[\frac{s_{12}}{z_{12}}\left(\frac{s_{13}}{z_{13}} + \frac{s_{23}}{z_{23}}\right)\left(\frac{s_{14}}{z_{14}} + \frac{s_{24}}{z_{24}} + \frac{s_{34}}{z_{34}}\right)A(1,2,3,4,5,6) + \mathrm{perm}(2,3,4)\right].
\end{aligned}$$

Higher points. Since all the key formulae above generalize to any multiplicity – the local version of the open string disk correlator (6.8), the trading identity (6.37), the IBP relations (6.36) and the pure spinor cohomology formula (5.14) for SYM tree amplitudes – we can state the following result [21]:

Proposition 19. *The massless n-point superstring disk amplitude is given by*

$$\mathcal{A}_n(P) = (2\alpha')^{n-3} \int_{D(P)} \prod_{j=2}^{n-2} dz_j \prod_{1\leq i<j}^{n-1} |z_{ij}|^{-2\alpha' s_{ij}} \left[\prod_{k=2}^{n-2} \sum_{m=1}^{k-1} \frac{s_{mk}}{z_{mk}} A(1,2,\ldots,n) + \text{perm}(2,3,\ldots,n-2) \right]. \tag{6.49}$$

We therefore see that the multiparticle superfield techniques and several related combinatorial identities, all inspired by the simplicity of the pure spinor formalism, lead to a striking simplification of the n-point superstring disk amplitude:

- All polarization dependence is carried by a linear combination of $(n-3)!$ field-theory SYM amplitudes $A(1,Q,n-1,n)$ with $Q = q_2 q_3 \ldots q_{n-2}$ a permutation of $2,3,\ldots,n-2$. These $(n-3)!$ permutations in fact form a basis under the BCJ relations (5.55) or (5.62).
- All the α'-dependence of the n-point amplitude (6.49) resides in the disk integrals over permutations of $X_{12\ldots n-2} = \prod_{k=2}^{n-2} \sum_{m=1}^{k-1} \frac{s_{mk}}{z_{mk}}$ multiplying the SYM amplitudes. Hence, all the string corrections to SYM field-theory are carried by scalar, i.e. polarization independent, integrals.

Additional structures become visible when restricting the integration domains of (6.49) to the $(n-3)!$-family of $D(1,P,n-1,n)$ with $P = p_2 p_3 \ldots p_{n-2}$ a permutation of $2,3,\ldots,n-2$, see the definition in (3.77). This $(n-3)!$-vector of color-ordered string amplitudes

$$\mathcal{A}(1,P,n-1,n;\alpha') = \sum_{Q\in S_{n-3}} F_P{}^Q(\alpha') A(1,Q,n-1,n) \tag{6.50}$$

can then be organized through the following square matrix of integrals

$$\begin{aligned} F_P{}^Q(\alpha') := (2\alpha')^{n-3} \int\limits_{0<z_{p_2}<z_{p_3}<\ldots<z_{p_{n-2}}<1} dz_2\, dz_3 \, \ldots \, dz_{n-2} \prod_{1\leq i<j}^{n-1} |z_{ij}|^{-2\alpha' s_{ij}} \, \frac{s_{1q_2}}{z_{1q_2}} \left(\frac{s_{1q_3}}{z_{1q_3}} + \frac{s_{q_2q_3}}{z_{q_2q_3}} \right) \\ \times \left(\frac{s_{1q_4}}{z_{1q_4}} + \frac{s_{q_2q_4}}{z_{q_2q_4}} + \frac{s_{q_3q_4}}{z_{q_3q_4}} \right) \cdots \left(\frac{s_{1q_{n-2}}}{z_{1q_{n-2}}} + \frac{s_{q_2q_{n-2}}}{z_{q_2q_{n-2}}} + \ldots + \frac{s_{q_{n-3}q_{n-2}}}{z_{q_{n-3}q_{n-2}}} \right), \end{aligned} \tag{6.51}$$

indexed by permutations P and Q of the $n-3$ labels $23\ldots n-2$. On the one hand, there is no obstruction to extending (6.50) beyond the $(n-3)!$ disk orderings $D(1,P,n-1,n)$ – more general choices would simply place some of the integration variables of (6.51) into the regions $(-\infty,0)$ and $(1,\infty)$. On the other hand, as will be elaborated in section 7.3, the $(n-3)!$ color-ordered open-string amplitudes (6.50) already form a basis of the complete $(n-1)!$-family of $\mathcal{A}_n(Q)$ in the color-dressed amplitude (3.78).

One can already anticipate from the symmetric footing of color-ordered open-string and SYM amplitudes in (6.50) that the field-theory limit of $F_P{}^Q(\alpha')$ yields a Kronecker delta in the permutations P,Q,

$$F_P{}^Q(\alpha') = \delta_P^Q + \mathcal{O}(\alpha'^2), \tag{6.52}$$

and we will study this relation and its α'-corrections from several perspectives.

6.4. The open superstring as a field-theory double copy

We shall now relate the form of the disk integrand in the n-point open-string amplitude (6.49) to the structure of the KLT formula (4.159) for gravitational tree amplitudes. In the same way as KLT formulae in field theories are hallmarks of double copy, the form of the disk amplitude is argued to identify the interactions of massless open-superstring excitations as a double copy of SYM with a theory of bicolored scalars dubbed Z-theory.

6.4.1. Parke–Taylor factors and Z-integrals

As pointed out above, the calculation of the string disk amplitudes was carried out in the $\mathrm{SL}_2(\mathbb{R})$ frames where (z_1, z_{n-1}, z_n) are fixed to one of $(0, 1, \infty)$ or $(1, 0, \infty)$ to account for the residual Möbius symmetry of the disk. In order to generalize the n-point formula (6.49) to arbitrary $\mathrm{SL}_2(\mathbb{R})$ frames, we need to undo the above fixing of z_1, z_{n-1}, z_n. The task is to identify an $\mathrm{SL}_2(\mathbb{R})$-*covariant uplift* of the worldsheet functions $\mathcal{Z}_{1A}\mathcal{Z}_{n-1,\tilde{B}}$ or X_{1Q} in the amplitude representations (6.8) or (6.49). In other words, it remains to reverse the $\mathrm{SL}_2(\mathbb{R})$-fixing ($D(P)$ is defined in (3.77))

$$\int_{D(P)} \frac{dz_1\ dz_2\ \cdots\ dz_n}{\mathrm{vol}(\mathrm{SL}(2,\mathbb{R}))} = |z_{1,n-1} z_{1,n} z_{n-1,n}| \int_{D(P)} dz_2\, dz_3\, \ldots\, dz_{n-2}\,, \tag{6.53}$$

and to identify a suitable function $f_{A,B}(z_1, \ldots, z_n)$ such that,

$$\lim_{(z_1,z_{n-1},z_n)\to(0,1,\infty)} |z_{1,n-1} z_{1,n} z_{n-1,n}| \cdot f_{A,B}(z_1, \ldots, z_n) = \mathcal{Z}_{1A}\mathcal{Z}_{n-1,B} \Big|_{z_{n-1}=1}^{z_1=0}\,, \tag{6.54}$$

or $z_1 \leftrightarrow z_{n-1}$. The Jacobian $|z_{1,n-1} z_{1,n} z_{n-1,n}|$ on the right-hand side of (6.53) is part of the prescription $1/\mathrm{vol}(\mathrm{SL}(2,\mathbb{R}))$ that avoids an infinite overcount of z_j-configurations that are related by Möbius transformations [10, 12, 16].[34] The desired uplift $f_{A,B}$ is uniquely determined by (6.54) and requiring $\mathrm{SL}_2(\mathbb{R})$-weight two in each variable: in the same way as

$$\frac{1}{z_i - z_j} \to \frac{(cz_i + d)(cz_j + d)}{z_i - z_j} \quad \text{under} \quad z_k \to \frac{az_k + b}{cz_k + d} \quad \text{with} \quad \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R}) \tag{6.55}$$

is said to have $\mathrm{SL}_2(\mathbb{R})$-weight one in z_i, z_j, the uplift $f_{A,B}$ is required to transform as

$$f_{A,B}(z_1, \ldots, z_n) \to f_{A,B}(z_1, \ldots, z_n) \prod_{j=1}^{n} (cz_j + d)^2 \tag{6.56}$$

to yield well-defined integrals with the measure on the left-hand side of (6.53). The simplest (though not the only) quantities with $\mathrm{SL}_2(\mathbb{R})$-weight two in all of $z_1, z_2, \ldots, z_n$ are the so-called *Parke–Taylor factors*

$$\mathrm{PT}(c_1, c_2, \ldots, c_n) = \frac{1}{z_{c_1c_2} z_{c_2c_3} \ldots z_{c_{n-1}c_n} z_{c_nc_1}}\,. \tag{6.57}$$

After adapting the permutation $C = c_1c_2 \ldots c_n$ to the target expression $\mathcal{Z}_{1A}\mathcal{Z}_{n-1,B}$, it is easy to check that the $\mathrm{SL}_2(\mathbb{R})$-covariant solution to (6.54) is given by

$$f_{A,B}(z_1, \ldots, z_n) = (-1)^{|B|-1}\, \mathrm{PT}(1, A, n, \tilde{B}, n{-}1) \tag{6.58}$$

with $\tilde{B}$ the reversal of B. In other words, the functions $\mathcal{Z}_{1A}\mathcal{Z}_{n-1,B}$ in the local representation (6.8) of the n-point amplitude descend from Parke–Taylor integrals

$$\int_{D(P)} \frac{dz_1\ dz_2\ \cdots\ dz_n}{\mathrm{vol}(SL(2,\mathbb{R}))} \mathrm{PT}(1, A, n, \tilde{B}, n{-}1) = (-1)^{|B|-1} \int_{D(P)} dz_2\, dz_3\, \ldots\, dz_{n-2}\, \mathcal{Z}_{1A}\mathcal{Z}_{n-1,B}\,, \tag{6.59}$$

and the simplest examples are given by

$$\mathrm{PT}(1, 2, 4, 3) \to -\frac{1}{z_{12}}\,, \qquad \mathrm{PT}(1, 4, 2, 3) \to \frac{1}{z_{32}} \tag{6.60}$$

[34] One could alternatively fix any other triplet of punctures z_a, z_b, z_c and change the Jacobian and measure on the right-hand side of (6.53) to $|z_{ab} z_{ac} z_{bc}|$ and $\prod_{j\neq a,b,c}^{n} dz_j$.

as well as

$$\begin{aligned}
&\mathrm{PT}(1,2,3,5,4) \to -\frac{1}{z_{12}z_{23}}\,, \qquad \mathrm{PT}(1,2,5,3,4) \to \frac{1}{z_{12}z_{43}}\,, \qquad \mathrm{PT}(1,5,2,3,4) \to -\frac{1}{z_{43}z_{32}}\,,\\
&\mathrm{PT}(1,3,2,5,4) \to -\frac{1}{z_{13}z_{32}}\,, \qquad \mathrm{PT}(1,3,5,2,4) \to \frac{1}{z_{13}z_{42}}\,, \qquad \mathrm{PT}(1,5,3,2,4) \to -\frac{1}{z_{42}z_{23}}\,.
\end{aligned} \tag{6.61}$$

Upon dressing with the Koba–Nielsen factor of $\mathrm{SL}_2(\mathbb{R})$-weight zero in each variable, the gauge-fixed integrals in (6.8) and (6.49) are found to be expressible in terms of Parke–Taylor- or Z-integrals defined by

$$Z(P|q_1,q_2,\ldots,q_n) := (2\alpha')^{n-3}\int_{D(P)} \frac{dz_1\; dz_2\; \cdots\; dz_n}{\mathrm{vol}(\mathrm{SL}(2,\mathbb{R}))}\frac{\prod_{i<j}^n |z_{ij}|^{-2\alpha' s_{ij}}}{z_{q_1q_2}z_{q_2q_3}\ldots z_{q_{n-1}q_n}z_{q_nq_1}}\,. \tag{6.62}$$

In this setting, Z-integrals are labelled by two permutations $P,Q \in S_n$ up to cyclic identifications in P or Q: the permutation $P := p_1p_2\ldots p_n$ in the first entry encodes the integration domain $D(P)$ in (3.77), while the second permutation $Q := q_1q_2\ldots q_n$ refers to a Parke–Taylor factor (6.57) in the integrand. In summary,

Lemma 12. *The $\mathrm{SL}_2(\mathbb{R})$-covariant uplift of the worldsheet integrals appearing in the local form of the superstring amplitude (6.8) is given by*

$$(2\alpha')^{n-3}\int d\mu_P^n\, \mathcal{Z}_{1A}\mathcal{Z}_{n-1\tilde{B}} = -(-1)^{|B|}Z(P|1,A,n,B,n{-}1)\,, \tag{6.63}$$

where the measure $d\mu_P^n$ is defined in (6.9).

Thus, the local form (6.10) of the superstring amplitude can be rewritten as

$$\mathcal{A}_n(P) = -\sum_{AB=23\ldots n-2}\langle V_{1A}V_{n-1,\tilde{B}}V_n\rangle(-1)^{|B|}Z(P|1,A,n,B,n{-}1) + \mathrm{perm}(23\ldots n{-}2)\,, \tag{6.64}$$

for instance

$$\begin{aligned}
\mathcal{A}_4(P) &= -\langle V_{12}V_3V_4\rangle Z(P|1,2,4,3) + \langle V_1V_{32}V_4\rangle Z(P|1,4,2,3)\,,\\
\mathcal{A}_5(P) &= -\langle V_{123}V_4V_5\rangle Z(P|1,2,3,5,4) + \langle V_{12}V_{43}V_5\rangle Z(P|1,2,5,3,4) - \langle V_1V_{432}V_5\rangle Z(P|1,5,2,3,4)\\
&\quad - \langle V_{132}V_4V_5\rangle Z(P|1,3,2,5,4) + \langle V_{13}V_{42}V_5\rangle Z(P|1,3,5,2,4) - \langle V_1V_{423}V_5\rangle Z(P|1,5,3,2,4)\,.
\end{aligned} \tag{6.65}$$

6.4.2. Open superstrings as a KLT formula

The ${F_P}^Q$-functions (6.51) in the n-point disk amplitude (6.50) are integrals of worldsheet functions X_{1Q} as one can readily check from their expressions in (6.24),

$${F_P}^Q(\alpha') = (2\alpha')^{n-3}\int_{D(1,P,n-1,n)} dz_2\, dz_3\, \ldots\, dz_{n-2}\prod_{1\le i<j}^{n-1}|z_{ij}|^{-2\alpha' s_{ij}}X_{1Q}\,. \tag{6.66}$$

In the same way as the X_{1Q} integrands are related to the $\mathcal{Z}_{1R}$ functions through the KLT matrix (6.20), one can represent the ${F_P}^Q(\alpha')$ via

$$\begin{aligned}
{F_P}^Q(\alpha') &= (2\alpha')^{n-3}\sum_{R\in S_{n-3}}S(Q|R)_1\int_{D(1,P,n-1,n)} dz_2\, dz_3\, \ldots\, dz_{n-2}\prod_{1\le i<j}^{n-1}|z_{ij}|^{-2\alpha' s_{ij}}\,\mathcal{Z}_{1R}\\
&= -(2\alpha')^{n-3}\sum_{R\in S_{n-3}}S(Q|R)_1\int_{D(1,P,n-1,n)}\frac{dz_1\, dz_2\, \ldots\, dz_n}{\mathrm{vol}(\mathrm{SL}_2(\mathbb{R}))}\prod_{1\le i<j}^{n}|z_{ij}|^{-2\alpha' s_{ij}}\;\mathrm{PT}(1,R,n,n{-}1)\,,
\end{aligned} \tag{6.67}$$

using (6.59) at $B=\emptyset$ in passing to the $\mathrm{SL}(2,\mathbb{R})$-covariant last line. This identifies the integrals ${F_P}^Q$ to be linear combinations of Z-integrals (6.62) selected by the KLT kernel [166],

$$
{F_P}^Q = -\sum_{R\in S_{n-3}} S(Q|R)_1 Z(1,P,n{-}1,n|1,R,n,n{-}1)\ . \tag{6.68}
$$

Hence, the n-point open-superstring amplitude (6.50) takes the form of the field-theory KLT relations [166]

$$
\mathcal{A}_n(P) = -\sum_{Q,R\in S_{n-3}} Z(P|1,R,n,n{-}1)S(R|Q)_1 A(1,Q,n{-}1,n) \tag{6.69}
$$

with one of the two SYM factors in the supergravity tree amplitude (4.159) replaced by disk integrals $A(1,R,n,n{-}1)\to Z(P|1,R,n,n{-}1)$. This is the case for any choice of the open-string color ordering P in (6.69) which is a spectator in the sum over permutations Q,R entering the KLT kernel. We have dropped the restriction of the disk integrals (6.49) to the $(n{-}3)!$-family of domains $D(1,P,n{-}1,n)$ which was convenient to organize the ${F_P}^Q$ in (6.51) into a square matrix.

Given that the KLT formula is a central tree-level incarnation of the double-copy structure in perturbative gravity, it is tempting to interpret (6.69) as signaling the open superstring to be a double copy. Indeed, we support the interpretation of the Z-integrals (6.62) as amplitudes in a theory of bi-colored scalars in sections 6.4.4, 7.4 and 8.6.

We conclude by illustrating the KLT products (6.68) and (6.69) through their four- and five-point examples:

Four points. Since the permutation sums over S_{n-3} trivialize at $n=4$ and the KLT kernel becomes a scalar $S(2|2)_1=s_{12}$, we find the simple results

$$
\begin{aligned}
{F_2}^2 &= -s_{12}Z(1,2,3,4|1,2,4,3) = \frac{\Gamma(1-2\alpha' s_{12})\Gamma(1-2\alpha' s_{23})}{\Gamma(1-2\alpha' s_{12}-2\alpha' s_{23})}\,,\\
\mathcal{A}_4(P) &= -Z(P|1,2,4,3)s_{12}A(1,2,3,4)\,,
\end{aligned} \tag{6.70}
$$

where we have imported the Gamma-function representation of the four-point disk integral from (6.13) in the first line.

Five points. At five points, the permutation-inequivalent entries of the symmetric KLT kernel are $S(23|23)_1 = s_{12}(s_{13}{+}s_{23})$ and $S(23|32)_1 = s_{12}s_{13}$. The resulting functions ${F_P}^Q$ and KLT representation of the disk amplitude are

$$
\begin{aligned}
{F_{23}}^{23} &= -s_{12}(s_{13}{+}s_{23})Z(1,2,3,4,5|1,2,3,5,4) - s_{12}s_{13}Z(1,2,3,4,5|1,3,2,5,4)\\
{F_{23}}^{32} &= -s_{12}s_{13}Z(1,2,3,4,5|1,2,3,5,4) - s_{13}(s_{12}{+}s_{23})Z(1,2,3,4,5|1,3,2,5,4)
\end{aligned} \tag{6.71}
$$

as well as

$$
\mathcal{A}_5(P) = -\begin{pmatrix} Z(P|1,2,3,5,4)\\ Z(P|1,3,2,5,4)\end{pmatrix}^T \begin{pmatrix} s_{12}(s_{13}{+}s_{23}) & s_{12}s_{13}\\ s_{12}s_{13} & s_{13}(s_{12}{+}s_{23})\end{pmatrix}\begin{pmatrix} A(1,2,3,4,5)\\ A(1,3,2,4,5)\end{pmatrix}\ . \tag{6.72}
$$

6.4.3. KK and BCJ relations of Z-integrals

The KLT formula (4.159) does not manifest the permutation symmetry of the supergravity amplitude by the sum over $(n{-}3)!$ rather than $(n{-}1)!$ color-orderings of the two types of SYM amplitudes.[35] One can verify on the basis of the KK and BCJ relations (5.40) and (5.62) of SYM amplitudes that different choices of the legs $(1,n{-}1,n)\leftrightarrow(a,b,c)$ excluded from the permutation sums yield equivalent KLT formulae.

[35] See [182] for alternative versions of the KLT formula with manifest permutation symmetry and [219] for an early discussion thereof in the mathematics literature.

In the context of disk amplitudes, the KLT formula in (6.69) ultimately applies to the n-point correlator (6.1) in the amplitude prescription (3.75),

$$\langle \mathcal{K}_n \rangle = -\frac{dz_1\, dz_{n-1}\, dz_n}{\mathrm{vol}(\mathrm{SL}(2,\mathbb{R}))} \sum_{Q,R\in S_{n-3}} \mathrm{PT}(1,R,n,n{-}1) S(R|Q)_1 A(1,Q,n{-}1,n) \ \mathrm{mod}\ \nabla_{z_k}\,, \tag{6.73}$$

where the total Koba–Nielsen derivatives $\nabla_{z_k} f$ discarded in the IBP procedure are not tracked,

$$\nabla_{z_k} f = \partial_{z_k} f - 2\alpha' f \sum_{\substack{j=1\\ j\neq k}}^{n} \frac{s_{kj}}{z_{kj}}\,. \tag{6.74}$$

By the discussion in section 3.4.2, the integrated correlator does not depend on the distribution of external legs to integrated and unintegrated vertices, so the KLT representation (6.73) of the correlator is bound to be permutation invariant. In the same way as the permutation invariance of supergravity amplitudes originates from KK and BCJ relations of its SYM constituents, the symmetry of the correlator requires KK and BCJ relations of the Parke–Taylor factors (6.57) modulo total Koba–Nielsen derivatives. The factors of $dz_1 dz_{n-1} dz_n$ on the right-hand side of (6.73) are merely a reminder of the $\mathrm{SL}(2,\mathbb{R})$ frame used to define $\mathcal{K}_n$ in (6.1) and lead to a permutation invariant measure upon insertion into (3.75).

For a fixed choice of the integration domain $D(P)$, the $Z(P|Q)$ integrals (6.62) associated with different permutations of $Q = q_1, q_2, \ldots, q_n$ indeed satisfy the same relations as color-ordered SYM amplitudes. First, cyclic symmetry and reflection parity immediately follow from the definition (6.57) of Parke–Taylor factors,

$$\begin{aligned} Z(P|q_1,q_2,q_3\ldots,q_n) &= Z(P|q_2,q_3,\ldots,q_n,q_1)\,,\\ Z(P|q_1,q_2,\ldots,q_n) &= (-1)^n Z(P|q_n,\ldots,q_2,q_1)\,. \end{aligned} \tag{6.75}$$

Second, partial-fraction rearrangements of the integrand imply,

$$Z(P|1,A,n,B) = (-1)^{|B|} Z(P|1,\tilde{B}\sqcup A,n)\,, \quad \forall\, A,B\,, \tag{6.76}$$

which is the direct $A(\cdot) \to Z(P|\cdot)$ analogue of the KK relation (5.47). Third, IBP relations as in section 6.3.1 take the form of the BCJ relations (5.62) with $A(\cdot) \to Z(P|\cdot)$

$$0 = \sum_{j=2}^{n-1} (k_{q_1}\cdot k_{q_2q_3\ldots q_j}) Z(P|q_2,q_3,\ldots,q_j,q_1,q_{j+1},\ldots,q_n)\,, \tag{6.77}$$

see appendix B of [166] for a proof. By analogy with (5.55), an equivalent system of integration-by-parts relations is furnished by

$$Z(P|\{Q,R\},n) = 0\,, \tag{6.78}$$

see section 4.3.3 for the S bracket. The fact that $Z(P|\cdot)$-integrals at fixed choice of P obey direct analogues (6.75) to (6.78) of field-theory amplitude relations supports the interpretation of disk integrals as amplitudes in a scalar theory.

Note that the expressions for n-point disk amplitudes in four-dimensional MHV helicities in [220] follow from the relabeling of (6.73) involving the Parke–Taylor basis $\mathrm{PT}(1,2,3,R)$ with permutations R of $4,5,\ldots,n$ as well as the Parke–Taylor formula [221] for the dimensional reduction of the SYM amplitudes.

6.4.4. Bi-adjoint scalars from the field-theory limit of Z-integrals

The interpretation of Z-integrals (6.62) in terms of scalar field theory is further substantiated by their low-energy limits $\alpha' \to 0$, where tree-level amplitudes of bi-adjoint scalars are recovered. More specifically, we encounter the theory of bi-adjoint scalars $\Phi := \Phi_{i|a} t^i \otimes \tilde{t}^a$ taking values in the tensor product $U(N)\times U(\tilde{N})$ of color groups with associated structure constants f_{ijk} and $\tilde{f}_{abc}$, i.e. $[t^i,t^j] = f_{ijk}t^k$ and $[\tilde{t}^a,\tilde{t}^b] = \tilde{f}_{abc}\tilde{t}^c$. The Lagrangian defining the bi-adjoint theory features a cubic interaction

$$\mathcal{L}_{\phi^3} = \frac{1}{2}\partial_m \Phi_{i|a} \partial^m \Phi_{i|a} + \frac{1}{3!} f_{ijk}\tilde{f}_{abc} \Phi_{i|a}\Phi_{j|b}\Phi_{k|c}\,, \tag{6.79}$$

and by the two types of adjoint indices of the scalars, its tree amplitudes can be expanded in terms of two species of independent traces involving either t^i or $\tilde{t}^a$,

$$M_n^{\phi^3} = \sum_{P,Q} \mathrm{Tr}(t^{1P})\mathrm{Tr}(\tilde{t}^{1Q}) m(1,P|1,Q)\,, \qquad t^{1P} := t^1 t^{p_1} t^{p_2} \ldots t^{p_{n-1}}\,, \qquad \tilde{t}^{1Q} := \tilde{t}^1 \tilde{t}^{q_1} \tilde{t}^{q_2} \ldots \tilde{t}^{q_{n-1}}\,. \tag{6.80}$$

This doubles the color-decomposition of open-string and gauge-theory tree amplitudes in (3.78), and the color-independent building blocks $m(A|B)$ are referred to as *doubly-partial* amplitudes [35]. From the Feynman rules of the Lagrangian (6.79), the doubly-partial amplitudes solely depend on the $s_{ij\ldots k}$ via propagators of tree-level diagrams with cubic vertices or *cubic diagrams* for short.

Also the field-theory limit of disk integrals (6.62) yields kinematic poles that correspond to the propagators of cubic diagrams (or planar binary trees) [147, 222, 192]. These poles appear only in the planar channels of the associated planar binary trees, corresponding to groups of adjacent external particles in the planar trees. Luckily, these adjacent poles in the field-theory limits of the disk integrals[36] (6.62) admit a nice combinatorial expansion encoded in the doubly-partial amplitudes [35]

$$\lim_{\alpha'\to 0} Z(P|Q) = m(P|Q)\,. \tag{6.81}$$

In other words, the bi-adjoint scalar theory (6.79) gives the field-theory limit of Z-integrals, as expected from the early discussions of [147, 222]. Hence, bi-adjoint scalars furnish the low-energy limit of Z-theory.

Berends–Giele double currents. While the kinematic poles in Z-integrals have been systematically studied at various multiplicities from several perspectives [216, 192, 166, 223], the straightforward Feynman-diagram expansion and rich combinatorial structure of the doubly-partial amplitudes make (6.81) a rewarding shortcut for the computation of field-theory limits. In particular, we shall now introduce *Berends–Giele double currents* that encode the planar-binary-tree expansion of $m(P|Q)$ and offer a highly efficient approach to the kinematic poles of Z-integrals.

The field equation following from the Lagrangian (6.79) can be written as

$$\Box\Phi = [\![\Phi, \Phi]\!]\,, \tag{6.82}$$

where we define $[\![\Phi,\Phi]\!] := (\Phi_{i|a}\Phi_{j|b} - \Phi_{j|a}\Phi_{i|b})t^i t^j \otimes \tilde{t}^a \tilde{t}^b$ for $\Phi := \Phi_{i|a} t^i \otimes \tilde{t}^a$. A solution to the non-linear field equation (6.82) can be constructed perturbatively in terms of *Berends–Giele double currents* $\phi_{P|Q}$ with the ansatz [224],

$$\Phi(X) := \sum_{P,Q} \phi_{P|Q}\, t^P \otimes \tilde{t}^Q\, e^{k_P\cdot X}, \qquad t^P := t^{p_1} t^{p_2} \ldots t^{p_{|P|}} \tag{6.83}$$

which generalizes the perturbiner expansion (4.93) in SYM to two species of Lie-algebra generators. Since the ansatz (6.83) contains the plane-wave factor $e^{k_P\cdot X}$ (as opposed to $e^{k_Q\cdot X}$), the coefficients $\phi_{P|Q}$ must vanish unless P is a permutation of Q in order to have a well-defined multiparticle interpretation, i.e.

$$\phi_{P|Q} := 0\,, \qquad \text{if } P \setminus Q \neq \emptyset\,. \tag{6.84}$$

Plugging the ansatz (6.83) into the field equation (6.82) leads to the following recursion [224]

$$\phi_{P|Q} = \frac{1}{s_P} \sum_{XY=P} \sum_{AB=Q} \big(\phi_{X|A}\phi_{Y|B} - (X \leftrightarrow Y)\big)\,, \qquad \phi_{i|j} = \delta_{ij}\,, \tag{6.85}$$

see (3.107) for the definition of multiparticle Mandelstam invariants s_P. The recursion terminates with the single-particle double current subject to the linearized equation $\Box\phi_{i|j}e^{k_i\cdot X} = 0$ such that $k_i^2 = 0$, and we

[36]The factor of $(2\alpha')^{n-3}$ in the definition (6.62) of $Z(P|Q)$ guarantees that the leading term in the low-energy expansion is of order s_{ij}^{3-n}, without any accompanying factors of α'.

can pick normalization conventions where $\phi_{i|j} = \delta_{ij}$. Given that the two entries P, Q of the currents enter the recursion (6.85) on equal footing, the symmetry $\phi_{i|j} = \phi_{j|i}$ propagates to arbitrary rank,

$$\phi_{P|Q} = \phi_{Q|P} \,. \tag{6.86}$$

Since the summands on the right-hand side of (6.85) are antisymmetric in both X, Y and A, B, the shuffle symmetry

$$\phi_{A\sqcup B|Q} = 0 \;\forall\; A, B \neq \emptyset \,, \qquad \phi_{A|P\sqcup Q} = 0 \;\forall\; P, Q \neq \emptyset \tag{6.87}$$

follows from the same type of combinatorial proof as given for the Berends–Giele currents of SYM below (4.132). In particular, Schocker's identity [176] can be applied to both slots to infer

$$\phi_{AiB|Q} = (-1)^{|A|}\phi_{i(\tilde{A}\sqcup B)|Q} \,, \qquad \phi_{A|PiQ} = (-1)^{|P|}\phi_{A|i(\tilde{P}\sqcup Q)} \tag{6.88}$$

from (6.87), see (4.105) for the analogous identity for SYM currents. Upon setting $B \to \emptyset$ in the first identity or $Q \to \emptyset$ in the second, (6.88) specializes to the reflection identities

$$\phi_{Ai|Q} = (-1)^{|A|}\phi_{i\tilde{A}|Q} \,, \qquad \phi_{A|Pi} = (-1)^{|P|}\phi_{A|i\tilde{P}} \,. \tag{6.89}$$

The symmetries (6.87) generalize the standard Berends–Giele symmetry to both sets of color generators and guarantee that the ansatz (6.83) is a (double) Lie series [167], thereby preserving the Lie-algebra-valued nature of $\Phi(X)$ in (6.82) w.r.t. both t^i and $\tilde{t}^a$.

Examples of Berends–Giele double currents. Based on $\phi_{i|j} = \delta_{ij}$, the simplest application of the recursion (6.85) leads to rank-two double currents:

$$\phi_{12|12} = \frac{1}{s_{12}}(\phi_{1|1}\phi_{2|2} - \phi_{2|1}\phi_{1|2}) = \frac{1}{s_{12}} \,, \quad \phi_{12|21} = \frac{1}{s_{12}}(\phi_{1|2}\phi_{2|1} - \phi_{2|2}\phi_{1|1}) = -\frac{1}{s_{12}} \,. \tag{6.90}$$

At rank three and four, it is straightforward to work out examples such as

$$\begin{aligned}
\phi_{123|123} &= \frac{1}{s_{123}}\Big(\frac{1}{s_{12}} + \frac{1}{s_{23}}\Big) \,, \quad & \phi_{1234|1234} &= \frac{1}{s_{1234}}\Big(\frac{1}{s_{123}s_{12}} + \frac{1}{s_{123}s_{23}} + \frac{1}{s_{12}s_{34}} + \frac{1}{s_{234}s_{23}} + \frac{1}{s_{234}s_{34}}\Big) \,, \\
\phi_{123|132} &= -\frac{1}{s_{23}s_{123}} \,, & \phi_{1234|1243} &= -\frac{1}{s_{1234}}\Big(\frac{1}{s_{12}s_{34}} + \frac{1}{s_{234}s_{34}}\Big) \,.
\end{aligned} \tag{6.91}$$

Berends–Giele double currents from planar binary trees. As pointed out in [156], the Berends–Giele double currents $\phi_{P|Q}$ can be obtained from the planar-binary-tree expansions given by the b map (4.125) as

$$\phi_{P|Q} = \langle b(P), Q\rangle = \langle P, b(Q)\rangle \,, \tag{6.92}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product of words (C.11), and the symmetry $\phi_{P|Q} = \phi_{Q|P}$ of (6.86) is a consequence of the self-adjoint property (4.127) of the b map. In addition, the shuffle symmetry (6.87) follows from the property $b(R \sqcup S) = 0$ proven in (4.132).

For an example application of (6.92), using the expansion (4.126) for $b(123)$ we get

$$\phi_{123|132} = \langle b(123), 132\rangle = \frac{1}{s_{12}s_{123}}\langle[[1,2],3],132\rangle + \frac{1}{s_{23}s_{123}}\langle[1,[2,3]],132\rangle = -\frac{1}{s_{23}s_{123}} \tag{6.93}$$

as $\langle[[1,2],3],132\rangle = \langle 123 - 213 - 312 + 321, 132\rangle = 0$ and $\langle[1,[2,3]],132\rangle = \langle 123 - 132 - 231 + 321, 132\rangle = -1$.

We will see later in (8.91) that there is a generalization of the relation (6.92) between planar binary trees and the Berends–Giele double current to a series expansion in α'.

Berends–Giele formula for doubly-partial amplitudes. Similar to the Berends–Giele formulae (5.1) in gauge theory, the double currents of bi-adjoint scalars yield their doubly-partial amplitudes via [224][37]

$$m(P,n|Q,n) = \lim_{s_P\to 0} s_P \phi_{P|Q} \,. \tag{6.94}$$

We reiterate that $\phi_{i|j} = \delta_{ij}$, and $\phi_{P|Q}$ vanishes unless P is a permutation of Q such that $s_P = s_Q$. By the cyclic symmetry of $m(R|S)$ in both words R and S, there is no loss of generality in assuming their n-point instances to take the form $m(P,n|Q,n)$, where $|P| = |Q| = n-1$.[38]

It is easy to see using the symmetries (6.87) obeyed by the double currents that the $m(P,n|Q,n)$ in (6.94) obey KK relations independently in both sets of color orderings. Moreover, the Cachazo–He–Yuan (CHY) representation of doubly-partial amplitudes lead to BCJ relations in both entries [35],

$$m(\{A,B\},n|Q) = m(A|\{P,Q\},n) = 0 \,. \tag{6.95}$$

As an earlier alternative to (6.94), doubly-partial amplitudes $m(R|S)$ can be determined from the algorithm described in [35] based on drawing polygons and collecting the products of propagators associated to cubic graphs which are compatible with both of R and S as planar orderings. Their overall sign, however, requires keeping track of the polygon orientations.

As the main result of this section, by combining (6.81) with (6.94) we get

Proposition 20. *The field-theory limit of the n-point disk integrals (6.62) is given by*

$$\lim_{\alpha'\to 0} Z(P,n|Q,n) = \lim_{s_P\to 0} s_P \phi_{P|Q} \,. \tag{6.96}$$

Examples of field-theory limits. Typical expressions for doubly-partial amplitudes or field-theory limits of Z-integrals are illustrated by the following examples at four points

$$\lim_{\alpha'\to 0} Z(1234|1234) = \frac{1}{s_{12}} + \frac{1}{s_{23}} \,, \quad \lim_{\alpha'\to 0} Z(1234|1243) = -\frac{1}{s_{12}} \,, \quad \lim_{\alpha'\to 0} Z(1234|1423) = -\frac{1}{s_{23}} \,, \tag{6.97}$$

at five points

$$\begin{aligned} \lim_{\alpha'\to 0} Z(12345|12345) &= \frac{1}{s_{12}s_{34}} + \frac{1}{s_{23}s_{45}} + \frac{1}{s_{34}s_{51}} + \frac{1}{s_{45}s_{12}} + \frac{1}{s_{51}s_{23}} \,, \\ \lim_{\alpha'\to 0} Z(12345|12354) &= -\frac{1}{s_{12}s_{45}} - \frac{1}{s_{23}s_{45}} \,, \quad \lim_{\alpha'\to 0} Z(12345|13524) = 0 \end{aligned} \tag{6.98}$$

and at six points

$$\lim_{\alpha'\to 0} Z(123456|134256) = -\frac{1}{s_{234}s_{34}} \left(\frac{1}{s_{56}} + \frac{1}{s_{61}} \right) . \tag{6.99}$$

Relation to the inverse KLT kernel. As another central result of [35], doubly-partial amplitudes of bi-adjoint scalars are related to the *inverse* of the KLT matrix (4.160). More specifically, bases of $m(P|Q)$ under BCJ relations (6.95) form invertible $(n-3)! \times (n-3)!$ matrices with entries given by

$$m^{-1}(1,R,n-1,n|1,Q,n,n-1) = -S(R|Q)_1 \,. \tag{6.100}$$

The particular choices of BCJ bases on the left-hand side are consistent with the fact that the recursion (4.160) for the KLT matrix is tailored to the same BCJ bases of SYM amplitudes in (4.159). By (6.81),

[37] The convention for the sign of the Mandelstam invariants here is such that $m^{\text{here}}(P,n|Q,n) = (-1)^{|P|} m^{\text{there}}(P,n|Q,n)$ in comparison with the normalization of [35].

[38] While cyclic symmetry of $m(R|S)$ is not manifest from the Berends–Giele formula (6.94), it is built in from the definition (6.80) of doubly-partial amplitudes due to the cyclicity of the traces in t^i and $\tilde{t}^a$.

this implies that field-theory limits of disk integrals can also be assembled from the KLT matrix as firstly pointed out in [166].

As another consequence of (6.100), the relation (6.68) between the disk integrals ${F_P}^Q$ in terms of their Parke–Taylor analogues Z can be inverted to give [225]

$$Z(1,P,n{-}1,n|Q) = \sum_R m(Q|1,R,n{-}1,n){F_P}^R\,, \tag{6.101}$$

which is for instance instrumental to convert results on the α'-expansions of both sides. Moreover, the appearance (6.81) of doubly-partial amplitudes in the field-theory limit of Z-integrals can be derived from (6.101) and the field-theory limit (6.52) of ${F_P}^R$ on the right-hand side. This is not a circular conclusion since (6.52) is a necessity for the consistent reduction of open-string amplitudes to those of SYM under (6.50), and we will furthermore substantiate (6.52) through the method for its α'-expansion in section 8.5.

Inverse KLT matrix and Berends–Giele double currents. In terms of the Berends–Giele double currents the statement in (6.100) translates to the observation in [156] (see also [224]) later proved in [157]:

Lemma 13. *The matrix of Berends–Giele double currents $\phi_{iP|iR}$ of (6.92) is the inverse to the standard KLT matrix $S(R|Q)_i$ of (4.171)*

$$\sum_R \phi_{iP|iR}S(R|Q)_i = \delta_{P,Q}\,. \tag{6.102}$$

Proof. Taking the scalar product with iP of the result $\ell(iR) = \sum_Q S^\ell(iR|iQ)b(iQ)$ from lemma (4.180), we get $\langle iP,\ell(iR)\rangle = \sum_Q S^\ell(iR|iQ)\langle iP, b(iQ)\rangle$. That is, $\delta_{P,R} = \sum_Q S(R|Q)_i\phi(iQ|iP)$, where we used that $S^\ell(iR|iQ) = S(R|Q)_i$ in (4.171) and $\langle iP, b(iQ)\rangle = \phi(iQ|iP)$ in (6.92). ☐

A positive aspect of the formula (6.102) identifying the Berends–Giele double current as the inverse of the KLT matrix is that there is no need to choose the relative positions of $1, n{-}1, n$ like in (6.100) as no extraneous labels are present in (6.102). Moreover, this identity allows us to invert the relation (4.182),

$$V_{iP} = \sum_Q S(P|Q)_i M_{iQ} \quad\Longrightarrow\quad M_{iP} = \sum_Q \phi_{iP|iQ}V_{iQ}\,, \tag{6.103}$$

directly without reference to extra labels.

6.5. The field-theory limit of the superstring disk amplitudes

On the one hand, as reviewed in section 5.2, a closed formula for SYM tree-level amplitudes can be obtained using pure spinor cohomology methods as

$$A(1,2,\ldots,n{-}1,n) = \langle E_{12\ldots n-1}V_n\rangle = \sum_{XY=12\ldots n-1}\langle M_X M_Y M_n\rangle\,. \tag{6.104}$$

On the other hand, we know that the SYM tree-level amplitudes are obtained as the limit $\alpha'\to 0$ of the superstring amplitude (6.64). In the *non-local* KLT-representation (6.69) of the string amplitude, this follows from the field-theory limit (6.81) together with the inverse relation (6.100) between the KLT matrix and a $(n{-}3)!^2$ basis of doubly-partial amplitudes of bi-adjoint scalars.

The goal of this section is to give an alternative proof and to recover the cohomology formula (6.104) of SYM amplitudes from the *local* representation of the string amplitude

$$\mathcal{A}(1,2,\ldots,n) = -\sum_{\rho\in S_{n-3}}\sum_{XY=\rho(23\ldots n-2)}\big\langle V_{1X}V_{n-1,\tilde Y}V_n\big\rangle(-1)^{|Y|}Z(1,2,\ldots,n|1,X,n,Y,n{-}1)\,. \tag{6.105}$$

Proposition 21. *The field-theory limit of the pure spinor superstring amplitude in its local representation (6.105) yields the SYM tree-level formula (6.104)*

$$\lim_{\alpha'\to 0}\mathcal{A}(1,2,\ldots,n) = A(1,2,\ldots,n)\,. \tag{6.106}$$

Proof. The field-theory limit of the Z-integral in (6.105) with the canonical ordering $P = 12\ldots n$ in the domain is given by (6.96),

$$\lim_{\alpha'\to 0} Z(1,2,\ldots,n|Y,n{-}1,1,X,n) = s_{12\ldots n-1}\phi_{12\ldots n-1|Y(n-1)1X}\,, \tag{6.107}$$

where we cyclically rotated $1, X, n, Y, n{-}1 \to Y, n{-}1, 1, X, n$ to attain the form of $Z(\ldots,n|\ldots,n)$ with matching end labels. Note that when the domain is the canonical ordering, the deconcatenation formula (6.85) for $\phi_{12\ldots n-1|Y(n-1)1X}$ simplifies due to the constraint (6.84) and we get $s_{12\ldots n-1}\phi_{12\ldots n-1|Y(n-1)1X} = -\phi_{1A|1X}\phi_{B(n-1)|Y(n-1)}$ where $AB = 23\ldots n{-}2$ with $|A| = |X|$, $|B| = |Y|$. This means that we can write

$$\lim_{\alpha'\to 0} Z(1,2,\ldots,n|Y,n{-}1,1,X,n) = -\sum_{AB=23\ldots n-2} \phi_{1A|1X}\phi_{(n-1)\tilde{B}|(n-1)\tilde{Y}}\,, \tag{6.108}$$

where we used the reflection property (6.89) to rewrite $\phi_{B(n-1)|Y(n-1)} = \phi_{(n-1)\tilde{B}|(n-1)\tilde{Y}}$, and only a single term contributes to (6.108) where $|A| = |X|$ and $|Y| = |B|$. Therefore the limit of the string tree amplitude (6.105) as $\alpha' \to 0$ becomes

$$\begin{aligned}
\lim_{\alpha'\to 0} \mathcal{A}(1,2,\ldots,n) &= \sum_{\rho\in S_{n-3}}\ \sum_{XY=\rho(23\ldots n-2)}\ \sum_{AB=23\ldots n-2} \big\langle \big(\phi_{1A|1X}V_{1X}\big)\big(\phi_{(n-1)\tilde{B}|(n-1)\tilde{Y}}V_{n-1,\tilde{Y}}\big)V_n\big\rangle(-1)^{|Y|}\\
&= \sum_{AB=23\ldots n-2} \langle\Big(\sum_C \phi_{1A|1C}V_{1C}\Big)(-1)^{|B|}\Big(\sum_D \phi_{(n-1)\tilde{B}|(n-1)D}V_{(n-1)D}\Big)V_n\rangle \\
&= \sum_{AB=23\ldots n-2} \langle M_{1A}(-1)^{|B|}M_{(n-1)\tilde{B}}V_n\rangle = \sum_{XY=12\ldots n-1}\langle M_X M_Y V_n\rangle\\
&= \langle E_{12\ldots n-1}V_n\rangle = A(1,2,\ldots,n)\,,
\end{aligned} \tag{6.109}$$

where (6.84) implies that C and D in the second line only need to be summed over permutations of A and B, respectively. We could therefore insert

$$\sum_C \phi_{1A|1C}V_{1C} = M_{1A}\,, \qquad \sum_D \phi_{(n-1)\tilde{B}|(n-1)D}V_{(n-1)D} = M_{(n-1)\tilde{B}}\,, \tag{6.110}$$

from (6.103) followed by $M_{(n-1)\tilde{B}} = (-1)^{|B|}M_{B(n-1)}$ which finishes the proof. □

The above proof is valid for the canonical ordering $P = 123\ldots n$ due to (6.107). The generalization of the relation (6.106) for a general color ordering P was proposed in [224], see (7.42).

7. String and field-theory amplitude relations

In this section, we review the rich interplay of the results in the previous section with amplitude relations in field and string theory. Section 7.1 is dedicated to the color-kinematics duality in gauge theory and explicit realizations of kinematic Jacobi identities through the local representation of disk amplitudes in the $\alpha' \to 0$ limit. We focus on gravitational amplitudes in section 7.2, briefly review the KLT relations between open- and closed-string tree amplitudes and extract the cubic-diagram organization of the gravitational double copy from different representations of closed-string amplitudes. In section 7.3 we shall review the monodromy relations between open-superstring amplitudes with different disk orderings which furnish an elegant derivation of the BCJ relations among gauge-theory amplitudes.

The structure of disk amplitudes has implications for field-theory double-copy relations beyond gauge theories and gravity. In section 7.4, we shall discuss different representations of Born–Infeld amplitudes and manifest the color-kinematics duality of the non-linear sigma model of Goldstone bosons. Finally, the applications of the disk correlator for heterotic string theories will be discussed in section 7.5, along with the resulting amplitude relations for Einstein–Yang–Mills theory.

The discussions of this section does not rely on the detailed structure of the low-energy expansion of string tree amplitudes. As will be detailed in section 8, the coefficients in the α'-expansion of disk and sphere integrals exhibit an elegant pattern of multiple zeta values (MZVs). By organizing the string-corrections to SYM and supergravity amplitude according to their MZV content, we will find echoes of the field-theory structures of this section at all orders in α', see section 8.4.

7.1. Color-kinematics duality

This section is dedicated to an explicit realization of the color-kinematics duality in SYM tree amplitudes, based on the $\alpha' \to 0$ limit of superstring disk amplitudes. As we will see, this field-theory limit will naturally generate parameterizations of SYM amplitudes in terms of cubic diagrams whose kinematic factors obey the same Jacobi relations as their color factors. The *BCJ numerators* we will derive are simple combinations of the local building blocks $\langle V_{1A} V_{n-1,B} V_n \rangle$ in pure spinor superspace descending from the $(n-2)!$-term representation (6.8) of disk amplitudes.

7.1.1. Review of the color-kinematics duality

Our perspective on scattering amplitudes in gauge theories dramatically changed due to the seminal conjecture of Bern, Carrasco and Johansson in 2008 that their kinematic dependence can be arranged to exhibit the same symmetries as the color factors [29]. This *color-kinematics duality* holds for a variety of tree-level amplitudes and loop integrands of gauge theories with different numbers of supersymmetries. Together with the closely related gravitational double copy to be reviewed in section 7.2 below, the color-kinematics duality led to a large web of connections between field and string theories, see [30, 31] for reviews and [32] for a white paper.

Already at tree level, the color-kinematics duality is obscured in a Feynman-diagrammatic computation of $(n \geq 5)$-point amplitudes. The string-theoretic approach reviewed in this section led to the first explicit realizations of the color-kinematics duality in multiparticle tree-level amplitudes in 2011 [226]. In the first place, these results apply to ten-dimensional SYM, but they straightforwardly propagate to dimensional reductions including $\mathcal{N} = 4$ SYM in four dimensions. In fact, the manifestations of color-kinematics duality in this section also apply to pure Yang–Mills since its gluon amplitudes are the same as in maximally supersymmetric gauge theories.

Cubic-diagram parameterization. The color-kinematics duality relies on an elementary observation on tree amplitudes or loop integrands of pure or supersymmetric YM theory: any dependence on the adjoint degrees of freedom (or *color dependence* in short) of the external states occurs via contractions of the structure constants f^{abc}. In Feynman diagrams with exclusively cubic vertices, these contractions are straightforwardly determined by dressing internal lines with δ^{ab} and vertices with f^{abc}.

While any non-abelian gauge-theory Lagrangian features a quartic interaction $\sim \mathrm{Tr}([\mathbb{A}^m, \mathbb{A}^n][\mathbb{A}_m, \mathbb{A}_n])$, its color structure $f^{abe} f^{ecd}$ still resembles cubic diagrams. Each quartic vertex bypasses one of the propagators of the cubic diagrams, but one can still enforce a uniform number of propagators for all gauge-theory diagrams by inserting $1 = \frac{p^2}{p^2}$ with suitably chosen momenta p for each quartic vertex. As illustrated in figure 5, this amounts to expanding each quartic-vertex contribution in a channel $1 = \frac{s_{i\ldots j}}{s_{i\ldots j}}$ that is compatible with the accompanying color factors.

Hence, it is always possible to parameterize gauge-theory trees and loop integrands in terms of cubic diagrams i whose propagators D_i and color factors c_i can be straightforwardly read off from the cubic vertices and internal lines. For color-dressed tree amplitudes, this parameterization reads[39]

$$M_n^{\mathrm{gauge}} = \sum_{i \in \Gamma_n} \frac{c_i N_i}{D_i} \,. \tag{7.1}$$

The associated kinematic numerators N_i encoding all the dependence on momenta and polarizations receive contributions from various Feynman diagrams with different numbers of quartic vertices. The symbol Γ_n in

[39] We depart from our notation M_n for color-dressed SYM amplitudes to later on compare M_n^{gauge} in (7.1) with gravitational amplitudes M_n^{grav} and those of bi-adjoint scalars in (6.80).

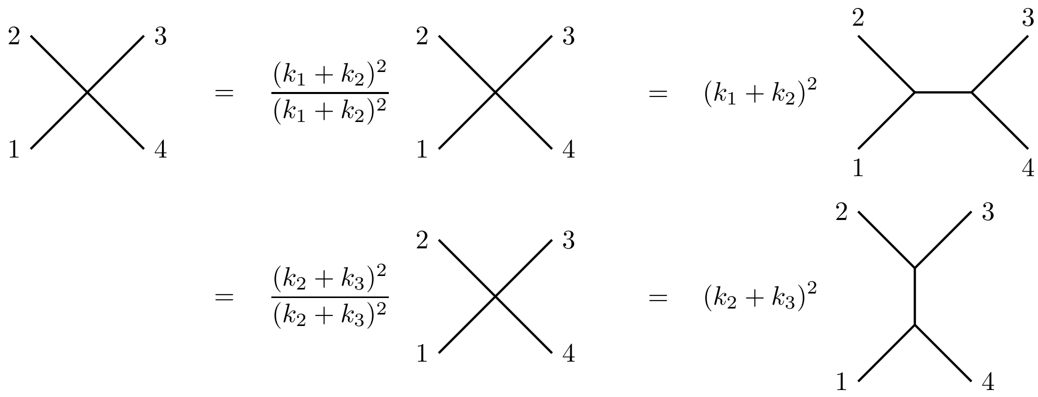


Figure 5: Two possibilities of expanding a quartic vertex: the first line is compatible with a color factor $f^{a_1a_2b}f^{ba_3a_4}$ while the second line captures the second term in rewriting the color factor as $f^{a_1a_3b}f^{ba_2a_4} - f^{a_2a_3b}f^{ba_1a_4}$ via the Jacobi identity.

the summation range of (7.1) denotes the set of $(2n-5)!!$ cubic tree diagrams with n external legs that are inequivalent under flips of cubic vertices.

However, the Jacobi identity

$$f^{abe}f^{ecd} + f^{bce}f^{ead} + f^{cae}f^{ebd} = 0 \tag{7.2}$$

introduces ambiguities in the alignment of quartic-vertex contributions with the propagator structure of cubic diagrams. These ambiguities illustrated in figure 5 lead to immense freedom in moving terms between the N_i of different cubic diagrams. This freedom was initially referred to as generalized gauge invariance [29, 227, 228] and later on related to non-abelian gauge transformations of perturbiners [84], for instance the transformation (4.112) mediating between Lorenz and BCJ gauge (see [85] for an all-order expression).

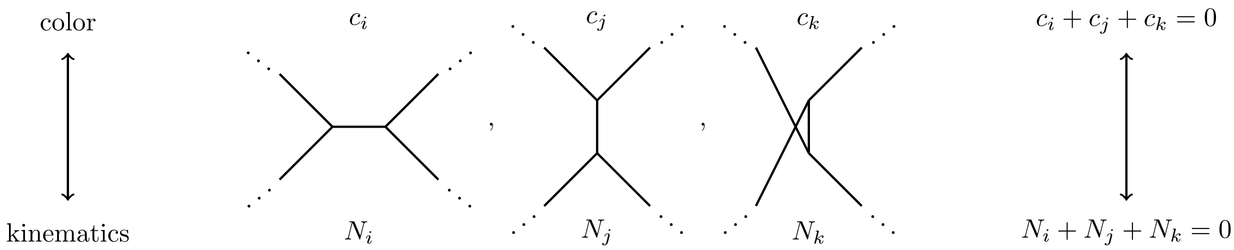


Figure 6: Triplets of cubic graphs whose color factors c_i and kinematic factors N_i are both related by a Jacobi identity if the duality between color and kinematics is manifest. The dotted lines at the corners represent arbitrary tree-level subdiagrams and are understood to be the same for all of the three cubic graphs.

Kinematic Jacobi identities. For all triplets of cubic diagrams $i, j, k \in \Gamma_n$ that share all propagators except for one, see figure 6, the Jacobi identity (7.2) implies that the associated color factors obey $c_i + c_j + c_k = 0$. According to the color-kinematics duality, one can choose the numerators N_l in (7.1) such that the *kinematic Jacobi identity* $N_i + N_j + N_k = 0$ holds for each such triplet i, j, k. Moreover, the antisymmetry $f^{abc} = f^{[abc]}$ implies that color factors c_i change their sign upon flipping any of the cubic vertices. Kinematic numerators with manifest color-kinematics duality are understood to also change $N_i \to -N_i$ under flips of cubic vertices in diagram i. In other words,

$$\text{manifest color-kinematics duality}: \left\{ \begin{array}{rcl} c_i + c_j + c_k = 0 & \Longrightarrow & N_i + N_j + N_k = 0 \ \forall \ i,j,k \in \Gamma_n \,, \\ c_i \to -c_i & \Longrightarrow & N_i \to -N_i \ \forall \ i \in \Gamma_n \,. \end{array} \right. \tag{7.3}$$

Examples up to four points. The three-point instance of the gauge-amplitude parameterization (7.1) in ten-dimensional SYM reduces to a single diagram without any propagators $D_i \to 1$, with color factor $c_i \to f^{123}$ and kinematic numerator

$$N_i \to \langle V_1 V_2 V_3 \rangle = (e_1 \cdot k_2)(e_2 \cdot e_3) + e_1^m (\chi_2 \gamma_m \chi_3) + \mathrm{cyc}(1,2,3) \,. \tag{7.4}$$

Here and below, we use the shorthand $a_i \to i$ for the adjoint indices of the $i^{\rm th}$ external state, e.g. write f^{123} in the place of $f^{a_1 a_2 a_3}$.

The first instance of quartic-vertex contributions arises at four points. The parameterization (7.1) comprises three diagrams in the s-, t- and u-channel associated with inverse propagators $s = s_{12},\ t = s_{23}$ and $u = s_{13} = -s - t$,

$$M_4^{\rm gauge} = \frac{N_s c_s}{s} + \frac{N_t c_t}{t} + \frac{N_u c_u}{u} \,. \tag{7.5}$$

The color factors are indexed by the relevant channel, and their Jacobi identity literally matches (7.2)

$$\left. \begin{array}{r} c_s = f^{12a} f^{a34} \\ c_t = f^{23a} f^{a14} \\ c_u = f^{31a} f^{a24} \end{array} \right\} \quad \Longrightarrow \quad c_s + c_t + c_u = 0 \,. \tag{7.6}$$

One admissible choice of numerators in ten-dimensional SYM reads

$$N_s = \langle V_{12} V_3 V_4 \rangle \,, \qquad N_t = \langle V_{23} V_1 V_4 \rangle \,, \qquad N_u = \langle V_{31} V_2 V_4 \rangle \,, \tag{7.7}$$

and they obey the kinematic Jacobi identity by BRST exactness of [159]

$$N_s + N_t + N_u = \langle (V_{12} V_3 + V_{23} V_1 + V_{31} V_2) V_4 \rangle = -\frac{1}{s_{12}} \langle Q (V_{123} V_4) \rangle = 0 \tag{7.8}$$

using (4.73) and $s_{13} + s_{23} = -s_{12}$ in the momentum phase space of four massless particles. Still, any other choice of $\{N_s, N_t, N_u\}$ besides (7.7) that yields the same amplitude (7.5) will obey kinematic Jacobi identities: this can be seen by adding $0 = K(\frac{s c_s}{s} + \frac{t c_t}{t} + \frac{u c_u}{u})$ to $M_4^{\rm gauge}$ with an arbitrary kinematic factor K which modifies the numerators in (7.5) by $\delta N_s = sK$, $\delta N_t = tK$ and $\delta N_u = uK$. The modification to the triplet in the kinematic Jacobi identity (7.8) then vanishes by momentum conservation,

$$\delta(N_s + N_t + N_u) = K(s + t + u) = 0 \,. \tag{7.9}$$

Examples at five points. At five points, the cubic-diagram parameterization (7.1) involves $5!! = 15$ terms

$$M_5^{\rm gauge} = \frac{N_{12|3|45} c_{12|3|45}}{s_{12} s_{45}} + \frac{N_{14|3|25} c_{14|3|25}}{s_{14} s_{25}} + \frac{N_{15|3|24} c_{15|3|24}}{s_{15} s_{24}} + \mathrm{cyc}(1,2,3,4,5) \tag{7.10}$$

with color factors $c_{ab|d|gh} = f^{abi} f^{idj} f^{jgh}$ subject to Jacobi identities $c_{ab|[d|gh]} = c_{[ab|d]|gh} = 0$. However, generic choices of kinematic numerators $N_{ab|d|gh}$ – say a naive Feynman-diagram computation or a crossing symmetric choice $N_{ab|d|gh} \to \langle V_{ab} V_d V_{gh} \rangle$ – will fail to obey kinematic Jacobi identities even though they yield the correct color-dressed amplitude (7.10).

Still, reparametrizations of the amplitude (7.10) will generically modify the three-term sum of numerators that decide about kinematic Jacobi identities: adding $0 = K_{45}(\frac{s_{12} c_{12|3|45}}{s_{12} s_{45}} + \frac{s_{23} c_{23|1|45}}{s_{23} s_{45}} + \frac{s_{13} c_{31|2|45}}{s_{13} s_{45}})$ with some kinematic factor K_{45} modifies three of the numerator factors, $\delta N_{12|3|45} = s_{12} K_{45}$, $\delta N_{23|1|45} = s_{23} K_{45}$ and $\delta N_{31|2|45} = s_{13} K_{45}$, while leaving the remaining 12 inert [229]. The sum of the three numerators which vanishes in a manifestly color-kinematics dual parameterization is modified by the above reparametrization via

$$\delta(N_{12|3|45} + N_{23|1|45} + N_{31|2|45}) = (s_{12} + s_{23} + s_{13}) K_{45} = s_{45} K_{45} \neq 0 \,. \tag{7.11}$$

One can similarly check that, in the naive crossing-symmetric choice $N_{ab|d|gh} \to \langle V_{ab} V_d V_{gh} \rangle$, the superspace expression $(V_{12} V_3 + V_{23} V_1 + V_{31} V_2) V_{45}$ is not BRST closed. In this way, the validity of the kinematic Jacobi identity $N_{[12|3]|45} = 0$ is seen to be gauge dependent. Hence, it is generically a matter of a suitable parameterization of the gauge-theory amplitude whether the color-kinematics duality is manifest or not.

Relation to the color decomposition. In order to extract color-ordered gauge-theory amplitudes $A(1,2,\ldots,n)$ from the cubic-diagram representation (7.1) of the color-dressed amplitude, one relies on the unique expansion of the color factors c_i in terms of traces of gauge-group generators $\mathrm{Tr}(t^P) = \mathrm{Tr}(t^{p_1}t^{p_2}\ldots t^{p_n})$. The above four- and five-point examples give[40]

$$\begin{aligned} c_s &= \mathrm{Tr}(t^1t^2t^3t^4 - t^1t^2t^4t^3 - t^2t^1t^3t^4 + t^2t^1t^4t^3)\,, \\ c_{12|3|45} &= \mathrm{Tr}(t^1t^2t^3t^4t^5 - t^1t^2t^3t^5t^4 - t^2t^1t^3t^4t^5 + t^2t^1t^3t^5t^4) - (t^1t^2 \leftrightarrow t^4t^5)\,, \end{aligned} \tag{7.12}$$

see section 2.1 of [230] for a general algorithm for arbitrary color factors c_i. Hence, the color-ordered amplitudes obtained from (7.1) are sums of diagrams

$$A(P) = \sum_{i\in\Gamma_n} \frac{N_i}{D_i} c_i \,\big|_{\mathrm{Tr}(t^P)}\,, \tag{7.13}$$

where the coefficients take values in $c_i \,\big|_{\mathrm{Tr}(t^P)} \in \{0,1,-1\}$. The four- and five-point instances in the above notation

$$\begin{aligned} A(1,2,3,4) &= \frac{N_s}{s} - \frac{N_t}{t}\,, \\ A(1,2,3,4,5) &= \frac{N_{12|3|45}}{s_{12}s_{45}} + \mathrm{cyc}(1,2,3,4,5) \end{aligned} \tag{7.14}$$

are clearly invariant under the reparametrizations $\sim K, K_{45}$ of the numerators in the color-dressed amplitude.

BCJ amplitude relations from the color-kinematics duality. The BCJ relations (5.62) among color-ordered gauge-theory amplitudes were firstly derived in [29] by assuming the existence of color-kinematics dual numerators (7.3). For instance, inserting $N_u = -N_s - N_t$ into

$$A(1,2,3,4) = \frac{N_s}{s} - \frac{N_t}{t}\,, \qquad A(2,3,1,4) = \frac{N_t}{t} - \frac{N_u}{u} \tag{7.15}$$

leads to the BCJ relation $A(2,3,1,4) = \frac{s}{u}A(1,2,3,4)$. However, BCJ relations are gauge-independent statements, i.e. they affect color-ordered amplitudes which do not depend on reparametrizations of (7.1). Hence, the gauge dependent kinematic Jacobi relations cannot be necessary conditions for BCJ amplitude relations. Instead, they are sufficient conditions as shown in [29].

7.1.2. DDM form of YM and bi-adjoint scalar amplitudes

In preparation for our proof that the $\alpha' \to 0$ limit of disk amplitudes yields numerators N_i subject to all kinematic Jacobi relations, we introduce the so-called Del Duca–Dixon–Maltoni (DDM) representation of color-dressed gauge-theory amplitudes. The color decomposition of M_n^{gauge} in terms of $(n{-}1)!$ color traces does not expose that the latter conspire to products of $n{-}2$ structure constants as required by Feynman rules. Only after exhaustive use of KK relations (5.40) among color-ordered amplitudes, one can see that the color decomposition simplifies to contracted f^{abc} in the coefficients of the KK independent $A(1,P,n)$ with $P \in S_{n-2}$ a permutation of $2,3,\ldots,n{-}1$. This kind of reduction by KK relations is known as the DDM form [195]

$$M_n^{\mathrm{gauge}} = \sum_{P\in S_{n-2}} c_{1|P|n} A(1,P,n)\,, \qquad c_{1|P|n} := f^{1p_2a}f^{ap_3b}\ldots f^{yp_{n-2}z}f^{zp_{n-1}n}\,, \tag{7.16}$$

where the color factor $c_{1|P|n}$ with $P = p_2p_3\ldots p_{n-1}$ corresponds to the cubic diagram of half-ladder topology in figure 7 below. In fact, the collection of $\{c_{1|P|n},\ P\in S_{n-2}\}$ furnishes an $(n{-}2)!$-element basis of all the $(2n{-}5)!!$ color factors c_i under Jacobi identities.[41] Accordingly the $(n{-}2)!$-family of half-ladders in figure 7 is referred to as the *master diagrams.*

[40]We are following normalization conventions $[t^a,t^b] = f^{abc}t^c$ and $\mathrm{Tr}(t^at^b) = \delta^{ab}$ which leads to $\mathrm{Tr}(t^1[t^2,t^3]) = f^{123}$ at three points and the coefficients ±1 in (7.12).

[41]This is a consequence of the fact that arbitrary Lie monomials built from non-commuting $t^{i_1}, t^{i_2},\ldots,t^{i_k}$ can be expanded in a Dynkin bracket basis of $[[\ldots[[t^1,t^{\rho(2)}],t^{\rho(3)}],\ldots,t^{\rho(k-1)}],t^{\rho(k)}]$, see section 4.1.5.

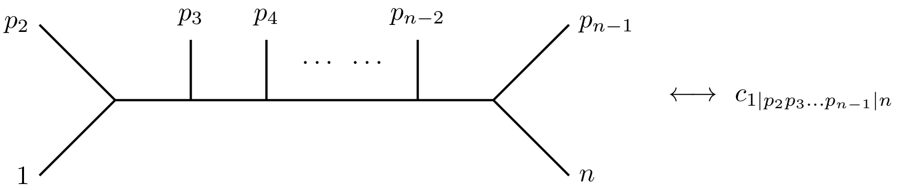


Figure 7: Master diagrams whose color factors $c_{1|P|n}$ defined in (7.16) are independent under Jacobi relations.

In the same way as (7.16) follows from KK relations of the color-order amplitudes in $M_n^{\rm gauge}$, one can start from the double color decomposition (6.80) of bi-adjoint scalars and exhaustively insert KK relations of their doubly-partial amplitudes $m(A|B)$. After expanding both entries A and B in $(n{-}2)!$-term KK bases, (6.80) takes the form [35]

$$M_n^{\phi^3} = \sum_{P,Q\in S_{n-2}} c_{1|P|n} m(1,P,n|1,Q,n)\tilde c_{1|Q|n} \tag{7.17}$$

by analogy with (7.16), where $\tilde c_{1|Q|n}$ is the half-ladder in figure 7 with $\tilde f^{abc}$ in the place of f^{ijk}.

At the same time, color-dressed ϕ^3 amplitudes can be written as

$$M_n^{\phi^3} = \sum_{i\in\Gamma_n} \frac{c_i \tilde c_i}{D_i} \tag{7.18}$$

as one can see from the straightforward Feynman-diagram computation with only one cubic vertex $\sim f^{ijk}\tilde f^{abc}$ in the Lagrangian (6.79). While (7.18) involves the complete $(2n{-}5)!!$-element collection of $c_i, \tilde c_i$ with $i\in\Gamma_n$ related by Jacobi identities, the equivalent form (7.17) of $M_n^{\phi^3}$ only features the color factors $c_{1|P|n}, \tilde c_{1|Q|n}$ of the master diagrams in figure 7 under Jacobi relations. Hence, by equating (7.17) and (7.18), the doubly-partial amplitudes $m(1,P,n|1,Q,n)$ turn out to summarize the net effect of solving *all* Jacobi relations.

The last observation can be used to rewrite the color-dressed gauge-theory amplitude (7.1): in a color-kinematics dual representation with N_i obeying the same Jacobi identities as c_i, the expansion of $M_n^{\rm gauge}$ in terms of color and kinematic factors of master diagrams must take the form

$$M_n^{\rm gauge} = \sum_{P,Q\in S_{n-2}} c_{1|P|n} m(1,P,n|1,Q,n) N_{1|Q|n}\,. \tag{7.19}$$

This can be understood from the fact that (7.17) follows from (7.18) solely by application of Jacobi identities among c_i, irrespective of their detailed form, and our assumption that the N_i obey the same Jacobi identities as the $\tilde c_i$. The $(n{-}2)!$-family of $N_{1|Q|n}$ in (7.19) is again associated with the half-ladder diagrams in figure 7 and referred to as *master numerators*. Indeed, all the $(2n{-}5)!!$ instances of N_i in a color-kinematics dual parameterization (7.1) must be combinations of $N_{1|Q|n}$ with coefficients in $\{0,1,-1\}$ determined by (7.19). In other words, any parameterization (7.19) of gauge-theory amplitudes implies all kinematic Jacobi relations of the cubic-diagram numerators since the same is evidently true in the ϕ^3 case (7.17) and (7.18).

In summary, we have encountered two representations of color-dressed tree amplitudes of gauge theories and bi-adjoint scalars: cubic-diagram expansions (7.1) and (7.18) related by trading kinematic numerators for another species of color factors $\tilde c_i \leftrightarrow N_i$. While cubic-diagram expansions still exist if some of the kinematic Jacobi relations are violated, the $(n{-}2)!^2$-term representations (7.17) and (7.19) are tied to Jacobi relations reducing all of $c_i, \tilde c_i, N_i$ to an $(n{-}2)!$ basis. One can again relate the gauge-theory amplitude (7.19) to the bi-adjoint scalar amplitude (7.17) by exchanging the kinematic master numerators with corresponding color factors, $N_{1|Q|n} \leftrightarrow \tilde c_{1|Q|n}$.

7.1.3. Local BCJ numerators from disk amplitudes

We shall now take advantage of the representation (7.19) of color-dressed gauge-theory amplitudes to retrieve the Jacobi relations of the kinematic numerators obtained from the $\alpha'\to 0$ limit of n-point disk amplitudes (6.8). By matching the DDM form (7.16) of color-dressed gauge-theory amplitudes with the expansion (7.19) in terms of master numerators, color-ordered n-point amplitudes are found to take the form

$$A_n(P) = \sum_{Q\in S_{n-2}} m(P|1,Q,n) N_{1|Q|n} \,. \tag{7.20}$$

We also made use of the linear independence of the color factors $c_{1|P|n}$ associated with the master diagrams in figure 7 and the fact that $A(P)$ and $m(P|1,Q,n)$ obey the same KK relations in P.

It turns out that (7.20) is precisely the form of $A(P)$ obtained in the field-theory limit of disk amplitudes: as detailed in section 6.4.1, adapting (6.8) to a generic $\mathrm{SL}_2(\mathbb{R})$ frame leads to the $(n-2)!$ term representation (6.64) of n-point disk amplitudes in terms of Parke–Taylor or Z-integrals (6.62) [166]

$$\mathcal{A}_n(P) = \sum_{XY=23\ldots n-2} (-1)^{|Y|+1} Z(P|1,X,n,Y,n-1)\langle V_{1X} V_{(n-1)\tilde{Y}} V_n\rangle + \mathrm{perm}(2,3,\ldots,n-2)\,, \tag{7.21}$$

for instance

$$\begin{aligned}
\mathcal{A}_4(P) &= -Z(P|1,2,4,3)\langle V_{12}V_3V_4\rangle + Z(P|1,4,2,3)\langle V_1V_{32}V_4\rangle\,, \qquad (7.22)\\
\mathcal{A}_5(P) &= -Z(P|1,2,3,5,4)\langle V_{123}V_4V_5\rangle + Z(P|1,2,5,3,4)\langle V_{12}V_{43}V_5\rangle - Z(P|1,5,2,3,4)\langle V_1V_{432}V_5\rangle + (2\leftrightarrow 3)\,,\\
\mathcal{A}_6(P) &= -Z(P|1,2,3,4,6,5)\langle V_{1234}V_5V_6\rangle + Z(P|1,2,3,6,4,5)\langle V_{123}V_{54}V_6\rangle\\
&\quad - Z(P|1,2,6,3,4,5)\langle V_{12}V_{543}V_6\rangle + Z(P|1,6,2,3,4,5)\langle V_1V_{5432}V_6\rangle + \mathrm{perm}(2,3,4)\,.
\end{aligned}$$

Here and in later equations, the sum over permutations of $2,3,\ldots,n-2$ is understood to not act on the labels in the integration domain P.

Given that the field-theory limit (6.81) of the Z-integrals yields doubly-partial amplitudes, the SYM amplitudes resulting from (7.21) are given by

$$\begin{aligned}
A_n(P) &= \sum_{XY=23\ldots n-2} (-1)^{|Y|+1} m(P|1,X,n,Y,n-1)\langle V_{1X}V_{(n-1)\tilde{Y}}V_n\rangle + \mathrm{perm}(2,3,\ldots,n-2)\\
&= \sum_{Q\in S_{n-2}} m(P|1,Q,n-1) N_{1|Q|n-1} \qquad (7.23)
\end{aligned}$$

with master numerators [226]

$$N_{1|XnY|n-1} = (-1)^{|Y|-1}\langle V_{1X}V_{(n-1)\tilde{Y}}V_n\rangle\,, \tag{7.24}$$

for instance

$$N_{1|23\ldots jn(j+1)\ldots n-2|n-1} = (-1)^{n-j-1}\langle V_{12\ldots j}V_{n-1,n-2\ldots j+1}V_n\rangle\,. \tag{7.25}$$

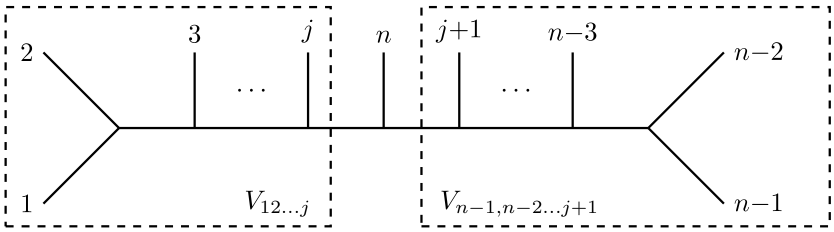
The second line of (7.23) exposes the expansion of $A(P)$ in an $(n-2)!$ family of $m(P|1,Q,n-1)$ (with Q a permutation of $2,3,\ldots,n-2,n$) characteristic to color-kinematic dual representations of SYM amplitudes. Since (7.23) is related to the color-kinematics dual form (7.20) via $n\leftrightarrow n-1$, we identify the local superfields (7.24) as the BCJ master numerators of the half-ladder diagrams in figure 7 with $n-1$ in the place of n. In fact, by the diagrammatic interpretation of V_P in figure 1, the right-hand side of (7.24) organizes the master diagrams into three subdiagrams as visualized in figure 8.

We spell out the simplest examples at four points

$$N_{1|24|3} = -\langle V_{12}V_3V_4\rangle\,, \qquad N_{1|42|3} = \langle V_1V_{32}V_4\rangle \tag{7.26}$$

and at five points

$$\begin{aligned}
N_{1|235|4} &= -\langle V_{123}V_4V_5\rangle\,, \quad N_{1|253|4} = \langle V_{12}V_{43}V_5\rangle\,, \quad N_{1|523|4} = -\langle V_1V_{432}V_5\rangle\,, \qquad (7.27)\\
N_{1|325|4} &= -\langle V_{132}V_4V_5\rangle\,, \quad N_{1|352|4} = \langle V_{13}V_{42}V_5\rangle\,, \quad N_{1|532|4} = -\langle V_1V_{423}V_5\rangle\,.
\end{aligned}$$



$$\longleftrightarrow \ \langle V_{12\ldots j} V_{n-1,n-2\ldots j+1} V_n \rangle$$

Figure 8: The mapping between master numerators and expressions in pure spinor superspace according to (7.25).

The superspace numerators (7.24) enter the representations (7.23) of color-ordered gauge-theory amplitudes that are hallmarks of manifest color-kinematics duality by the discussion of section 7.1.2. Hence, the master numerators (7.24) determine all other cubic-diagram numerators in (7.1) by a sequence of kinematic Jacobi identities, and the coefficients can be conveniently determined by isolating the propagators of interest from the doubly-partial amplitudes in (7.23). Moreover, the master numerators are local, i.e. free of kinematic poles, by the construction of multiparticle superfields A^P_α in BCJ gauge that enter $V_P = \lambda^\alpha A^P_\alpha$, see section 4.1. On these grounds, the superspace expressions (7.24) are referred to as local BCJ numerators [226].

7.1.4. Component expansion of BCJ numerators

In order to extract the superspace components from master numerators $\langle V_X V_Y V_Z \rangle$, it is convenient to combine BCJ gauge for the superfields with the non-linear version of Harnad–Shnider gauge for their θ-expansion, see section 4.3 of [84] and Appendix F. In this BCJ–Harnad–Shnider gauge, the relevant orders in θ are,

$$V_P = \frac{1}{2}(\lambda\gamma_m\theta)e^m_P + \frac{1}{3}(\lambda\gamma_m\theta)(\theta\gamma^m\chi_P) - \frac{1}{32}(\lambda\gamma^m\theta)(\theta\gamma_{mnq}\theta)f^{nq}_P + \mathcal{O}(\theta^4)\,, \tag{7.28}$$

with local multiparticle polarizations $e^m_P, \chi^\alpha_P, f^{mn}_P$ defined by (4.122) in the place of the single-particle polarizations $e^m_i, \chi^\alpha_i, f^{mn}_i$ in (2.17). Similar to the discussion in section 5.2.2, this organization of the θ-expansion reduces the component expansion at all multiplicities,

$$\langle V_X V_Y V_Z \rangle = \frac{1}{2} e^m_X f^{mn}_Y e^n_Z + (\chi_X\gamma_m\chi_Y)e^m_Z + \mathrm{cyc}(XYZ)\,, \tag{7.29}$$

to the $\lambda^3\theta^5$ correlators (3.96) and (3.100) of the three-point amplitude [165].

7.1.5. The Möbius product

While the SYM amplitudes $A(P,n) = \langle E_P V_n \rangle$ satisfy the BCJ amplitude relations, a naive relabeling of P does not lead to numerators that satisfy the color-kinematics duality. As discussed in [226] and reviewed above, the way string theory disk amplitudes give rise to a local representation of numerators satisfying the color-kinematics duality is via the field-theory limit of the pure spinor parameterization of the correlator with $(n{-}2)!$ numerators of the form $\langle V_{1P} V_{(n-1)Q} V_n \rangle$ with each multiparticle vertex V_R in the BCJ gauge reviewed in section 4.1.6. This parameterization is generated by (7.20), and its essential feature is the distribution of the labels $1, n{-}1, n$ into three separate superfields V_R within master numerators (7.24). This splitting can be traced back to the fixing of the Möbius invariance of the disk correlator in (3.75).

The field-theory limit of the disk integrals with different disk orderings, given by bi-adjoint amplitudes (6.96), does not modify this label distribution in the numerators, while relabeling the color ordering of $A(P,n)$ in (5.14) does. Note, however, that (5.14) and the field-theory limit of the string disk amplitude manifestly coincide for the canonical ordering $P = 12\ldots n$ (and in fact for a $(n{-}3)!$ basis of color-orderings $P = 1R(n{-}1)n$), as demonstrated in (6.109). Let us illustrate the above point with an example.

The pure spinor formula (5.14) for the ordering $P = 12435$ in $A(1,2,4,3,5) = \langle E_{1243} V_5 \rangle$ yields

$$A(1,2,4,3,5) = \frac{\langle V_{124}V_3V_5\rangle}{s_{12}s_{124}} + \frac{\langle V_{421}V_3V_5\rangle}{s_{24}s_{124}} + \frac{\langle V_{12}V_{43}V_5\rangle}{s_{12}s_{34}} + \frac{\langle V_1V_{243}V_5\rangle}{s_{24}s_{34}} + \frac{\langle V_1V_{342}V_5\rangle}{s_{34}s_{234}}\,, \tag{7.30}$$

while the $\alpha' \to 0$ limit of the superstring amplitude (6.64) with the same ordering gives, after using (6.96),

$$\begin{aligned}\lim_{\alpha'\to 0} \mathcal{A}(1,2,4,3,5) = \frac{\langle (V_{12}V_{43} + V_{123}V_4)V_5\rangle}{s_{12}s_{124}} - \frac{\langle (V_1V_{423} + V_{13}V_{42})V_5\rangle}{s_{24}s_{124}} + \frac{\langle V_{12}V_{43}V_5\rangle}{s_{34}s_{12}} \\ - \frac{\langle V_1V_{432}V_5\rangle}{s_{34}s_{234}} - \frac{\langle V_1V_{423}V_5\rangle}{s_{24}s_{234}} \,.\end{aligned} \tag{7.31}$$

These two amplitudes must be equal, but the numerators differ as only the latter preserves the superfield structure $V_{1A}V_{(n-1)\tilde{B}}V_n$ due to fixing $(z_1, z_{n-1}, z_n) \to (0, 1, \infty)$ by Möbius invariance. Comparing both amplitudes we see the correspondence

$$\begin{aligned} &V_{124}V_3 \to V_{12}V_{43} + V_{123}V_4\,, && V_1V_{243} \to -V_1V_{423}\,, \\ &V_{421}V_3 \to -V_1V_{423} - V_{13}V_{42}\,, && V_1V_{342} \to -V_1V_{432}\,. \end{aligned} \tag{7.32}$$

In [224] an algorithm was proposed that reproduces the above correspondences. The idea is to guarantee that the two labels i and j (usually $i = 1$ and $j = n-1$) always appear in two separate vertices[42]. Therefore the algorithm redistributes the labels i and j between two vertices if they originally appear inside a single vertex V_{iAjB} and does nothing otherwise. To this effect, we define the *Möbius product* $\circ_{ij}$ as [224]

$$V_{iAjB} \circ_{ij} V_C := \sum_{\delta(B\dot{\ell}(C)) = R\otimes S} V_{iAR}V_{jS}\,, \qquad V_{AiB} \circ_{ij} V_{CjD} := V_{AiB}V_{CjD}\,, \tag{7.33}$$

where $\dot{\ell}(C)$ denotes the letterification (C.12) of the Dynkin bracket $\ell(C)$ of (C.1) and $\delta(P)$ is the deshuffle map (C.10). Note that it is always possible to move label i to the front using (4.58), $V_{PiQ} = -V_{i\ell(P)Q}$, so these two rules are sufficient. In summary, the mapping (7.33) ensures that the labels i and j are split between the two vertices V_A and V_B in the product $V_A \circ_{ij} V_B$. The choice $i = 1$ and $j = n-1$ corresponds to the usual fixing of the Möbius symmetry of the disk. For example applications of (7.33) we list

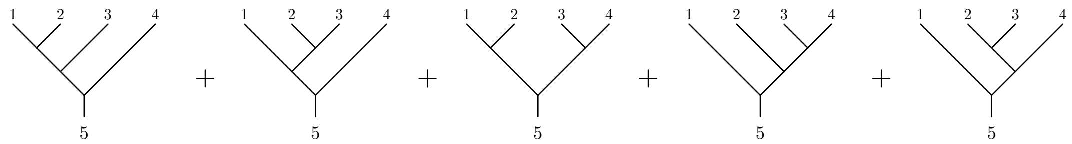
$$\begin{aligned}
V_{124} \circ_{14} V_3 &= V_{12}V_{43} + V_{123}V_4\,, \\
V_{142} \circ_{14} V_3 &= V_1V_{423} + V_{12}V_{43} + V_{123}V_4 + V_{13}V_{42}\,, \\
V_{421} \circ_{14} V_3 &= -V_1V_{423} - V_{13}V_{42}\,, \\
V_{143} \circ_{14} V_2 &= V_1V_{432} + V_{13}V_{42} + V_{12}V_{43} + V_{132}V_4\,, \\
V_{134} \circ_{14} V_2 &= V_{13}V_{42} + V_{132}V_4\,, \\
V_{1235} \circ_{15} V_4 &= V_{123}V_{54} + V_{1234}V_5\,, \\
V_{1253} \circ_{15} V_4 &= V_{12}V_{534} + V_{123}V_{54} + V_{1234}V_5 + V_{124}V_{53}\,, \\
V_{1523} \circ_{15} V_4 &= V_1V_{5234} + V_{12}V_{534} + V_{123}V_{54} + V_{1234}V_5 \\
&\quad + V_{124}V_{53} + V_{13}V_{524} + V_{134}V_{52} + V_{14}V_{523}\,, \\
V_{1235} \circ_{15} V_{46} &= V_{123}V_{546} - V_{123}V_{564} + V_{12346}V_5 - V_{12364}V_5\,, \\
V_{152} \circ_{15} V_{34} &= V_1V_{5234} - V_1V_{5243} + V_{12}V_{534} - V_{12}V_{543} + V_{1234}V_5 \\
&\quad - V_{1243}V_5 + V_{134}V_{52} - V_{143}V_{52}\,.
\end{aligned} \tag{7.34}$$

From the translation $V_AV_B \to [\ell(A), \ell(B)]$ we obtain the free-Lie-algebra interpretation the above mapping: it is a rewriting system of nested commutators from $[\ell(iAjB), \ell(C)]$ to a basis[43] of brackets of the form $[\ell(iP), \ell(jQ)]$. For instance, the first example in (7.34) is equivalent to

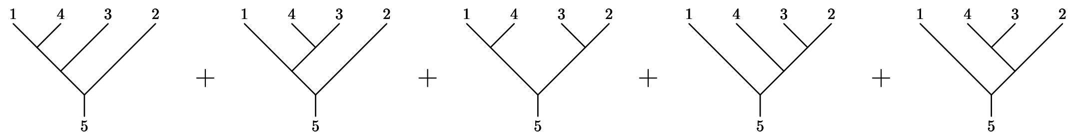
$$[[[1,2],4],3] = [[1,2],[4,3]] + [[[1,2],3],4]\,, \tag{7.35}$$

[42] The third vertex V_n can always be fixed with label n by cyclic invariance.

[43] Despite appearances, this is not a $(n-2)!$ dimensional basis of the free Lie algebra but a $(n-1)!$ one, even after the fixing of two letters i and j. The reason is that the lengths of P and Q are not fixed. The simplest example is the case $n = 3$ where the $(n-1)! = 2$ dimensional basis is $[\ell(1), \ell(23)]$ and $[\ell(13), \ell(2)]$.



$$A(12345) = \frac{\langle V_{[[1,2],3]}\circ_{14}V_4V_5\rangle}{s_{12}s_{123}} + \frac{\langle V_{[1,[2,3]]}\circ_{14}V_4V_5\rangle}{s_{23}s_{123}} + \frac{\langle V_{[1,2]}\circ_{14}V_{[3,4]}V_5\rangle}{s_{12}s_{34}} + \frac{\langle V_1\circ_{14}V_{[2,[3,4]]}V_5\rangle}{s_{34}s_{234}} + \frac{\langle V_1\circ_{14}V_{[[2,3],4]}V_5\rangle}{s_{23}s_{234}}$$



$$A(14325) = \frac{\langle V_{[[1,4],3]}\circ_{14}V_2V_5\rangle}{s_{14}s_{134}} + \frac{\langle V_{[1,[4,3]]}\circ_{14}V_2V_5\rangle}{s_{34}s_{134}} + \frac{\langle V_{[1,4]}\circ_{14}V_{[3,2]}V_5\rangle}{s_{14}s_{23}} + \frac{\langle V_1\circ_{14}V_{[4,[3,2]]}V_5\rangle}{s_{23}s_{234}} + \frac{\langle V_1\circ_{14}V_{[[4,3],2]}V_5\rangle}{s_{34}s_{234}}$$

Figure 9: The amplitudes $A(12345)$ and $A(14325)$ parameterized with BCJ numerators according to the Möbius map (7.36) with $i,j=1,4$. The expanded numerators after applying the Möbius product (7.33) are given in (7.37). See figure 2 for the binary tree expansion of $b(1234)$.

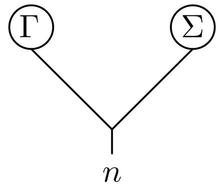
which can be explicitly verified by expanding the commutators. The correctness of the other examples can be checked similarly. In [224], a similar interpretation was used to map the product $V_{iAjB}V_CV_n$ to a multi-peripheral color factor composed from a string of structure constants. The map (7.33) was then shown to correspond to a closed formula to rewrite the multi-peripheral factors in the DDM basis of Del Duca, Dixon and Maltoni [195].

7.1.6. Local BCJ numerators from the Möbius product

Using the Möbius product (7.33) it is easy to obtain local numerators for SYM tree amplitudes satisfying the color-kinematics duality, and in fact the full tree amplitudes in BCJ form. To this end, for an n-point tree amplitude, we map the planar binary trees in the expansion of $b(P)$ in (4.125) with $|P|=n-1$ and the root identified as the n-th leg to pure spinor superspace numerators as follows [224]

$$[\Gamma,\Sigma] \longrightarrow \langle V_\Gamma \circ_{ij} V_\Sigma V_n\rangle \tag{7.36}$$

with superfields in the BCJ gauge and for arbitrary choices for i,j (usually $i,j=1,n-1$). The graphical depiction is the following:



$$\langle V_\Gamma \circ_{ij} V_\Sigma V_n\rangle$$

where the blobs Γ and Σ represent arbitrary cubic trees. For example, the expression for the amplitudes $A(1,2,3,4,5)$ and $A(1,4,3,2,5)$ are obtained from $s_{1234}b(1234)$ and $s_{1234}b(1432)$ from (4.125) using the prescription (7.36) with $i,j=1,4$, see figure 9. More explicitly, after applying the Möbius product to the numerators one gets

$$\begin{aligned} A(1,2,3,4,5) &= \frac{\langle V_{123}V_4V_5\rangle}{s_{12}s_{123}} + \frac{\langle V_{123}V_4V_5 - V_{132}V_4V_5\rangle}{s_{23}s_{123}} - \frac{\langle V_{12}V_{43}V_5\rangle}{s_{12}s_{34}} + \frac{\langle V_1V_{432}V_5\rangle}{s_{34}s_{234}} + \frac{\langle V_1V_{432}V_5 - V_1V_{423}V_5\rangle}{s_{34}s_{234}}\,, \\ A(1,4,3,2,5) &= \frac{\langle V_1V_{432}V_5 + V_{12}V_{43}V_5 + V_{13}V_{42}V_5 + V_{132}V_4V_5\rangle}{s_{14}s_{134}} + \frac{\langle V_1V_{432}V_5 + V_{12}V_{43}V_5\rangle}{s_{34}s_{134}} \end{aligned} \tag{7.37}$$

$$+\frac{\langle V_1V_{432}V_5 - V_1V_{423}V_5 - V_{123}V_4V_5 + V_{132}V_4V_5\rangle}{s_{14}s_{23}} + \frac{\langle V_1V_{432}V_5 - V_1V_{423}V_5\rangle}{s_{23}s_{234}} + \frac{\langle V_1V_{432}V_5\rangle}{s_{34}s_{234}}\,,$$

which agree with the results of [226]. For instance, the numerator of the pole $1/(s_{34}s_{134})$ in the amplitude $A(1,4,3,2,5)$ is given by $V_{[1,[4,3]]}\circ_{14} V_2V_5$, whose evaluation via (7.33) yields

$$\big(V_{143}\circ_{14} V_2 - V_{134}\circ_{14} V_2\big)V_5 = V_1V_{432}V_5 - V_{12}V_{43}V_5\,, \tag{7.38}$$

where we used (4.55) and (4.56) to rewrite $V_{[1,[4,3]]} = V_{143} - V_{134}$ followed by the examples in (7.34). Comparing with the parameterization of the five-point numerators $n_{j=1,2,\ldots,15}$ in [29]

$$\begin{aligned} A(1,2,3,4,5) &= \frac{n_1}{s_{12}s_{123}} + \frac{n_2}{s_{23}s_{234}} + \frac{n_3}{s_{34}s_{12}} + \frac{n_4}{s_{123}s_{23}} + \frac{n_5}{s_{234}s_{34}}\,, \\ A(1,4,3,2,5) &= \frac{n_6}{s_{14}s_{134}} + \frac{n_5}{s_{234}s_{34}} + \frac{n_7}{s_{23}s_{14}} + \frac{n_8}{s_{134}s_{34}} + \frac{n_2}{s_{234}s_{23}}\,, \end{aligned} \tag{7.39}$$

it is easy to verify that the BCJ triplet identity $n_3 - n_5 + n_8 = 0$ is satisfied:

$$-\langle V_{12}V_{43}V_5\rangle - \langle V_1V_{432}V_5\rangle + \langle V_1V_{432}V_5 + V_{12}V_{43}\rangle = 0\,. \tag{7.40}$$

All the other BCJ numerator identities can be similarly verified.

7.1.7. The field-theory limit of the superstring disk amplitude for arbitrary orderings

Finally, we define the Möbius product of Berends–Giele currents $M_X\circ_{ij} M_Y$ by the action on the products $V_A\circ_{ij} V_B$ arising from the expansion (4.83) of M_X and M_Y which extends to

$$E_P^{(ij)} := \sum_{XY=P} M_X\circ_{ij} M_Y\,. \tag{7.41}$$

It was argued in [224] that the field-theory limit of the superstring amplitude with arbitrary color ordering can be written as

$$\lim_{\alpha'\to 0}\mathcal{A}(P,n) = \langle E_P^{(1,n-1)}V_n\rangle =: A^{(1,n-1)}(P,n)\,, \tag{7.42}$$

such that the right-hand side can be viewed as a closed-formula yielding field-theory amplitudes whose (local) numerators satisfy the color-kinematics duality.

7.1.8. Local BCJ numerators from finite gauge transformations

In [157] a straightforward parameterization of YM tree amplitudes satisfying the color-kinematics duality was obtained. The idea is to map the Lie-polynomial numerators $\Gamma = [\Gamma_1,\Gamma_2]$ of the planar-binary-tree-expansions generated by (4.125) into kinematic numerators. This can be done using the θ=0 component of the local vector potential $A^m_{[\Gamma_1,\Gamma_2]}$ in the BCJ gauge of [84, 85]. In this gauge, the vector potential A^m_Γ is associated to a cubic-tree Lie polynomial Γ and satisfies the same Jacobi identities of the associated color factors but in kinematic space[44]. The BCJ gauge at arbitrary multiplicity was shown in [85] to be equivalent to a standard finite gauge transformation of the SYM field $\mathbb{A}^m$.

Starting from the binary-tree expansion $b(P)b(n)$, where $b(n)$=n is a single letter, the YM tree amplitude $A^{\rm YM}(P,n)$ is obtained from the map[45]

$$N(\Gamma n) = (e_\Gamma\cdot e_n)\,, \tag{7.43}$$

where e^m_Γ is the local θ=0 component (4.122) of the superfield A^m_Γ in the BCJ gauge reviewed in section 4.1.6. More precisely,

$$A^{\rm YM}(P,n) = \lim_{s_P\to 0} s_P N\big(b(P)b(n)\big)\,, \tag{7.44}$$

[44] More recently, mapping binary trees to kinematic numerators was proposed in [231, 232] exploiting a beautiful connection to free Lie algebras via the quasi-shuffle product [233].

[45] Similar maps were considered in [85].

where the expansion of the binary-tree map $b(P)$ decorates the color-kinematics dual numerators with the cubic-diagram propagators. While the earlier expressions (7.29) for the components of n-point BCJ numerators involve multiparticle polarizations of rank $\leq n{-}2$, the numerators in (7.43) involve rank-$(n{-}1)$ building blocks.

For example, the four-point amplitudes in the KK basis of color ordering following from (7.44) are given by

$$A^{\rm YM}(1234) = \left(\frac{e^m_{[[1,2],3]}}{s_{12}} + \frac{e^m_{[1,[2,3]]}}{s_{23}}\right) e^m_4 \, , \qquad (7.45)$$
$$A^{\rm YM}(1324) = \left(\frac{e^m_{[[1,3],2]}}{s_{13}} + \frac{e^m_{[1,[3,2]]}}{s_{23}}\right) e^m_4 \, ,$$

from which all BCJ kinematic numerator identities map one-to-one to the Jacobi identities of the associated Lie polynomials. Similarly, the five-point amplitudes

$$A^{\rm YM}(12345) = \left(\frac{e^m_{[[[1,2],3],4]}}{s_{12}s_{45}} + \frac{e^m_{[1,[[2,3],4]]}}{s_{23}s_{51}} + \frac{e^m_{[[1,2],[3,4]]}}{s_{12}s_{34}} + \frac{e^m_{[[1,[2,3]],4]}}{s_{45}s_{23}} + \frac{e^m_{[1,[2,[3,4]]]}}{s_{51}s_{34}}\right) e^m_5 \, , \qquad (7.46)$$
$$A^{\rm YM}(14325) = \left(\frac{e^m_{[[[1,4],3],2]}}{s_{14}s_{25}} + \frac{e^m_{[1,[[4,3],2]]}}{s_{43}s_{51}} + \frac{e^m_{[[1,4],[3,2]]}}{s_{14}s_{32}} + \frac{e^m_{[[1,[4,3]],2]}}{s_{25}s_{43}} + \frac{e^m_{[1,[4,[3,2]]]}}{s_{51}s_{32}}\right) e^m_5 \, ,$$
$$A^{\rm YM}(13425) = \left(\frac{e^m_{[[[1,3],4],2]}}{s_{13}s_{25}} + \frac{e^m_{[1,[[3,4],2]]}}{s_{43}s_{51}} + \frac{e^m_{[[1,3],[4,2]]}}{s_{13}s_{42}} + \frac{e^m_{[[1,[3,4]],2]}}{s_{25}s_{43}} + \frac{e^m_{[1,[3,[4,2]]]}}{s_{51}s_{42}}\right) e^m_5 \, ,$$
$$A^{\rm YM}(12435) = \left(\frac{e^m_{[[[1,2],4],3]}}{s_{12}s_{35}} + \frac{e^m_{[1,[[2,4],3]]}}{s_{24}s_{51}} + \frac{e^m_{[[1,2],[4,3]]}}{s_{12}s_{34}} + \frac{e^m_{[[1,[2,4]],3]}}{s_{35}s_{24}} + \frac{e^m_{[1,[2,[4,3]]]}}{s_{51}s_{34}}\right) e^m_5 \, ,$$
$$A^{\rm YM}(14235) = \left(\frac{e^m_{[[[1,4],2],3]}}{s_{14}s_{35}} + \frac{e^m_{[1,[[4,2],3]]}}{s_{24}s_{51}} + \frac{e^m_{[[1,4],[2,3]]}}{s_{14}s_{32}} + \frac{e^m_{[[1,[4,2]],3]}}{s_{35}s_{24}} + \frac{e^m_{[1,[4,[2,3]]]}}{s_{51}s_{32}}\right) e^m_5 \, ,$$
$$A^{\rm YM}(13245) = \left(\frac{e^m_{[[[1,3],2],4]}}{s_{13}s_{45}} + \frac{e^m_{[1,[[3,2],4]]}}{s_{23}s_{51}} + \frac{e^m_{[[1,3],[2,4]]}}{s_{13}s_{24}} + \frac{e^m_{[[1,[3,2]],4]}}{s_{45}s_{23}} + \frac{e^m_{[1,[3,[2,4]]]}}{s_{51}s_{24}}\right) e^m_5$$

have kinematic numerators that manifestly satisfy the color-kinematics duality. Similar expressions can be written down at arbitrary multiplicity, and their form closely resembles the form of the amplitudes in the Berends–Giele method, but now they arise from the planar binary tree expansion $b(P)$.

The above BCJ representations are equivalent to

$$A^{\rm YM}(P) = \sum_{Q \in S_{n-2}} m(P|1,Q,n)(e_{1Q}\cdot e_n) \, , \qquad (7.47)$$

where the propagators are now organized into doubly-partial amplitudes instead of $b(P)$. This representation was studied in section 5 of [201] and generalized to tree-level matrix elements for the effective operators $\alpha'\mathbb{F}^3$ and $\alpha'^2\mathbb{F}^4$ of the open bosonic string.

7.1.9. An explicit solution to BCJ relations in KLT form

The process of obtaining the field-theory limit (7.23) from the local $(n{-}2)!$-term representation of disk amplitudes (7.21) can be repeated for the non-local form (6.69) with $(n{-}3)!$ terms. From the low-energy limit (6.81), we arrive at an explicit representation of the BCJ amplitude relations in terms of $(n{-}3)!$ SYM amplitudes,

$$A(P) = -\sum_{Q,R \in S_{n-3}} m(P|1,R,n,n{-}1) S(R|Q)_1 A(1,Q,n{-}1,n) \, , \qquad (7.48)$$

of both practical and conceptual appeal.

At the practical level, (7.48) is a closed-form solution to the entirety of BCJ relations (5.55) or (5.62), i.e. for the expansion of arbitrary color-ordered amplitudes in a prescribed $(n{-}3)!$ BCJ basis. The BCJ relations by themselves do not offer any guidance on how to solve the huge equation system to rewrite the $(n-1)!$ permutations of $A(P,n)$ in terms of the $(n-3)!$ linearly independent $A(1,Q,n-1,n)$. Hence, it is beneficial to have the closed formula for the expansion coefficients in (7.48), in particular since the entries of $m(\cdot|\cdot)$ and $S(\cdot|\cdot)_1$ can be efficiently generated from the recursions (4.160) and (6.85), respectively. For example, using (7.48) to rewrite the SYM amplitude $A(24315)$ in the BCJ basis $\{A(12345), A(13245)\}$ we get

$$\begin{aligned} A(24315) &= -\big(m(24315|12354)S(23|23)_1 + m(24315|13254)S(32|23)_1\big)A(12345) \\ &\quad - \big(m(24315|12354)S(23|32)_1 + m(24315|13254)S(32|32)_1\big)A(13245) \\ &= -\frac{s_{12}}{s_{134}}A(12345) - \frac{(s_{12}+s_{23})}{s_{134}}A(13245)\,, \end{aligned} \tag{7.49}$$

where we used that $m(24315|12354) = 0$, $m(24315|13254) = 1/(s_{13}s_{134})$, as well as (4.161) for the KLT matrix.

At a conceptual level, the KLT form of (7.48) leads to the conclusion that SYM is a double copy of bi-adjoint scalars with SYM itself. Since this statement carries over to any other field or string theory subject to tree-level BCJ relations, bi-adjoint scalars can be viewed as the identity operator under taking double copies. This can of course be anticipated from the identification (6.100) of doubly-partial amplitudes as the inverse KLT kernel [35]. The realization of SYM as a double copy of bi-adjoint scalars with SYM is the $\alpha' \to 0$ limit of the double-copy formula (6.69) for disk amplitudes: when interpreting open superstrings as a double copy of Z-theory with SYM, bi-adjoint scalars are recovered from the low-energy limit of the more general Z-theory of bi-colored scalars, see section 8.6 for an α'-expansion of their equations of motion.

We conclude by mentioning a quick consistency check of (7.48): for permutations $P \to 1, A, n-1, n$ within the BCJ basis on the right-hand side, (7.48) holds trivially since $m(\cdot|\cdot)$ and $S(\cdot|\cdot)_1$ are inverse to each other by (6.100). For any other permutation P outside the BCJ basis of $A(1,Q,n-1,n)$, SYM amplitudes $A(P)$ obey the same BCJ relations in P as $m(P|B)$ at fixed B.

7.2. String-theory KLT relations and the double-copy form of gravity numerators

This section is dedicated to gravitational amplitudes in string and field theories. We review the string-theory incarnation of the KLT formula, identify closed-string analogues of the Z-integrals along with their field-theory limits and deduce the local form of the gravitational double copy with cubic-graph numerators given by perfect squares $N_i\tilde{N}_i$. This is the tree-level case [29] of the conjecture due to Bern, Carrasco and Johansson [228] that representations of gauge-theory amplitudes with manifest color-kinematics duality induce explicit loop integrands in double-copy form for a variety of gravitational theories. The BCJ double copy radically changed the computational reach for multiloop amplitudes in supergravity and drives precision calculations of gravitational-wave observables, see [30, 31] for reviews and [32] for a white paper.

7.2.1. String-theory KLT relations

The opening line for closed-string tree-level amplitudes in the pure spinor formalism is given by

$$\mathcal{M}_n^{\rm closed} = \left(-\frac{\alpha'}{2\pi}\right)^{n-3} \int\limits_{\mathbb{C}^{n-3}\setminus\{z_a=z_b\}} d^2z_2\, d^2z_3 \,\ldots\, d^2z_{n-2}\, \langle\!\langle V_1^{\rm cl}(z_1)U_2^{\rm cl}(z_2)\ldots U_{n-2}^{\rm cl}(z_{n-2})V_{n-1}^{\rm cl}(z_{n-1})V_n^{\rm cl}(z_n)\rangle\!\rangle\,, \tag{7.50}$$

where the integration of $z_2, z_3, \ldots, z_{n-2}$ over the Riemann sphere realizes the moduli-space integral over genus-zero surfaces with n marked points in the $\mathrm{SL}_2(\mathbb{C})$ frame with z_1, z_{n-1}, z_n fixed to $(0, 1, \infty)$. In comparison to the disk-amplitude prescription (3.75), the closed-string vertex operators $V_i^{\rm cl}, U_i^{\rm cl}$ are double copies of the open-string ones V_i, U_i,

$$V_i^{\rm cl} = |\lambda^\alpha A_\alpha(\theta)|^2 e^{k_i\cdot X}\,, \qquad U_i^{\rm cl} = \big|\partial\theta^\alpha A_\alpha(\theta) + A_m(\theta)\Pi^m + d_\alpha W^\alpha(\theta) + \tfrac{1}{2}N_{mn}F^{mn}(\theta)\big|^2 e^{k_i\cdot X}\,, \tag{7.51}$$

where $K(\theta)$ denote the SYM superfields without their plane-wave factor, see (3.61). Moreover, $|\lambda^\alpha A_\alpha(\theta)|^2 = \lambda^\alpha A_\alpha(\theta)\tilde{\lambda}^{\hat\beta}\tilde{A}_{\hat\beta}(\theta)$ introduces right-moving counterparts $\tilde{\lambda}^{\hat\alpha}, \tilde{\theta}^{\hat\beta}$ of the left-moving worldsheet variables $\lambda^\alpha, \theta^\beta$

whose spinor indices $\hat{\alpha},\hat{\beta},\ldots$ have same (opposite) chirality as $\alpha,\beta,\ldots$ in the case of the type IIB (type IIA) theory. The $\tilde{\theta}$-expansion of $\tilde{A}_{\hat{\beta}}(\tilde{\theta})$ and all the other superfields $\tilde{K}(\tilde{\theta})$ in $U_i^{\rm cl}$ again takes the form of (2.17) with independent gauge-multiplet polarizations $\tilde{e}_m,\tilde{\chi}^{\hat{\alpha}}$ in the place of e_m,χ^α in A_β.

The correlator $\langle\!\langle\ldots\rangle\!\rangle$ in (7.50) is adapted to the sphere rather than the disk: apart from the plane-wave factors $e^{k_i\cdot X}$ in (7.51), the OPEs for the left- and right-moving parts of $V_i^{\rm cl},U_i^{\rm cl}$ are performed separately, and the zero-mode integral (3.80) applies independently to $\lambda^\alpha,\theta^\beta$ and $\tilde{\lambda}^{\hat{\alpha}},\tilde{\theta}^{\hat{\beta}}$. Hence, the sphere correlator in (7.50) factorizes into two copies of the correlators $\mathcal{K}_n$ on the disk defined by (6.1),

$$\mathcal{M}_n^{\rm closed}=\Big(-\frac{\alpha'}{2\pi}\Big)^{n-3}\int\limits_{\mathbb{C}^{n-3}\backslash\{z_a=z_b\}} d^2z_2\,d^2z_3\ldots d^2z_{n-2}\,\langle\mathcal{K}_n\rangle\,\langle\tilde{\mathcal{K}}_n\rangle\prod_{i<j}^n|z_{ij}|^{-\alpha' s_{ij}}\,, \tag{7.52}$$

where the closed-string polarizations are obtained from the tensor products of the superfields in $\mathcal{K}_n$ and $\tilde{\mathcal{K}}_n$. The OPE singularities in $\tilde{\mathcal{K}}_n$ are the complex conjugates $\overline{z}_{ij}^{-1}$ of the z_{ij}^{-1} in $\mathcal{K}_n$.

At three points, the absence of integrated punctures immediately leads to the factorization of the string amplitudes into color-ordered open-string ones

$$\mathcal{M}_3^{\rm closed}=\mathcal{A}(1,2,3)\tilde{\mathcal{A}}(1,2,3)\,. \tag{7.53}$$

At $n\geq 4$ points, one can even decompose the closed-string Koba–Nielsen factor into products of meromorphic and antimeromorphic functions according to $|z_{ij}|^{-\alpha' s_{ij}}=(z_{ij})^{-\frac{\alpha'}{2}s_{ij}}(\overline{z}_{ij})^{-\frac{\alpha'}{2}s_{ij}}$. The integrand of (7.52) is a holomorphic square of a meromorphic but multivalued function $\langle\mathcal{K}_n\rangle\prod_{i<j}^n(z_{ij})^{-\frac{\alpha'}{2}s_{ij}}$ with branch points at the diagonals $z_a=z_b$. Hence, it requires care to extend the double-copy structure of (7.52) to the sphere integrals: the multivaluedness of $(z_{ij})^{-\frac{\alpha'}{2}s_{ij}}$ introduces monodromy phases $e^{\pm\frac{i\pi}{2}\alpha' s_{ij}}$ in relating different integration contours which also take center stage in the discussion of monodromy relations in section 7.3.

The monodromy phases in unwinding the sphere integrals (7.52) over closed-string Koba–Nielsen factors into products of disk integrals (with open-string Koba–Nielsen factors at $\alpha'\to\frac{\alpha'}{4}$) have been firstly determined by Kawai, Lewellen and Tye (KLT) in 1986 [182]. At four points, the phases conspire to a single trigonometric factor in

$$\begin{aligned}\mathcal{M}_4^{\rm closed}&=-\frac{\alpha'}{2\pi}\int_{\mathbb{C}^{n-3}\backslash\{0,1,\infty\}} d^2z_2\,\langle\mathcal{K}_4\rangle\,\langle\tilde{\mathcal{K}}_4\rangle\,|z_2|^{-\alpha' s_{12}}|1{-}z_2|^{-\alpha' s_{23}}\\&=-\frac{\alpha'}{2\pi}\sin\Big(\frac{\pi\alpha'}{2}s_{12}\Big)\int_0^1 dz_2\,z_2^{-\frac{\alpha'}{2}s_{12}}(1{-}z_2)^{-\frac{\alpha'}{2}s_{23}}\,\langle\mathcal{K}_4\rangle\\&\quad\times\int_0^{-\infty}d\overline{z}_2\,(-\overline{z}_2)^{-\frac{\alpha'}{2}s_{12}}(1{-}\overline{z}_2)^{-\frac{\alpha'}{2}s_{23}}\,\langle\tilde{\mathcal{K}}_4\rangle\\&=-\frac{2}{\pi\alpha'}\sin\Big(\frac{\pi\alpha'}{2}s_{12}\Big)\mathcal{A}(1,2,3,4;\tfrac{\alpha'}{4})\,\tilde{\mathcal{A}}(2,1,3,4;\tfrac{\alpha'}{4})\,,\end{aligned} \tag{7.54}$$

where the rescaling $\alpha'\to\frac{\alpha'}{4}$ in the open-string amplitudes on the right-hand side can be seen by comparison with the Koba–Nielsen exponents in $|z_{ij}|^{-2\alpha' s_{ij}}$ in (6.51). Note that one can employ the Gamma-function representation (6.13) of open-string amplitudes together with $\sin(\pi x)=\frac{\pi x}{\Gamma(1+x)\Gamma(1-x)}$ and the field-theory KLT relations (4.159) to factor out the supergravity amplitude

$$\mathcal{M}_4^{\rm closed}=M_4^{\rm grav}\frac{\Gamma(1{-}\frac{\alpha'}{2}s_{12})\Gamma(1{-}\frac{\alpha'}{2}s_{23})\Gamma(1{-}\frac{\alpha'}{2}s_{13})}{\Gamma(1{+}\frac{\alpha'}{2}s_{12})\Gamma(1{+}\frac{\alpha'}{2}s_{23})\Gamma(1{+}\frac{\alpha'}{2}s_{13})}\,, \tag{7.55}$$

but this is no longer possible at five points.

The analogous trigonometric phase factors in the n-point KLT formula furnish an α'-dependent generalization of the field-theory momentum kernel $S(P|Q)_i$ (4.160) involving $n{-}3$ trigonometric factors

$$\mathcal{S}_{\alpha'}(Aj|BjC)_i=\frac{2}{\pi\alpha'}\sin\Big(\frac{\pi\alpha'}{2}k_j\cdot k_{iB}\Big)\,\mathcal{S}_{\alpha'}(A|BC)_i\,,\qquad \mathcal{S}_{\alpha'}(\emptyset|\emptyset)_1=1\,, \tag{7.56}$$

for example,

$$\begin{aligned}
\mathcal{S}_{\alpha'}(2|2)_1 &= \frac{2}{\pi\alpha'}\sin\left(\frac{\pi\alpha'}{2}k_1\cdot k_2\right), \\
\mathcal{S}_{\alpha'}(23|23)_1 &= \left(\frac{2}{\pi\alpha'}\right)^2\sin\left(\frac{\pi\alpha'}{2}k_3\cdot k_{12}\right)\sin\left(\frac{\pi\alpha'}{2}k_1\cdot k_2\right), \\
\mathcal{S}_{\alpha'}(23|32)_1 &= \left(\frac{2}{\pi\alpha'}\right)^2\sin\left(\frac{\pi\alpha'}{2}k_1\cdot k_2\right)\sin\left(\frac{\pi\alpha'}{2}k_1\cdot k_3\right) = \mathcal{S}_{\alpha'}(32|23)_1\,, \\
\mathcal{S}_{\alpha'}(32|32)_1 &= \left(\frac{2}{\pi\alpha'}\right)^2\sin\left(\frac{\pi\alpha'}{2}k_2\cdot k_{13}\right)\sin\left(\frac{\pi\alpha'}{2}k_1\cdot k_3\right).
\end{aligned} \tag{7.57}$$

This generalizes the recursion (4.160) of the field-theory momentum kernel, and the normalization factors are engineered to have $\mathcal{S}_{\alpha'}(P|Q)_i = S(P|Q)_i + \mathcal{O}(\alpha'^2)$.

The n-point KLT formula for closed-string tree amplitudes then takes the compact form [182, 184, 185]

$$\mathcal{M}_n^{\rm closed} = -\sum_{P,Q\in S_{n-3}} \mathcal{A}(1,P,n,n{-}1;\tfrac{\alpha'}{4})\mathcal{S}_{\alpha'}(P|Q)_1\tilde{\mathcal{A}}(1,Q,n{-}1,n;\tfrac{\alpha'}{4}) \tag{7.58}$$

and evidently reduces to the field-theory KLT relation (4.159) as $\alpha' \to 0$. The KLT formula (7.58) with type I amplitudes on the right-hand side computes tree amplitudes of the type IIB (type IIA) superstring if the chiralities of the fermions in $\mathcal{A}(\ldots)$ and $\tilde{\mathcal{A}}(\ldots)$ are the same (opposite). Similarly (7.58) relates tree amplitudes of closed and open bosonic strings.

As will be discussed in section 7.3, the $(n{-}3)!$ permutations of $\mathcal{A}(\ldots)$ and $\tilde{\mathcal{A}}(\ldots)$ on the right-hand side of (7.58) furnish bases under the monodromy relations of color-ordered open-string amplitudes. Accordingly, the four-point KLT relations (7.54) can be written in the alternative form

$$\mathcal{M}_4^{\rm closed} = -\frac{2}{\pi\alpha'}\sin\left(\frac{\pi\alpha'}{2}s_{23}\right)\mathcal{A}(1,2,3,4;\tfrac{\alpha'}{4})\,\tilde{\mathcal{A}}(1,3,2,4;\tfrac{\alpha'}{4})\,. \tag{7.59}$$

These two equivalent forms stem from different ways of deforming integration contours in [182]. The systematic study of the analogous n-point integration contours on the sphere led to the momentum-kernel formalism in [185].

One can also manifest the symmetry $\mathcal{A}\leftrightarrow\tilde{\mathcal{A}}$ of the KLT formula by repeated use of monodromy relations, but already the four-point example

$$\mathcal{M}_4^{\rm closed} = -\frac{2}{\pi\alpha'}\frac{\sin\left(\frac{\pi\alpha'}{2}s_{12}\right)\sin\left(\frac{\pi\alpha'}{2}s_{23}\right)}{\sin\left(\frac{\pi\alpha'}{2}s_{13}\right)}\mathcal{A}(1,2,3,4;\tfrac{\alpha'}{4})\,\tilde{\mathcal{A}}(1,2,3,4;\tfrac{\alpha'}{4}) \tag{7.60}$$

shows that the locality of the KLT kernel in (7.56) is lost in this way. This motivates the choice of asymmetric bases for $\mathcal{A}$ and $\tilde{\mathcal{A}}$ in (7.58) which lead the simple and local entries (7.56) of the n-point KLT kernel.

7.2.2. Sphere integrals and their field-theory limit

The derivation of the KLT formula is independent on the polarizations accompanying the sphere integrals and the rational functions of $z_j,\overline{z}_j$ entering the correlators $\langle\mathcal{K}_n\rangle,\langle\tilde{\mathcal{K}}_n\rangle$ in (7.52). Hence, one can rewrite it at the level of Parke–Taylor integrals

$$J(P|Q) := \left(-\frac{\alpha'}{2\pi}\right)^{n-3}\int\limits_{\mathbb{C}^{n-3}}\frac{d^2z_1\;d^2z_2\;\cdots\;d^2z_n}{{\rm vol}({\rm SL}(2,\mathbb{C}))}\prod_{i<j}^n|z_{ij}|^{-\alpha' s_{ij}}\,{\rm PT}(Q)\overline{{\rm PT}(P)} \tag{7.61}$$

that furnish the closed-string counterparts of the Z-integrals (6.62): In both of $Z(P|Q)$ and $J(P|Q)$, the second entry refers to the meromorphic Parke–Taylor factor ${\rm PT}(Q)$ in the integrand. The role of first word P in turn changes in passing from the disk to the sphere – instead of a disk ordering $D(P)$, it refers to an antimeromorphic (i.e. complex conjugate) Parke–Taylor factor $\overline{\mathrm{PT}(P)}$ in the sphere integrand of (7.61) which does not arise in disk correlators.

The equivalent of the KLT formula (7.58) for the sphere integrals (7.61) takes a universal form for any pair of Parke–Taylor factors $\mathrm{PT}(Q)\overline{\mathrm{PT}(P)}$,

$$J(P|Q) = -\sum_{A,B\in S_{n-3}} Z(1,A,n,n{-}1|P)\mathcal{S}_{\alpha'}(A|B)_1 Z(1,B,n{-}1,n|Q) \tag{7.62}$$

and in fact for any other pair of rational functions in $z_j,\overline{z}_j$ of the same $\mathrm{SL}_2(\mathbb{C})$-weight. Here and below, the rescaling $\alpha'\to\frac{\alpha'}{4}$ within the disk integrals $Z(1,A,n,n{-}1|P)$, $Z(1,B,n{-}1,n|Q)$ is implicit. This rescaling rule applies whenever disk integrals are imported into closed-string computations as in (7.58) or (7.62).

The field-theory limit of (7.62) reveals another striking parallel between the $Z(P|Q)$ and $J(P|Q)$ integrals: Given that the $\alpha'\to 0$ limit (6.81) of disk integrals introduces doubly-partial amplitudes $m(P|Q)$ and therefore the inverse field-theory KLT matrix by (6.100), we conclude that [137]

$$\lim_{\alpha'\to 0} J(P|Q) = m(P|Q)\,, \tag{7.63}$$

i.e. the disk and sphere integrals $Z(P|Q)$ and $J(P|Q)$ have the same field-theory limit. As will be detailed in section 8.7, this parallel between $Z(P|Q)$ and $J(P|Q)$ even has an echo at all orders in their α'-expansions.

7.2.3. The local form of the gravitational double copy

The sphere integrals (7.61) of Parke–Taylor type and their field-theory limit (7.63) admit an elegant proof of the gravitational double copy at the level of cubic tree diagrams, cf. section 7.1. The starting point is the local representation of the disk correlator in the form of (7.21)

$$\langle\mathcal{K}_n\rangle = \frac{dz_1\,dz_{n-1}\,dz_n}{\mathrm{vol}(\mathrm{SL}(2,\mathbb{R}))}\sum_{P\in S_{n-2}} N_{1|P|n-1}\,\mathrm{PT}(1,P,n{-}1) \bmod \nabla_{z_k}\,, \tag{7.64}$$

where the superfield representation $\sim\langle V_X V_Y V_Z\rangle$ of the master numerators $N_{1|P|n-1}$ (with P a permutation of $2,3,\ldots,n{-}2,n$) can be found in (7.24). The rescaling $\alpha'\to\frac{\alpha'}{4}$ in a closed-string context also applies to the expression (6.74) for Koba–Nielsen derivatives ∇_{z_k}. The DDM-type formula (7.64) was already at the heart of deriving BCJ numerators from disk amplitudes in section 7.1.3. Upon insertion into the closed-string amplitude representation (7.52) and identifying the J-integrals (7.61), it leads to the $(n{-}2)!^2$-term expression

$$\mathcal{M}_n^{\mathrm{closed}} = \sum_{P,Q\in S_{n-2}} \tilde{N}_{1|P|n-1} J(1,P,n{-}1|1,Q,n{-}1) N_{1|Q|n-1}\,, \tag{7.65}$$

where P,Q are again permutations of $2,3,\ldots,n,n{-}2$ (our choice of SL_2 frame led to a swap $n\leftrightarrow n{-}1$ relative to the DDM-type formulae in section 7.1.2). With the field-theory limit (7.63) of the sphere integrals and relabeling of $n\leftrightarrow n{-}1$, one readily obtains gravity amplitudes in the form

$$M_n^{\mathrm{grav}} = \sum_{P,Q\in S_{n-2}} \tilde{N}_{1|P|n} m(1,P,n|1,Q,n) N_{1|Q|n} \tag{7.66}$$

analogous to the color-dressed tree amplitudes of bi-adjoint scalars and SYM in (7.17) and (7.19). By Jacobi identities of both color factors c_i and kinematic numerators N_i, (7.17) and (7.19) were explained to be equivalent to the cubic-diagram expansions (7.18) and (7.1). Since these rewritings solely depend on the properties of the universal building block $m(1,P,n|1,Q,n)$, the same equivalence must hold for (7.66) and

$$M_n^{\mathrm{grav}} = \sum_{i\in\Gamma_n} \frac{N_i\tilde{N}_i}{D_i}\,, \tag{7.67}$$

where both types of kinematic numerators N_i and $\tilde{N}_i$ obey Jacobi relations. Hence, we have derived the prescription of [29, 228, 227] that kinematic Jacobi relations among the numerators are sufficient to obtain gravity amplitudes from SYM via

$$M_n^{\rm grav} = M_n^{\rm gauge}\Big|_{c_i\to\tilde{N}_i} = \sum_{i\in\Gamma_n}\frac{N_i c_i}{D_i}\Big|_{c_i\to\tilde{N}_i}\,, \tag{7.68}$$

i.e. by replacing color factors by another copy $\tilde{N}_i$ of kinematic numerators. In fact, the Jacobi identities of the color factors c_i imply that the cubic-diagram expansion (7.1) of gauge-theory amplitudes may still accommodate Jacobi-violating numerators N_i, see the discussion below (7.10). Accordingly, the color-kinematics dual representation (7.67) of gravity amplitudes is still valid if only one of the sets of numerator $\{N_i\}$ or $\{\tilde{N}_i\}$ obeys Jacobi identities.

Note that (7.66) in combination with (7.20) yields another manifestly local formulation of the double copy

$$M_n^{\rm grav} = \sum_{P\in S_{n-2}}\tilde{N}_{1|P|n}A(1,P,n)\,, \tag{7.69}$$

which is obtained from the DDM form (7.16) through the same replacement $c_{1|P|n}\to\tilde{N}_{1|P|n}$ as in (7.68).

7.2.4. Another derivation of the field-theory KLT relation

As exemplified in section 7.1.9, it is rewarding to also insert the non-local form (6.73) of the disk correlator into the field-theory limit of string amplitudes. In the closed-string case, (7.52) together with a relabeling of $n\leftrightarrow n{-}1$ in $\tilde{\mathcal{K}}_n$ leads to the amplitude representation

$$\mathcal{M}_n^{\rm closed} = \sum_{P,Q,A,B\in S_{n-3}}\tilde{A}(1,P,n,n{-}1)S(P|Q)_1 J(1,Q,n{-}1,n|1,A,n,n{-}1)S(A|B)_1 A(1,B,n{-}1,n)\,, \tag{7.70}$$

where both left- and right-movers contribute one copy of the field-theory KLT kernel. By the field-theory limit (7.63) and the inverse relation (6.100) between $m(1,Q,n{-}1,n|1,A,n,n{-}1)$ and $S(A|B)_1$, we have

$$S(P|B)_1 = \sum_{Q,A\in S_{n-3}} S(P|Q)_1\lim_{\alpha'\to0}J(1,Q,n{-}1,n|1,A,n,n{-}1)S(A|B)_1\,, \tag{7.71}$$

and arrive at the KLT formula (4.159) for gravity amplitudes by taking $\alpha'\to0$ in (7.70). In conclusion, the combinatorial structure of the field-theory KLT formula can be understood from two perspectives – either from the monodromy phases in manipulating integration cycles on the sphere and disk or from the expansion of disk correlators in a basis of Parke–Taylor factors.

7.3. Monodromy relations

Color-ordered open-string amplitudes $\mathcal{A}(P)$ associated with different orderings P of the vertex operators on the disk boundary obey monodromy relations [134, 135]. Similar to the KLT relations (7.58) for closed-string amplitudes, they solely rely on analytic properties of the disk worldsheet and are therefore universal to the bosonic theory and type I superstrings. Monodromy relations can be equivalently formulated at the level of the $Z(P|Q)$-integrals (6.62): while section 6.4.3 featured relations between different "integrands Q" at fixed "integration domain P", monodromy relations concern different choices of the domain P at fixed integrand Q. To begin with, the procedure of fixing $\mathrm{SL}_2(\mathbb{R})$ frames in section 6.4.1 leads to the following cyclicity and reflection properties,

$$Z(p_1p_2\ldots p_n|Q) = Z(p_2p_3\ldots p_np_1|Q) = (-1)^n Z(p_n\ldots p_2p_1|Q)\,, \tag{7.72}$$

yielding a naive upper bound of $\frac{1}{2}(n{-}1)!$ independent disk orderings. However, the actual basis dimensions for color-ordered disk amplitudes identified by monodromy relations are considerably smaller with only $(n{-}3)!$ choices of P at fixed Q [134, 135]. The proof relies on the following simple analytic property of the disk integrand and thereby extends to integrands of suitable $\mathrm{SL}_2(\mathbb{R})$-weight beyond Parke–Taylor factors: the only non-meromorphic dependence on the integration variables in (6.62) occurs through the Koba–Nielsen factor $\prod_{1\leq i<j}^n |z_{ij}|^{-2\alpha' s_{ij}}$. The latter can be related to the meromorphic but multivalued function $\prod_{1\leq i<j}^n (z_{ij})^{-2\alpha' s_{ij}}$ by monodromy phases $e^{\pm 2\pi i\alpha' s_{ij}}$ which differ from one ordering P to another. The same type of monodromy phases gives rise to the trigonometric factor in the four-point KLT relation (7.54). By applying Cauchy's theorem as detailed in [134, 135], one obtains,

$$0 = \sum_{j=1}^{n-1} \exp\big[2\pi i\alpha'(k_{p_1}\cdot k_{p_2p_3\ldots p_j})\big] Z(p_2p_3\ldots p_jp_1p_{j+1}\ldots p_n|Q)\,, \tag{7.73}$$

and the associated relation among color-ordered amplitudes of open superstrings,

$$0 = \sum_{j=1}^{n-1} \exp\big[2\pi i\alpha'(k_{p_1}\cdot k_{p_2p_3\ldots p_j})\big] \mathcal{A}(p_2p_3\ldots p_jp_1p_{j+1}\ldots p_n)\,, \tag{7.74}$$

which take an identical form for open bosonic strings. Since different choices of branches yield identical relations with opposite phases as compared to (7.73) and (7.74), one can take sums and differences of both options and obtain [134, 135]

$$\begin{aligned}0 &= \mathcal{A}(p_1p_2p_3\ldots p_n) + \sum_{j=2}^{n-1} \cos\big[2\pi\alpha'(k_{p_1}\cdot k_{p_2p_3\ldots p_j})\big]\mathcal{A}(p_2p_3\ldots p_jp_1p_{j+1}\ldots p_n)\,,\\ 0 &= \sum_{j=2}^{n-1} \sin\big[2\pi\alpha'(k_{p_1}\cdot k_{p_2p_3\ldots p_j})\big]\mathcal{A}(p_2p_3\ldots p_jp_1p_{j+1}\ldots p_n)\,.\end{aligned} \tag{7.75}$$

For real kinematics, these are simply the real and imaginary parts of (7.74). At leading order in α', the trigonometric coefficients reduce to $\cos(\alpha' x)\to 1$ and $\sin(\alpha' x)\to\alpha' x$, respectively. As a result, one obtains (special instances of) KK relations (5.40) and the fundamental BCJ relations (5.62) as the low-energy limit of the first and second line of (7.75), respectively [134, 135]. Moreover, the fact that $\lim_{\alpha'\to 0} Z(P|Q)$ obeys KK and BCJ relations in P at fixed Q is consistent with the relations of $m(P|Q)$ obtained in the field-theory limit.

In the canonical ordering $p_j = j$ at four points, (7.75) reduce to

$$\begin{aligned}0 &= \mathcal{A}(1,2,3,4) + \cos(2\pi\alpha' k_1\cdot k_2)\mathcal{A}(2,1,3,4) + \cos(2\pi\alpha' k_1\cdot k_{23})\mathcal{A}(2,3,1,4)\,,\\ 0 &= \sin(2\pi\alpha' k_1\cdot k_2)\mathcal{A}(2,1,3,4) + \sin(2\pi\alpha' k_1\cdot k_{23})\mathcal{A}(2,3,1,4)\,.\end{aligned} \tag{7.76}$$

The second relation together with $k_1\cdot k_{23} = -s_{23}$ and $\mathcal{A}(2,3,1,4) = \mathcal{A}(1,3,2,4)$ establishes the equivalence between the two forms (7.54) and (7.59) of the four-point KLT relations (taking the usual rescaling $\alpha'\to\frac{\alpha'}{4}$ into account). More generally, one may view monodromy relations as a consistency condition for permutation invariance of the n-point KLT formula (7.58): the $(n{-}3)!\times(n{-}3)!$ pairs of $\mathcal{A}(1,P,n,n{-}1)\tilde{\mathcal{A}}(1,Q,n{-}1,n)$ on the right-hand side need to generate bilinears $\mathcal{A}(X)\tilde{\mathcal{A}}(Y)$ with arbitrary orderings X,Y through linear combinations.

The monodromy relations of individual disk integrals

$$\begin{aligned}0 &= Z(p_1p_2p_3\ldots p_n|Q) + \sum_{j=2}^{n-1} \cos\big[2\pi\alpha'(k_{p_1}\cdot k_{p_2p_3\ldots p_j})\big] Z(p_2p_3\ldots p_jp_1p_{j+1}\ldots p_n|Q)\,,\\ 0 &= \sum_{j=2}^{n-1} \sin\big[2\pi\alpha'(k_{p_1}\cdot k_{p_2p_3\ldots p_j})\big] Z(p_2p_3\ldots p_jp_1p_{j+1}\ldots p_n|Q)\end{aligned} \tag{7.77}$$

equivalent to (7.75) underpin our viewpoint on the disk integrals (6.62) as the doubly-partial amplitudes of Z-theory. By (6.76) and (6.77), Z-theory amplitudes $Z(P|Q)$ satisfy the color-kinematics duality in the integrand orderings Q at fixed P to all orders in α'. The converse relations (7.77) among the integration domain orderings P at fixed Q, on the other hand, exhibit additional trigonometric α'-dependence. These trigonometric functions imprint the monodromy properties of the disk worldsheets on the S-matrix of Z-theory upon dressing with the relevant Chan–Paton factors $\sum_P Z(P,n|Q)\mathrm{Tr}(t^P t^n)$.

7.3.1. The $(n-2)!$ form of color-dressed open-string amplitudes

Since the first relation of (7.75) deforms the KK relations of gauge-theory amplitudes by $\cos(2\pi\alpha' k_P \cdot k_Q)$, one may wonder about the string-theory uplift of the DDM decomposition (7.16) in gauge theory. By repeated use of monodromy relations, one can express the color-dressed open-superstring amplitude (3.78) in terms of the $(n-2)!$ color orderings $\mathcal{A}(1,P,n)$ with $P = p_2 p_3 \ldots p_{n-1}$ [234],

$$\mathcal{M}_n = \sum_{P\in S_{n-2}} \mathrm{Tr}([[\ldots[[t^1,t^{p_2}]_{\alpha'},t^{p_3}]_{\alpha'},\ldots,t^{p_{n-2}}]_{\alpha'},t^{p_{n-1}}]_{\alpha'}t^n)\mathcal{A}(1,\rho(2,3,\ldots,n-1),n)\,. \tag{7.78}$$

Their coefficients generalize the color factors in the field-theory DDM formula (7.16)

$$c_{1|23\ldots n-1|n} = \mathrm{Tr}([[\ldots[[t^1,t^2],t^3],\ldots,t^{n-2}],t^{n-1}]t^n) \tag{7.79}$$

to involve an α'-dependent bracket instead of the conventional commutator

$$[t^p,t^q]_{\alpha'} = e^{i\pi\alpha' k_p\cdot k_q}t^p t^q - e^{-i\pi\alpha' k_p\cdot k_q}t^q t^p\,. \tag{7.80}$$

For matrix products in the entries of $[\cdot,\cdot]_{\alpha'}$ or nested commutators, the exponentials are understood to involve multiparticle momenta, e.g. $[t^1t^2,t^3]_{\alpha'} = e^{i\pi\alpha' k_{12}\cdot k_3}t^1t^2t^3 - e^{-i\pi\alpha' k_{12}\cdot k_3}t^3t^1t^2$. At four points, for instance

$$\begin{aligned}
\mathcal{M}_4 &= \mathrm{Tr}([[t^1,t^2]_{\alpha'},t^3]_{\alpha'}t^4)\mathcal{A}(1,2,3,4) + (2\leftrightarrow 3)\\
&= \big[e^{i\pi\alpha'(k_1\cdot k_2+k_{12}\cdot k_3)}\mathrm{Tr}(t^1t^2t^3t^4) - e^{i\pi\alpha'(-k_1\cdot k_2+k_{12}\cdot k_3)}\mathrm{Tr}(t^2t^1t^3t^4)\\
&\quad - e^{i\pi\alpha'(k_1\cdot k_2-k_{12}\cdot k_3)}\mathrm{Tr}(t^3t^1t^2t^4) + e^{i\pi\alpha'(-k_1\cdot k_2-k_{12}\cdot k_3)}\mathrm{Tr}(t^3t^2t^1t^4)\big]\mathcal{A}(1,2,3,4) + (2\leftrightarrow 3)\\
&= \mathrm{Tr}(t^1t^2t^3t^4{+}t^3t^2t^1t^4)\mathcal{A}(1,2,3,4) - \mathrm{Tr}(t^2t^1t^3t^4)\big[e^{-2\pi i\alpha' s_{12}}\mathcal{A}(1,2,3,4){+}e^{2\pi i\alpha' s_{13}}\mathcal{A}(1,3,2,4)\big] + (2\leftrightarrow 3)
\end{aligned} \tag{7.81}$$

can be checked to reproduce the conventional form (3.78) of the color-dressed amplitude by means of $k_{12}\cdot k_3 = -s_{12}$ and the monodromy relation $e^{-2\pi i\alpha' s_{12}}\mathcal{A}(1,2,3,4) + e^{2\pi i\alpha' s_{13}}\mathcal{A}(1,3,2,4) = -\mathcal{A}(2,1,3,4)$.

By isolating the coefficient of a given $\mathrm{Tr}(t^{p_1}t^{p_2}\ldots t^{p_n})$ on the right-hand side of (7.78), this DDM-type decomposition of open-string tree amplitude generates the expansion of arbitrary $\mathcal{A}(P)$ in a prescribed $(n-2)!$-element set of disk orderings. However, the expansion coefficients are not unique since the $\mathcal{A}(1,\ldots,n)$ on the right hand side of (7.78) are still related by monodromy relations. As we will see in section 7.4, the specialization of (7.78) to abelian Chan–Paton generators $t^i \to \mathbb{1}$ has valuable applications to Born–Infeld theory and its double-copy structure.

7.3.2. The $(n-3)!$ form of color-dressed open-string amplitudes

The next step after identifying the string-theory uplift (7.78) of the $(n-2)!$-term DDM decomposition is to reduce color-ordered string amplitudes to an $(n-3)!$-element basis under monodromy relations. As can be anticipated from the reduction of gauge-theory amplitudes into a BCJ basis via (7.48), an elegant solution of the monodromy relations is offered by the string-theory KLT kernel and its inverse.

In view of the interpretation (6.100) of the inverse field-theory KLT kernel as doubly-partial amplitudes of bi-adjoint scalars, the inverse of the string-theory KLT kernel $\mathcal{S}_{\alpha'}$ has been firstly studied in [235]. Its entries w.r.t. the $(n-3)!\times(n-3)!$ basis of $\mathcal{S}_{\alpha'}$ in (7.56) are given by

$$m^{-1}_{\alpha'}(1,R,n-1,n|1,Q,n,n-1) = -\mathcal{S}_{\alpha'}(R|Q)_1\,, \tag{7.82}$$

and one can infer more general entries $m_{\alpha'}(A|B)$ by inverting the kernel in different representations of the KLT relations (7.58) with other $(n-3)!$ bases $\mathcal{B}_1, \mathcal{B}_2$ of $\mathcal{A}(\ldots)$, $\tilde{\mathcal{A}}(\ldots)$,

$$\mathcal{M}_n^{\text{closed}} = \sum_{P,Q\in\mathcal{B}_1,\mathcal{B}_2} \mathcal{A}(P) m_{\alpha'}^{-1}(P|Q)\tilde{\mathcal{A}}(Q)\,. \tag{7.83}$$

At four and five points, for example, we have,

$$\begin{aligned}
m_{\alpha'}(1,2,3,4|1,2,4,3) &= -\frac{\pi\alpha'}{2\sin\left(\frac{\pi\alpha'}{2}s_{12}\right)}\,, \qquad\qquad (7.84)\\
m_{\alpha'}(1,2,3,4|1,2,3,4) &= \frac{\pi\alpha'}{2}\left\{\cot\left(\frac{\pi\alpha'}{2}s_{12}\right)+\cot\left(\frac{\pi\alpha'}{2}s_{23}\right)\right\},\\
m_{\alpha'}(1,5,3,2,4|1,2,3,5,4) &= \Big(\frac{\pi\alpha'}{2}\Big)^2\frac{1}{\sin(\frac{\pi\alpha'}{2}s_{14})}\left\{\cot\left(\frac{\pi\alpha'}{2}s_{23}\right)+\cot\left(\frac{\pi\alpha'}{2}s_{35}\right)\right\},\\
m_{\alpha'}(1,5,3,2,4|1,3,2,5,4) &= -\Big(\frac{\pi\alpha'}{2}\Big)^2\frac{1}{\sin(\frac{\pi\alpha'}{2}s_{14})}\frac{1}{\sin(\frac{\pi\alpha'}{2}s_{23})}\,, \qquad\qquad (7.85)
\end{aligned}$$

and permutations. The entries of higher-multiplicity $m_{\alpha'}$ are efficiently generated by the `Mathematica` notebook in the ancillary file of [235].

At any multiplicity, $m_{\alpha'}(A|B)$ enjoys cyclicity and monodromy relations of open-string amplitudes in both slots A, B (while holding the other one fixed) after performing the usual conversion $\alpha' \to 4\alpha'$ between closed- and open-string settings. This leads to the elegant formula to expand disk amplitudes or Z-integrals with an arbitrary integration cycle P in a prescribed basis w.r.t. monodromy relations [235],

$$\begin{aligned}
\mathcal{A}_n(P) &= -\sum_{Q,R\in S_{n-3}} m_{4\alpha'}(P|1,R,n,n{-}1)\mathcal{S}_{4\alpha'}(R|Q)_1\mathcal{A}(1,Q,n{-}1,n)\,, \qquad (7.86)\\
Z(P|C) &= -\sum_{Q,R\in S_{n-3}} m_{4\alpha'}(P|1,R,n,n{-}1)\mathcal{S}_{4\alpha'}(R|Q)_1 Z(1,Q,n{-}1,n|C)\,,
\end{aligned}$$

which readily follows from the logic of section 7.1.9. The first line of (7.86) furnishes the string-theory uplift of the BCJ reduction of SYM amplitudes in (7.48), and the second line is valid for arbitrary Parke–Taylor orderings C in the Z-integrals (see (6.101) for the converse formula for attaining a prescribed basis of Parke–Taylor factors at fixed disk ordering). For example, using the expansions of $m_{\alpha'}$ in (7.84) and the KLT matrix in (7.57) we get

$$\begin{aligned}
\mathcal{A}(1,5,3,2,4) &= \frac{\sin(2\pi\alpha' s_{12})\sin\big(2\pi\alpha'(s_{13}{+}s_{23})\big)}{\sin(2\pi\alpha' s_{14})\sin(2\pi\alpha' s_{23})}\mathcal{A}(1,2,3,4,5)+\frac{\sin(2\pi\alpha' s_{12})\sin(2\pi\alpha' s_{13})}{\sin(2\pi\alpha' s_{14})\sin(2\pi\alpha' s_{23})}\mathcal{A}(1,3,2,4,5)\\
&\quad-\Big\{\cot(2\pi\alpha' s_{23})+\cot(2\pi\alpha' s_{35})\Big\}\frac{\sin(2\pi\alpha' s_{12})\sin(2\pi\alpha' s_{13})}{\sin(2\pi\alpha' s_{14})}\mathcal{A}(1,2,3,4,5) \qquad (7.87)\\
&\quad-\Big\{\cot(2\pi\alpha' s_{23})+\cot(2\pi\alpha' s_{35})\Big\}\frac{\sin(2\pi\alpha' s_{13})\sin\big(2\pi\alpha'(s_{12}{+}s_{23})\big)}{\sin(2\pi\alpha' s_{14})}\mathcal{A}(1,3,2,4,5)\,,
\end{aligned}$$

consistent with the five-point examples in [134, 135].

In fact, Mizera identified the entries of $m_{\alpha'}$ with intersection numbers of twisted cycles [188] and thereby opened up a fascinating connection between string perturbation theory and twisted deRham theory. In this framework, the KLT relations are a consequence of *twisted period relations* [236], and their representation (7.83) follows from elementary linear algebra in twisted homologies and cohomologies [188]. Similarly, the field-theory KLT relations can be understood from intersection numbers of twisted cocycles [189].

Since $m_{\alpha'}$ can be algorithmically computed from intersection numbers, there is no circular logic in solving monodromy relations via (7.86): the entries of $m_{\alpha'}$ beyond the $(n-3)! \times (n-3)!$ basis in (7.82) do not necessitate any prior knowledge of the solutions to the monodromy relations.

7.4. Double copies beyond gravity from string amplitudes

The color-kinematics duality and double copy apply to a much wider classes of theories beyond gauge theories and (super-)gravity [30, 31, 32]. In this section, we will review the input of superstring tree amplitudes on the double-copy structure of Born–Infeld theory and its supersymmetrizations. The reasoning will be based on the KLT-form (6.73) of the disk correlator which implies that all tree amplitudes of Born–Infeld are double copies involving a BCJ basis of SYM tree amplitudes. The other double-copy component of Born–Infeld amplitudes turns out to be a non-linear sigma model (NLSM) of Goldstone bosons even though the latter are not part of the naive string spectrum (but can be engineered to arise as massless excitations in the setup of [237]).

7.4.1. Born–Infeld and NLSM

The low-energy limit of abelian open-superstring tree-level interactions gives rise to the Born–Infeld theory [238], also see [239] for a review and [240, 241] for its supersymmetrizations to so-called Dirac–Born–Infeld–Volkov–Akulov theories. Tree-level amplitudes $\mathcal{M}_n^{\rm BI}$ of Born–Infeld were identified as a field-theory double copy of SYM with scalar amplitudes in the NLSM of Goldstone bosons [242] as can be stated through the KLT formula

$$\mathcal{M}_n^{\rm BI} = - \sum_{Q,R\in S_{n-3}} A_{\rm NLSM}(1,R,n,n{-}1) S(R|Q)_1 A(1,Q,n{-}1,n)\,. \tag{7.88}$$

In contrast to the gravitational KLT formula (4.159), the polarizations of the colorless spin-one multiplets in $\mathcal{M}_n^{\rm BI}$ stem entirely from those in the color-ordered SYM amplitudes $A(R)$.

The study of the NLSM [243, 244, 245, 246, 247] and its tree-level amplitudes [248, 249, 250, 251] has a long history. Within the modern amplitudes program, the interest in the NLSM was fueled by the observation that its tree amplitudes obey KK and BCJ relations [252] and qualify to enter field-theory double copies. Just like the Born–Infeld amplitudes, color-ordered tree amplitudes of the NLSM vanish for odd multiplicity, and their simplest non-zero instances are

$$\begin{aligned} A_{\rm NLSM}(1,2,3,4) &= s_{12}+s_{23}\,, \\ A_{\rm NLSM}(1,2,3,4,5,6) &= s_{12} - \frac{(s_{12}{+}s_{23})(s_{45}{+}s_{56})}{2s_{123}} + {\rm cyc}(1,2,3,4,5,6)\,. \end{aligned} \tag{7.89}$$

In order to compute Born–Infeld amplitudes from the low-energy limit of abelian open-superstring states, we specialize the color-dressed disk amplitude $\mathcal{M}_n$ in (3.78) to $t^i \to \mathbb{1}$ and insert the KLT formula (6.69) for color-ordered disk amplitudes:

$$\begin{aligned} \mathcal{M}_n^{\rm BI} &= - \lim_{\alpha'\to 0} \frac{1}{(2\pi\alpha')^{n-2}} \mathcal{M}_n \big|_{t^i\to\mathbb{1}} = - \lim_{\alpha'\to 0} \frac{1}{(2\pi\alpha')^{n-2}} \sum_{Q\in S_{n-1}} \mathcal{A}_n(Q,n) \\ &= - \sum_{Q,R\in S_{n-3}} \Big(\lim_{\alpha'\to 0} \frac{1}{(2\pi\alpha')^{n-2}} Z_\times(1,R,n,n{-}1) \Big) S(R|Q)_1 A(1,Q,n{-}1,n)\,, \end{aligned} \tag{7.90}$$

where we introduce the following shorthand for symmetrized disk integrals or abelian Z-theory amplitudes

$$Z_\times(P) = \sum_{Q\in S_{n-1}} Z(Q,n|P)\,. \tag{7.91}$$

The inverse $n{-}2$ factors of $2\pi\alpha'$ in (7.90) compensate for the leading order $\sim \alpha'^{n-2}$ in the low-energy expansion of the abelian open-string amplitudes $\mathcal{M}_n\big|_{t^i\to\mathbb{1}}$ that will be exposed in the discussion below. Based on the reflection property (7.72), the symmetrization in (7.91) annihilates $Z_\times(P)$ of odd multiplicity.

Since the disk integrals $Z_\times$ solely depend on Mandelstam invariants, the KLT formula (7.90) implies that Born–Infeld is a double-copy involving SYM. For consistency with the alternative KLT formula (7.88) which identifies the NLSM as the second double-copy component [242], the coefficients of the linearly independent $A(1,Q,n-1,n)$ have to agree. Hence, the conclusion from equating (7.88) with (7.90) is that NLSM amplitudes arise from the low-energy limit of symmetrized disk integrals [187],

$$A_{\rm NLSM}(P) = \lim_{\alpha'\to 0} \frac{1}{(2\pi\alpha')^{n-2}} Z_\times(P)\,. \tag{7.92}$$

This adds support to the interpretation of disk integrals as tree-level amplitudes in a bi-colored scalar Z-theory: when abelianizing the gauge-group generators $t^i \to \mathbb{1}$ dressing the disk ordering P of $Z(P|Q)$, the low-energy limit reproduces the tree amplitudes of the NLSM, a well-known scalar field theory. The appearance of NLSM amplitudes in the low-energy limit of abelian Z-theory is here deduced from the double copy (7.88) of Born–Infeld in [242] and does not rely on Goldstone bosons in the superstring spectrum. On the other hand, toroidal compactifications of ten-dimensional superstrings along with worldsheet boundary condensates indeed give rise to NLSM Goldstone bosons among the massless excitations [237].

7.4.2. BCJ numerators of the NLSM from disk integrals

The definition (7.91) of symmetrized disk integrals does not manifest the leading term in its α'-expansion, so it may appear surprising that the limit (7.92) does not diverge. In order to expose the leading order α'^{n-2} of the $Z_\times(Q)$, we shall employ a variant of the DDM-type decomposition (7.78) of color-ordered open-string amplitudes. In reading this decomposition at the level of the disk integrals and specializing to abelian gauge-group generators, we obtain

$$Z_\times(Q) = \sum_{P\in S_{n-2}} {\rm Tr}([[\ldots[[\mathbb{1}^1,\mathbb{1}^{p_2}]_{\alpha'},\mathbb{1}^{p_3}]_{\alpha'},\ldots,\mathbb{1}^{p_{n-2}}]_{\alpha'},\mathbb{1}^{p_{n-1}}]_{\alpha'}\mathbb{1}^n) Z(1,P,n|Q)\,. \tag{7.93}$$

In slight abuse of notation, we have indicated through the superscripts of $\mathbb{1}^j$ that these unit matrices arose from the abelianization of t^j. In this way, the information about the momentum dependence in the α'-deformed bracket (7.80) is preserved and we can evaluate

$${\rm Tr}([[\ldots[[\mathbb{1}^1,\mathbb{1}^{p_2}]_{\alpha'},\mathbb{1}^{p_3}]_{\alpha'},\ldots,\mathbb{1}^{p_{n-2}}]_{\alpha'},\mathbb{1}^{p_{n-1}}]_{\alpha'}\mathbb{1}^n) = \prod_{j=2}^{n-1}(e^{i\pi\alpha' k_{1p_2\ldots p_{j-1}}\cdot k_{p_j}} - e^{-i\pi\alpha' k_{1p_2\ldots p_{j-1}}\cdot k_{p_j}})\,. \tag{7.94}$$

Upon converting the exponentials to sine functions, this implies

$$Z_\times(Q) = (2i)^{n-2} \sum_{P\in S_{n-2}} Z(1,P,n|Q) \prod_{j=2}^{n-1} \sin\left(\pi\alpha' k_{1p_2\ldots p_{j-1}}\cdot k_{p_j}\right)\,, \tag{7.95}$$

which leads to vanishing $Z_\times(Q)$ of odd multiplicity and the following examples at even n:

$$\begin{aligned} Z_\times(q_1,q_2,q_3,q_4) &= 4\sin^2\big(\pi\alpha' k_1\cdot k_2\big) Z(1,2,3,4|q_1,q_2,q_3,q_4) + 4\sin^2\big(\pi\alpha' k_1\cdot k_3\big) Z(1,3,2,4|q_1,q_2,q_3,q_4)\,,\\ Z_\times(q_1,q_2,\ldots,q_6) &= 16\sum_{P\in S_4} \sin\big(\pi\alpha' k_1\cdot k_{p_2}\big)\sin\big(\pi\alpha' k_{1p_2}\cdot k_{p_3}\big)\\ &\qquad\times \sin\big(\pi\alpha' k_{1p_2p_3}\cdot k_{p_4}\big)\sin\big(\pi\alpha' k_{1p_2p_3p_4}\cdot k_{p_5}\big) Z(1,P,6|q_1,q_2,\ldots,q_6)\,. \end{aligned} \tag{7.96}$$

Given that the low-energy limit of $Z(1,P,n|Q)$ yields doubly-partial amplitudes m at order α'^0 and each sine function introduces leading low-energy order α'^1, one can easily identify the low-energy limit of (7.95) to be

$$Z_\times(Q) = (2i\pi\alpha')^{n-2}\bigg\{ \sum_{P\in S_{n-2}} m(1,P,n|Q) \prod_{j=2}^{n-1}(k_{1p_2\ldots p_{j-1}}\cdot k_{p_j}) + \mathcal{O}(\alpha')\bigg\}\,. \tag{7.97}$$

Hence, the representation (7.92) of NLSM amplitudes in terms of low-energy limits of symmetrized disk integrals is non-singular and simplifies to

$$
\begin{aligned}
A_{\rm NLSM}(P) &= i^{n-2} \sum_{Q\in S_{n-2}} m(P|1,Q,n) \prod_{j=2}^{n-1} (k_{1q_2\ldots q_{j-1}} \cdot k_{q_j}) \\
&= \sum_{Q\in S_{n-2}} m(P|1,Q,n) N^{\rm NLSM}_{1|Q|n} \, .
\end{aligned} \tag{7.98}
$$

In passing to the second line, we have manifested the formal similarity with the DDM form (7.20) of gauge-theory amplitudes in a BCJ form. Given that the coefficients in the $(n-2)!$-term sum (7.20) over doubly-partial amplitudes $m(P|1,Q,n)$ are BCJ master numerators of SYM (cf. section 7.1.3), the analogous coefficients $N^{\rm NLSM}_{1|Q|n}$ in the second line of (7.98) are bound to be local BCJ numerators of the NLSM,

$$
N^{\rm NLSM}_{1|Q|n} = i^{n-2} \prod_{j=2}^{n-1} (k_{1q_2\ldots q_{j-1}} \cdot k_{q_j}) = i^{n-2} S(Q|Q)_1 \, . \tag{7.99}
$$

In the second step, we have identified the BCJ master numerators of the NLSM as diagonal entries of the field-theory KLT matrix as initially conjectured in [187] and then derived as outlined above in [253]. A Lagrangian for the NLSM with manifest color-kinematics duality was presented in [254] which reproduces the numerators in (7.99) from Feynman rules. Earlier explicit BCJ numerators for the NLSM in terms of the KLT kernel can be found in [186].

7.4.3. Coupling NLSM to bi-adjoint scalars

The behavior of NLSM and Born–Infeld amplitudes under soft limits $k_j \to 0$ in the external momenta singles out preferred ways of coupling Goldstone bosons to bi-adjoint scalars and supersymmetric Born–Infeld theories to SYM [255]. These extended theories to be referred to as NLSM+ϕ^3 and BI+SYM are related by KLT formulae

$$
A^{\rm BI+SYM}_n(P) = -\sum_{Q,R\in S_{n-3}} A_{{\rm NLSM}+\phi^3}(P|1,R,n,n{-}1) S(R|Q)_1 A(1,Q,n{-}1,n) \, , \tag{7.100}
$$

which can also be studied from disk amplitudes in the pure spinor formalism:

The BI+SYM theory of [255] results from the low-energy limit of open-superstring amplitudes where a subset of the Chan–Paton generators is abelianized. This amounts to keeping some of the t^j non-abelian in (7.90) and isolating an appropriate order in α' as the low-energy limit. The non-abelian t^j give rise to a color-decomposition w.r.t. $|P| < n$ legs, and $A^{\rm BI+SYM}_n(P)$ refers to the coefficient of ${\rm Tr}(t^{p_1}t^{p_2}\ldots t^{p_{|P|}})$.

Similar to the doubly-partial amplitudes $m(A|B)$ of bi-adjoint scalars defined by (6.80), the amplitudes $A_{{\rm NLSM}+\phi^3}(P|1,R,n,n{-}1)$ on the right-hand side of (7.100) are the coefficients of two types of traces – one over generators $\tilde t^a$ shared by the bi-adjoint scalars $\Phi = \Phi_{j|a} t^j \otimes \tilde t^a$ and the Goldstone bosons of the NLSM as well as one over the t^j exclusive to the $|P|$ external bi-adjoint scalars. The simplest examples that do not coincide with pure NLSM or ϕ^3 amplitudes occur at five points, where for instance [255]

$$
\begin{aligned}
A_{{\rm NLSM}+\phi^3}(3,4,5|1,2,3,4,5) &= 1 - \frac{s_{51}{+}s_{12}}{s_{34}} - \frac{s_{12}{+}s_{23}}{s_{45}} \, , \\
A_{{\rm NLSM}+\phi^3}(2,3,5|1,2,3,4,5) &= 1 - \frac{s_{45}{+}s_{51}}{s_{23}} \, .
\end{aligned} \tag{7.101}
$$

Note that the coupling of Goldstone bosons to bi-adjoint scalars can be accommodated in the NLSM Lagrangian of [254] with manifest color-kinematics duality.

In computing the SYM+BI amplitudes (7.100) from the low-energy limit of the open superstring, the NLSM+ϕ^3 amplitudes on the right-hand side arise via partially symmetrized disk integrals $Z_P(Q)$ [253]. As detailed in the reference, the latter can be defined by starting from Chan–Paton-dressed Z-theory in the DDM-type form (7.78)

$$\mathcal{Z}_n(Q) = \sum_{P \in S_{n-2}} \mathrm{Tr}([[\ldots[[t^1, t^{p_2}]_{\alpha'}, t^{p_3}]_{\alpha'}, \ldots, t^{p_{n-2}}]_{\alpha'}, t^{p_{n-1}}]_{\alpha'} t^n) Z(1, P, n|Q) \tag{7.102}$$

and setting a subset of the generators t^j to be unit matrices. We then obtain partially symmetrized disk integrals such as the ($|Q| = 5$)-point example

$$\begin{aligned} Z_{345}(Q) &= \mathcal{Z}_5(Q)\,\big|_{\mathrm{Tr}(t^3 t^4 t^5)} \\ &= 4\sin(\pi\alpha' k_1 \cdot k_2)\sin(\pi\alpha' k_{12} \cdot k_4)\cos(\pi\alpha' k_{124} \cdot k_3) Z(1,2,4,3,5|Q) \\ &\quad + 4\sin(\pi\alpha' k_1 \cdot k_4)\sin(\pi\alpha' k_{14} \cdot k_2)\cos(\pi\alpha' k_{124} \cdot k_3) Z(1,4,2,3,5|Q) \\ &\quad + 4\sin(\pi\alpha' k_1 \cdot k_4)\sin(\pi\alpha' k_{134} \cdot k_2)\cos(\pi\alpha' k_{14} \cdot k_3) Z(1,4,3,2,5|Q) \\ &\quad - 4\sin(\pi\alpha' k_1 \cdot k_2)\sin(\pi\alpha' k_{12} \cdot k_3)\cos(\pi\alpha' k_{123} \cdot k_4) Z(1,2,3,4,5|Q) \\ &\quad - 4\sin(\pi\alpha' k_1 \cdot k_3)\sin(\pi\alpha' k_{13} \cdot k_2)\cos(\pi\alpha' k_{123} \cdot k_4) Z(1,3,2,4,5|Q) \\ &\quad - 4\sin(\pi\alpha' k_1 \cdot k_3)\sin(\pi\alpha' k_{134} \cdot k_2)\cos(\pi\alpha' k_{13} \cdot k_4) Z(1,3,4,2,5|Q) \end{aligned} \tag{7.103}$$

from the color-decomposition of $\mathcal{Z}_n(Q)$ w.r.t. the non-abelian t^j. In the low-energy limit where $Z(P|Q) \rightarrow m(P|Q)$ as well as $\sin(\pi\alpha' k_P \cdot k_Q) \rightarrow \pi\alpha' k_P \cdot k_Q$ and $\cos(\pi\alpha' k_P \cdot k_Q) \rightarrow 1$, we recover $(2\pi\alpha')^2$ times the first line of (7.101) from (7.103). A variety of further examples and the systematics for general numbers of abelianized and non-abelian t^j can be found in [253]. Among other things, the results in the reference give rise to local BCJ numerators for the NLSM+ϕ^3 theory from the monodromy properties of disk integrals along the lines of section 7.4.2.

By gradually converting some of the Chan–Paton dressings of the Z-amplitudes (7.102) to become abelian $t^j \rightarrow \mathbb{1}$, the low-energy limits interpolate between pure ϕ^3-amplitudes and pure NLSM amplitudes. In the "semi-abelian" case, Z-theory can be viewed as an ultraviolet completion of the NLSM+ϕ^3 theory in [255].

7.5. Heterotic strings and Einstein–Yang–Mills

The $(n-3)!$ form of the disk correlators in (6.73) also has crucial implications for massless tree amplitudes of heterotic string theories since their supersymmetric chiral halves can be described through the left-moving modes of the pure spinor formalism. The massless sector of the heterotic string incorporates both gauge multiplets and the half-maximal supergravity multiplet in ten dimensions. In contrast to the type I theory, heterotic strings already feature gauge-gravity couplings at tree level due to worldsheets of sphere topology. Hence, one can study Einstein–Yang–Mills amplitudes from the correlators of the heterotic string and the field-theory limit of the sphere integrals in section 7.2.2.

7.5.1. Basics of heterotic-string amplitudes

The opening line for tree amplitudes of the heterotic string [256]

$$\mathcal{M}_n^{\mathrm{het}} = \left(-\frac{\alpha'}{2\pi}\right)^{n-3} \int\limits_{\mathbb{C}^{n-3}\setminus\{z_a=z_b\}} d^2 z_2\, d^2 z_3\, \ldots\, d^2 z_{n-2}\, \langle\!\langle V_1^{\mathrm{het}}(z_1) U_2^{\mathrm{het}}(z_2) \ldots U_{n-2}^{\mathrm{het}}(z_{n-2}) V_{n-1}^{\mathrm{het}}(z_{n-1}) V_n^{\mathrm{het}}(z_n) \rangle\!\rangle \tag{7.104}$$

is almost identical to that of type II superstrings in (7.50) up to the choice of vertex operators for the gauge and gravity multiplet: both the integrated and the unintegrated variant involve a chiral half from the bosonic string

$$V_i^{\mathrm{het}} = \lambda^\alpha A_\alpha^i(\theta) e^{k_i \cdot X} \overline{c} \cdot \begin{cases} \overline{\mathcal{J}}^{a_i} & : \text{ gauge multiplets}, \\ \sqrt{\frac{2}{\alpha'}}\, \tilde{\epsilon}_i^m \overline{\partial} X_m & : \text{ gravity multiplets}, \end{cases} \tag{7.105}$$

$$U_i^{\rm het} = \big(\partial\theta^\alpha A^i_\alpha(\theta) + \Pi_p A_i^p(\theta) + d_\alpha W_i^\alpha(\theta) + \tfrac{1}{2}N_{pq}F_i^{pq}(\theta)\big)e^{k_i\cdot X} \cdot \begin{cases} \overline{\mathcal{J}}^{a_i} & : \ \text{gauge multiplets}\,, \\ \sqrt{\frac{2}{\alpha'}}\,\tilde\epsilon_i^m \overline\partial X_m & : \ \text{gravity multiplets}\,, \end{cases}$$

where $\overline{\mathcal{J}}^{a_i}$ are Kac–Moody currents of antiholomorphic conformal weight $h=1$, the $\overline{c}$ ghost known from bosonic strings has conformal weight $h=-1$, and the polarization vectors of the gravity multiplets are transversal, $\tilde\epsilon_i\cdot k_i = 0$. The tree-level correlators are determined by $\langle\!\langle \overline{c}(z_1)\overline{c}(z_2)\overline{c}(z_3)\rangle\!\rangle = \overline{z}_{12}\overline{z}_{13}\overline{z}_{23}$ and the OPEs

$$\overline{\mathcal{J}}^a(z)\overline{\mathcal{J}}^b(w) \sim \frac{\delta^{ab}}{(\overline{z}-\overline{w})^2} + \frac{f^{abc}\overline{\mathcal{J}}^c(w)}{\overline{z}-\overline{w}} \tag{7.106}$$

as well as

$$\overline\partial X^m(z)e^{k\cdot X}(w) \sim -\frac{\alpha' k^m}{2(\overline{z}-\overline{w})}e^{k\cdot X}(w)\,, \qquad \overline\partial X^m(z)\overline\partial X^n(w) \sim -\frac{\alpha'\delta^{mn}}{2(\overline{z}-\overline{w})^2}\,. \tag{7.107}$$

Similar to the organization of type II amplitudes in (7.52), the correlator in (7.104) is guaranteed to comprise the Koba–Nielsen factor on the sphere and one copy of the disk correlator $\langle\mathcal{K}_n\rangle$ in (6.73) from the supersymmetric chiral half,

$$\mathcal{M}_n^{\rm het} = \Big(-\frac{\alpha'}{2\pi}\Big)^{n-3} \int\limits_{\mathbb{C}^{n-3}\backslash\{z_a=z_b\}} d^2z_2\, d^2z_3 \ldots d^2z_{n-2}\, \tilde{\mathcal{K}}_n^{\rm bos}\,\langle\mathcal{K}_n\rangle \prod_{i<j}^n |z_{ij}|^{-\alpha' s_{ij}}\,. \tag{7.108}$$

The bosonic chiral half in turn contributes the rational function $\tilde{\mathcal{K}}_n^{\rm bos}$ of the $\overline{z}_i$ that depends on the color degrees of freedom a_i of the external gauge multiplets as well as the half-polarizations $\tilde\epsilon_i^m$ and momenta of the external gravity multiplets. The non-vanishing three-point examples $\tilde{\mathcal{K}}_n^{\rm bos}(P|Q)$ of the bosonic correlator in (7.108) with external gauge and gravity multiplets in the sets P and Q are

$$\begin{aligned} &\tilde{\mathcal{K}}_3^{\rm bos}(1,2,3|\emptyset) = f^{a_1a_2a_3}\,, \qquad \tilde{\mathcal{K}}_3^{\rm bos}(1,2|3) = -\sqrt{\frac{\alpha'}{2}}\delta^{a_1a_2}(\tilde\epsilon_3\cdot k_1)\,, \\ &\tilde{\mathcal{K}}_3^{\rm bos}(\emptyset|1,2,3) = \sqrt{\frac{\alpha'}{2}}\Big\{\big[(\tilde\epsilon_1\cdot\tilde\epsilon_2)(\tilde\epsilon_3\cdot k_1) + {\rm cyc}(1,2,3)\big] - \frac{\alpha'}{2}(\tilde\epsilon_1\cdot k_2)(\tilde\epsilon_2\cdot k_3)(\tilde\epsilon_3\cdot k_1)\Big\}\,. \end{aligned} \tag{7.109}$$

As one can see from the appearance of the color factors in these examples, the heterotic-string amplitudes in (7.108) are automatically color dressed. In fact, the OPEs (7.106) of the Kac–Moody currents also introduce products of traces with a maximum of $\lfloor\frac{n}{2}\rfloor$ trace factors at n points as expected from coupling between colored gauge multiplets and uncolored supergravity multiplets. One can still isolate color-ordered single-trace amplitudes by picking up the antiholomorphic Parke–Taylor factors (6.57) in [257]

$$\langle\!\langle \overline{\mathcal{J}}^{a_1}(z_1)\overline{\mathcal{J}}^{a_2}(z_2)\ldots\overline{\mathcal{J}}^{a_n}(z_n)\rangle\!\rangle\,\big|_{{\rm Tr}(t^{a_1}t^{a_2}\ldots t^{a_n})} = -\overline{{\rm PT}(1,2,\ldots,n)}\,. \tag{7.110}$$

Accordingly, multi-trace amplitudes associated with ${\rm Tr}(t^{P_1}){\rm Tr}(t^{P_2})\ldots{\rm Tr}(t^{P_k})$ and $t^{p_1p_2\ldots p_{|P|}} = t^{p_1}t^{p_2}\ldots t^{p_{|P|}}$ are determined by isolating the product $(-1)^k\,{\rm PT}(P_1)\,{\rm PT}(P_2)\ldots{\rm PT}(P_k)$ from the current correlator.

7.5.2. Heterotic double copy and Einstein–Yang–Mills

The expansion (6.73) of the supersymmetric correlator $\langle\mathcal{K}_n\rangle$ in terms of SYM trees can be readily applied to the heterotic-string amplitude (7.108): in this way, all color-dressed tree amplitudes involving arbitrary combinations of gauge and gravity multiplets conspire to a field-theory double copy with one SYM constituent,

$$\begin{aligned} \mathcal{M}_n^{\rm het} &= -\sum_{Q,R\in S_{n-3}} \mathcal{B}(1,R,n,n{-}1)S(R|Q)_1 A(1,Q,n{-}1,n)\,, \\ \mathcal{B}(1,R,n,n{-}1) &= -\Big(-\frac{\alpha'}{2\pi}\Big)^{n-3}\int\limits_{\mathbb{C}^{n-3}\backslash\{z_a=z_b\}} d^2z_2\,d^2z_3\ldots d^2z_{n-2}\,\tilde{\mathcal{K}}_n^{\rm bos}\,\mathcal{Z}_{1R}\prod_{i<j}^n|z_{ij}|^{-\alpha' s_{ij}}\,. \end{aligned} \tag{7.111}$$

Given that the rational functions $\mathcal{Z}_{1R}$ on the right-hand side correspond to SL$(2,\mathbb{R})$-fixed Parke–Taylor factors $-\,\mathrm{PT}(1,R,n,n{-}1)$, see (6.59), one can uplift $\mathcal{B}(1,R,n,n{-}1)$ to a cyclic object $\mathcal{B}(P)$ by integrating over $-\,\mathrm{PT}(P)$ in the place of $\mathcal{Z}_{1R}$.

In order to understand the significance of the α'-dependent building block $\mathcal{B}(P)$ in the double copy (7.111), it is instructive to compare with the Einstein–Yang–Mills amplitudes obtained from the low-energy limit of heterotic strings: Einstein–Yang–Mills theories are double copies of SYM with the so-called YM+ϕ^3 theory [258]. Similar to the NLSM+ϕ^3 theory in section 7.4.3, YM+ϕ^3 is characterized by a minimal coupling of bi-adjoint scalars to pure (i.e. non-supersymmetric) Yang–Mills theory such that the BCJ relations are preserved. More precisely, in a color-decomposition of YM+ϕ^3 tree amplitudes w.r.t. the generators $\tilde{t}^j$ common to the gauge bosons and the bi-adjoint scalars,

$$M_n^{\mathrm{YM}+\phi^3} = \sum_{P\in S_{n-1}} \mathrm{Tr}(\tilde{t}^{1P}) A_{\mathrm{YM}+\phi^3}(1,P)\,, \tag{7.112}$$

the color-ordered amplitudes $A_{\mathrm{YM}+\phi^3}$ obey KK and BCJ relations. Accordingly, they qualify to enter the following KLT formula that encodes the double copy of Einstein–Yang–Mills:

$$M_n^{\mathrm{EYM}} = -\sum_{Q,R\in S_{n-3}} A_{\mathrm{YM}+\phi^3}(1,R,n,n{-}1) S(R|Q)_1 A(1,Q,n{-}1,n)\,. \tag{7.113}$$

Similar to the bosonic correlators $\tilde{\mathcal{K}}_n^{\mathrm{bos}}$ and thereby the string-theory building blocks $\mathcal{B}$ in (7.111), the $A_{\mathrm{YM}+\phi^3}$ still depend on the color-factors t^j exclusive to the bi-adjoint scalars because only the $\tilde{t}^j$ are stripped in (7.112). Since the KLT formula (7.113) is obtained from the low-energy limit of (7.111) with all α'-dependence carried by $\mathcal{B}(P)$, we can identify color-ordered YM+ϕ^3 amplitudes as its low-energy limit

$$\mathcal{B}(P) = A_{\mathrm{YM}+\phi^3}(P)\,\big|_{g_{\mathrm{YM}}\to\sqrt{\frac{\alpha'}{2}}} \big(1+\mathcal{O}(\alpha')\big)\,. \tag{7.114}$$

As exemplified by (7.109), the bosonic correlators $\tilde{\mathcal{K}}_n^{\mathrm{bos}}$ and hence the heterotic-string amplitudes (7.111) may already carry integer powers of $\sqrt{\frac{\alpha'}{2}}$ in their low-energy limits. These prefactors are interpreted as realizing the gravitational coupling κ of the heterotic string which, in the double copy of [258], translates into the gauge coupling g_{YM} of the YM+ϕ^3 theory.[46]

7.5.3. Heterotic strings as a field-theory double copy

By the bosonic origin of $\mathcal{B}(P)$ in (7.111), its α'-dependence can be further streamlined by expanding the bosonic correlator $\tilde{\mathcal{K}}_n^{\mathrm{bos}}$ in a Parke–Taylor basis. Even though the computation of $\tilde{\mathcal{K}}_n^{\mathrm{bos}}$ is straightforward from the OPEs (7.106) and (7.107), the Parke–Taylor decomposition relies on a way more intricate cascade of integration by parts than encountered in section 6.3 for supersymmetric correlators, see for instance [210, 259, 260, 261]. By the arguments in section 4.2 of [211], the coefficients $A_{(DF)^2+\mathrm{YM}+\phi^3}$ in an $(n{-}3)!$-term reduction (discarding total Koba–Nielsen derivatives $\nabla_{\overline{z}_k} f = \partial_{\overline{z}_k} f - \frac{\alpha'}{2} f \sum_{j=1, j\neq k}^n \frac{s_{kj}}{\overline{z}_{kj}}$ on the sphere)

$$\tilde{\mathcal{K}}_n^{\mathrm{bos}} = -\frac{d\overline{z}_1\, d\overline{z}_{n-1}\, d\overline{z}_n}{\mathrm{vol}(\mathrm{SL}(2,\mathbb{R}))} \sum_{Q,R\in S_{n-3}} \overline{\mathrm{PT}(1,R,n,n{-}1)} S(R|Q)_1 A_{(DF)^2+\mathrm{YM}+\phi^3}(1,Q,n{-}1,n) \bmod \nabla_{\overline{z}_k} \tag{7.115}$$

are given by field-theory amplitudes in a massive gauge theory dubbed $(DF)^2+\mathrm{YM}+\phi^3$ [262], see the discussion below (6.73) for the $d\overline{z}_i$ in the prefactor. Just like for YM+ϕ^3 theory, the massless states of $(DF)^2+\mathrm{YM}+\phi^3$ are bi-adjoint scalars and gauge bosons. The massive states in the $(DF)^2+\mathrm{YM}+\phi^3$ theory are tachyons $m^2=-\frac{4}{\alpha'}$ as expected for the open bosonic string, so the $A_{(DF)^2+\mathrm{YM}+\phi^3}$ are still

[46] In any other section of this review, we have stripped off the uniform prefactors g_{YM}^{n-2} from n-point tree-level amplitudes of SYM. In the YM+ϕ^3 theory, on the other hand, generic tree-level amplitudes mix different powers of g_{YM} according to the trace structure of the color factors: only the gluon vertices and the minimal coupling of two scalars ϕ to gluons carry powers of g_{YM} whereas the coefficient of the ϕ^3 interaction is taken to be independent on g_{YM}.

rational functions of α' [211]. Similar to (7.114), the gauge coupling of the $(DF)^2 + \mathrm{YM} + \phi^3$ theory is understood to be converted to the gravitational one in the double copy (7.115), i.e. $g_{\mathrm{YM}} \rightarrow \kappa = \sqrt{\frac{\alpha'}{2}}$.

The three-point amplitudes of the $(DF)^2+\mathrm{YM}+\phi^3$ theory (with subscripts ϕ and g for external scalars and gluons) reproduce the simplest bosonic correlators in (7.109),

$$A_{(DF)^2+\mathrm{YM}+\phi^3}(1_\phi, 2_\phi, 3_\phi) = f^{a_1 a_2 a_3}\,, \qquad A_{(DF)^2+\mathrm{YM}+\phi^3}(1_\phi, 2_\phi, 3_g) = -\sqrt{\frac{\alpha'}{2}}\delta^{a_1 a_2}(\tilde\epsilon_3 \cdot k_1)\,, \tag{7.116}$$
$$A_{(DF)^2+\mathrm{YM}+\phi^3}(1_g, 2_g, 3_g) = \sqrt{\frac{\alpha'}{2}}\left\{\big[(\tilde\epsilon_1 \cdot \tilde\epsilon_2)(\tilde\epsilon_3 \cdot k_1) + \mathrm{cyc}(1,2,3)\big] - \frac{\alpha'}{2}(\tilde\epsilon_1 \cdot k_2)(\tilde\epsilon_2 \cdot k_3)(\tilde\epsilon_3 \cdot k_1)\right\},$$

where the α'-correction in the last line can be traced back to the $\mathrm{Tr}(F^3)$ vertex in the $(DF)^2 + \mathrm{YM} + \phi^3$ Lagrangian [262]. External scalars ϕ in legs $1,3$ and gauge multiplets g in legs $2,4$ in turn give rise to

$$A_{(DF)^2+\mathrm{YM}+\phi^3}(1_\phi, 2_g, 3_\phi, 4_g) = \delta^{a_1 a_3}\frac{\alpha'}{2}\left\{\frac{(\tilde\epsilon_2 \cdot k_1)(\tilde\epsilon_4 \cdot k_3)}{s_{12}} + \frac{(\tilde\epsilon_2 \cdot k_3)(\tilde\epsilon_4 \cdot k_1)}{s_{23}} + (\tilde\epsilon_2 \cdot \tilde\epsilon_4) + \frac{\alpha'\, \tilde f_2^{mn}\tilde f_4^{mn}}{2+\alpha' s_{13}}\right\}. \tag{7.117}$$

By the double copy with SYM in (7.111), the external states ϕ and g in $(DF)^2 + \mathrm{YM} + \phi^3$ amplitudes translate into gauge multiplets g and gravity multiplets h in heterotic-string amplitudes.

The decomposition (7.115) is a useful way to disentangle the α'-dependence of (7.111) into the sphere integrals $J(P|R)$ in (7.61) with all the poles of massive-state exchange and the $A_{(DF)^2+\mathrm{YM}+\phi^3}$ with only massless and tachyonic poles,

$$\mathcal{B}(P) = -\sum_{Q,R \in S_{n-3}} J(P|1,R,n,n{-}1)S(R|Q)_1 A_{(DF)^2+\mathrm{YM}+\phi^3}(1,Q,n{-}1,n)\,. \tag{7.118}$$

The sphere integrals $J(P|Q)$ have zeros at $s_{ij\ldots k} = -\frac{2}{\alpha'}$ that prevent the tachyon poles of $A_{(DF)^2+\mathrm{YM}+\phi^3}$ from entering the heterotic-string amplitude. At four points, this can be anticipated from the Gamma functions in the denominator of (7.55). By combining (7.118) with (7.111), heterotic-string amplitudes can be brought into the form

$$\mathcal{M}_n^{\mathrm{het}} = \sum_{P,Q,A,B \in S_{n-3}} A_{(DF)^2+\mathrm{YM}+\phi^3}(1,P,n,n{-}1)S(P|Q)_1 J(1,Q,n{-}1,n|1,A,n,n{-}1)S(A|B)_1 A(1,B,n{-}1,n)\,, \tag{7.119}$$

which is directly analogous to the representation (7.70) of type II amplitudes, with $A_{(DF)^2+\mathrm{YM}+\phi^3}$ in the place of a second copy $\tilde A$ of SYM. External gauge and gravity multiplets in $\mathcal{M}_n^{\mathrm{het}}$ are represented by external scalars and gauge bosons in $A_{(DF)^2+\mathrm{YM}+\phi^3}$, respectively, with the conversion $g_{\mathrm{YM}} \rightarrow \sqrt{\frac{\alpha'}{2}}$ of the gauge coupling as in (7.114). Given that the matrix product $\sum_{Q,A \in S_{n-3}} S(P|Q)_1 J(1,Q,n{-}1,n|1,A,n,n{-}1)S(A|B)_1$ reduces to $S(P|B)_1$ in the field-theory limit, we arrive at the following refinement of (7.114)

$$A_{(DF)^2+\mathrm{YM}+\phi^3}(P) = A_{\mathrm{YM}+\phi^3}(P)\big(1 + \mathcal{O}(\alpha')\big)\,. \tag{7.120}$$

In the four-point example (7.117), for instance, the last term $\frac{\alpha'}{2}\tilde f_2^{mn}\tilde f_4^{mn}$ is subleading in α', and the resulting YM+ϕ^3 amplitude entering the Einstein–Yang–Mills double copy (7.113) is

$$A_{\mathrm{YM}+\phi^3}(1_\phi, 2_g, 3_\phi, 4_g) = \delta^{a_1 a_3} g_{\mathrm{YM}}^2\left\{\frac{(\tilde\epsilon_2 \cdot k_1)(\tilde\epsilon_4 \cdot k_3)}{s_{12}} + \frac{(\tilde\epsilon_2 \cdot k_3)(\tilde\epsilon_4 \cdot k_1)}{s_{23}} + (\tilde\epsilon_2 \cdot \tilde\epsilon_4)\right\}, \tag{7.121}$$

where g_{YM}^2 translates into the prefactor $\frac{\alpha'}{2}$ of the corresponding $(DF)^2 + \mathrm{YM} + \phi^3$ amplitude.

7.5.4. Einstein–Yang–Mills amplitude relations from string theories

As a key implication of the Einstein–Yang–Mills double copy (7.113), any tree amplitude of Einstein–Yang–Mills (regardless on the number of external gauge & gravity multiplets or traces in the t^i) can be expanded in terms of SYM trees. On top of the manifestly gauge invariant KLT form (7.113), one can explicitly realize the double copy with manifest locality,

$$M_n^{\rm EYM} = \sum_{P\in S_{n-2}} \tilde N^{{\rm YM}+\phi^3}_{1|P|n} A(1,P,n)\,, \tag{7.122}$$

with BCJ master numerators $\tilde N^{{\rm YM}+\phi^3}_{1|P|n}$ of YM+ϕ^3. This DDM form of Einstein–Yang–Mills amplitudes is analogous to the representation (7.69) of gravitational amplitudes. Field-theoretic computations of the master numerators $\tilde N^{{\rm YM}+\phi^3}_{1|P|n}$ from gauge invariance and color-kinematics duality and a discussion of the resulting Einstein–Yang–Mills amplitude relations (7.122) can be found in [263]. We shall here review the worldsheet approach to derive manifestly local Einstein–Yang–Mills amplitudes relations that amount to the DDM-type decomposition

$$\tilde{\mathcal K}_n^{\rm bos} = \frac{d\overline z_1\, d\overline z_{n-1}\, d\overline z_n}{{\rm vol(SL}(2,\mathbb R))} \sum_{P\in S_{n-2}} \tilde N^{(DF)^2+{\rm YM}+\phi^3}_{1|P|n} \overline{{\rm PT}(1,P,n)} \bmod \nabla_{\overline z_k} \tag{7.123}$$

analogous to (7.64). The YM+ϕ^3 master numerators in (7.122) can then be simply read off from the leading α'-order of the $(DF)^2+{\rm YM}+\phi^3$ numerators or Parke–Taylor coefficients in (7.123):

- For single-trace amplitudes $\mathcal A^{\rm het}(1,2,\ldots,n;p)$ with one external graviton p as well as n external gluons and associated trace ${\rm Tr}(t^1t^2\ldots t^n)$, the bosonic correlator due to (7.107) and (7.110) is given by

$$\begin{aligned}\tilde{\mathcal K}_n^{\rm bos} &\sim \overline{{\rm PT}(1,2,\ldots,n)} \sum_{j=1}^n \frac{\tilde\epsilon_p\cdot k_j}{\overline z_{jp}} = \overline{{\rm PT}(1,2,\ldots,n)} \sum_{j=1}^{n-1} (\tilde\epsilon_p\cdot k_{12\ldots j}) \frac{\overline z_{j,j+1}}{\overline z_{j,p}\overline z_{p,j+1}}\\ &= \sum_{j=1}^{n-1} (\tilde\epsilon_p\cdot k_{12\ldots j}) \overline{{\rm PT}(1,2,\ldots,j,p,j{+}1,\ldots,n{-}1,n)}\,,\end{aligned} \tag{7.124}$$

  where we used $\overline{{\rm PT}(1,2,\ldots,j,p,j{+}1,\ldots,n)} = \frac{\overline z_{j,j+1}}{\overline z_{j,p}\overline z_{p,j+1}} \overline{{\rm PT}(1,2,\ldots,n)}$ in passing to the last line. By matching with (7.123), one can read off master numerators

$$\tilde N^{(DF)^2+{\rm YM}+\phi^3}_{1|23\ldots jp(j+1)\ldots(n-1)|n} \to \tilde\epsilon_p\cdot k_{12\ldots j} \quad\Rightarrow\quad \tilde N^{{\rm YM}+\phi^3}_{1|23\ldots jp(j+1)\ldots(n-1)|n} \to \tilde\epsilon_p\cdot k_{12\ldots j} \tag{7.125}$$

  which result in the following amplitude relation (7.122) (with gravity multiplet p and single-trace ordering ${\rm Tr}(t^1t^2\ldots t^n)$)

$$A^{\rm EYM}(1,2,\ldots,n;p) = \sum_{j=1}^{n-1} (\tilde\epsilon_p\cdot k_{12\ldots j}) A(1,2,\ldots,j,p,j{+}1,\ldots,n{-}1,n)\,. \tag{7.126}$$

  This relation has been firstly derived from disk amplitudes of type I superstrings with one closed-string insertion [264] and generalized to single-trace amplitudes with multiple external gravitons using CHY methods[47] [266] and heterotic strings [259]. For single-trace amplitudes with an arbitrary number of gravity multiplets, a decomposition formula (7.122) in terms of intersection numbers of twisted cocycles can be found in [267]. Note that generic $(DF)^2+{\rm YM}+\phi^3$ numerators are rational functions in α', so their α'-independent instances in (7.125) are rather atypical.

- For n-gluon double-trace amplitudes associated with ${\rm Tr}(t^1t^P){\rm Tr}(t^Qt^n)$ and no external gravitons, the bosonic correlator determined by the current algebra is

$$\tilde{\mathcal K}_n^{\rm bos} \sim \overline{{\rm PT}(1,P)}\,\overline{{\rm PT}(Q,n)} = \frac{\alpha'}{2+\alpha' s_{1P}} \overline{{\rm PT}(1,\{P,Q\},n)} \bmod \nabla_{\overline z_k} \tag{7.127}$$

[47] The prescription for Einstein–Yang–Mills amplitudes in the CHY formalism has been given in [265, 242].

$$= -\frac{\alpha'(-1)^{|P|}}{2+\alpha' s_{1P}} \sum_{i=1}^{|P|} \sum_{j=1}^{|Q|} (-1)^{i-j} s_{ij} \sum_{\substack{A \in p_1 p_2 \cdots p_{i-1} \\ \shuffle p_{|P|} \cdots p_{i+1}}} \sum_{\substack{B \in q_{j+1} \cdots q_{|Q|} \\ \shuffle q_{j-1} \cdots q_2 q_1}} \overline{\mathrm{PT}(1, A, p_i, q_j, B, n)} \bmod \nabla_{\overline{z}_k} \, ,$$

where the second line is known from [259] and follows from expanding out the S-bracket via (4.144). The master numerators (7.123) can be read off from the Parke–Taylor coefficients, and we now have a non-trivial α'-dependence of the $\tilde{N}^{(DF)^2+\mathrm{YM}+\phi^3}_{1|P|n}$ through the geometric series

$$\frac{\alpha'}{2+\alpha' s_{1P}} = \frac{\alpha'}{2} \sum_{n=0}^{\infty} \left(-\frac{\alpha'}{2} s_{1P} \right)^n . \tag{7.128}$$

The master numerators of YM+ϕ^3 are obtained from the leading terms $\frac{\alpha'}{2+\alpha' s_{1P}} \to \frac{\alpha'}{2}$ and result in the following expansion of the Einstein–Yang–Mills double-trace amplitude $A^{\mathrm{EYM}}(1, P|Q, n)$ along with $\mathrm{Tr}(t^1 t^P)\mathrm{Tr}(t^Q t^n)$:

$$A^{\mathrm{EYM}}(1, P|Q, n) = \tfrac{\alpha'}{2} A(1, \{P, Q\}, n) \, , \tag{7.129}$$

for instance

$$\begin{aligned} A^{\mathrm{EYM}}(1,2|3,4) &= \tfrac{\alpha'}{2} s_{23} A(1,2,3,4) \, , \\ A^{\mathrm{EYM}}(1,2,3|4,5) &= \tfrac{\alpha'}{2} \big[s_{34} A(1,2,3,4,5) - s_{24} A(1,3,2,4,5) \big] \, , \\ A^{\mathrm{EYM}}(1,2,3,4|5,6) &= \tfrac{\alpha'}{2} \big[s_{45} A(1,2,3,4,5,6) - s_{35} A(1,2,4,3,5,6) + (2 \leftrightarrow 4) \big] \, , \\ A^{\mathrm{EYM}}(1,2,3|4,5,6) &= \tfrac{\alpha'}{2} \big[s_{34} A(1,2,3,4,5,6) - s_{24} A(1,3,2,4,5,6) - (4 \leftrightarrow 5) \big] \, . \end{aligned} \tag{7.130}$$

The prefactor is identified with the gravitational coupling in Einstein–Yang–Mills theory according to $\kappa^2 = \frac{\alpha'}{2}$ and signals two gravitational vertices as expected: each diagram contributing to double-trace amplitudes with external gauge multiplets involves one gravitational propagator ending on vertices with one factor of κ each.

The double-trace relations (7.129) have been derived and extended to one external gravity multiplet via CHY methods [266] and heterotic strings [259]. Their generalizations to arbitrary number of traces can be found in [268].

As illustrated by the tachyon pole $(2+\alpha' s_{1P})^{-1}$ in (7.127), the numerators $\tilde{N}^{(DF)^2+\mathrm{YM}+\phi^3}_{1|P|n}$ still feature propagators of the massive states in the $(DF)^2 + \mathrm{YM} + \phi^3$ theory. Still, they are free of massless poles and therefore yield a local low-energy limit $\tilde{N}^{\mathrm{YM}+\phi^3}_{1|P|n}$ for the Einstein–Yang–Mills amplitudes relations (7.122).

7.5.5. Reducing heterotic-string amplitudes to the single-trace sector

In fact, the Einstein–Yang–Mills amplitude relations (7.122) uplift to exact-in-α' relations for heterotic-string amplitudes in passing to the numerators of the $(DF)^2 + \mathrm{YM} + \phi^3$ theory

$$\mathcal{M}^{\mathrm{het}}_n = \sum_{P \in S_{n-2}} \tilde{N}^{(DF)^2+\mathrm{YM}+\phi^3}_{1|P|n} \mathcal{A}^{\mathrm{het}}(1, P, n) \, , \tag{7.131}$$

where $\mathcal{A}^{\mathrm{het}}(Q)$ denote the single-trace amplitudes for $|Q|$ external gauge multiplets. This follows from (7.123) in combination with

$$\mathcal{A}^{\mathrm{het}}(1, R, n, n{-}1) = -\left(-\frac{\alpha'}{2\pi} \right)^{n-3} \int_{\mathbb{C}^{n-3} \setminus \{z_a = z_b\}} d^2 z_2 \, d^2 z_3 \dots d^2 z_{n-2} \, \mathcal{Z}_{1R} \, \langle \mathcal{K}_n \rangle \prod_{i<j}^{n} |z_{ij}|^{-\alpha' s_{ij}} \tag{7.132}$$

and its $\mathrm{SL}(2, \mathbb{C})$ covariant uplift $\mathcal{Z}_{1R} \to -\mathrm{PT}(1, R, n, n{-}1)$ in (6.59) [137],

$$\mathcal{A}^{\mathrm{het}}(P) = -\sum_{Q, R \in S_{n-3}} J(P|1, R, n, n{-}1) S(R|Q)_1 A(1, Q, n{-}1, n) \, , \tag{7.133}$$

see (7.61) for the sphere integrals $J(P|R)$. One may view (7.131) as an alternative to the double copy (7.119) with locality w.r.t. the massless propagators but not w.r.t. the massive ones. For instance, specializing (7.131) to double-trace amplitudes $\mathcal{A}^{\rm het}(1,P|Q,n)$ of the heterotic string yields [259]

$$\mathcal{A}^{\rm het}(1,P|Q,n) = \frac{\alpha'}{2+\alpha' s_{1P}} \mathcal{A}^{\rm het}(1,\{P,Q\},n)\,, \tag{7.134}$$

with the Einstein–Yang–Mills relation (7.129) in its low-energy limit. The only case where (7.131) is free of massive propagators is the following single-trace amplitude with one external gravity multiplet [259]

$$\mathcal{A}^{\rm het}(1,2,\ldots,n;p) = \sum_{j=1}^{n-1} (\tilde{\epsilon}_p \cdot k_{12\ldots j}) \mathcal{A}^{\rm het}(1,2,\ldots,j,p,j{+}1,\ldots,n{-}1,n)\,, \tag{7.135}$$

which is the α'-uplift of (7.126).

8. α'-expansion of superstring tree-level amplitudes

In the previous sections, we have reviewed the derivation and structure of the expression (6.49) for the n-point disk amplitude in terms of SYM tree amplitudes. By disentangling the contributions of left- and right-moving worldsheet degrees of freedom, similar decompositions (7.70) and (7.119) were deduced for sphere amplitudes of type II and heterotic strings. On the one hand, these genus-zero results only cover the leading order in string perturbation theory and still receive loop corrections. On the other hand, (6.49), (7.70) and (7.119) are exact in α', i.e. they incorporate all orders in the low-energy expansion at genus zero.

This section is dedicated to the α'-expansion of n-point disk and sphere amplitudes, with a detailed review of the structure and explicit computation of the string corrections to the field-theory limits discussed in the previous sections. These string corrections are organized into infinite series in the dimensionless Mandelstam invariants $\alpha' k_i{\cdot}k_j$ with rational combinations of multiple zeta values (MZVs) in their coefficients. The appearance of MZVs unravels elegant mathematical properties of and striking connections between tree-level amplitudes in different perturbative string theories. Moreover, the interplay of MZVs with the accompanying polynomials in $\alpha' k_i{\cdot}k_j$ identifies several echos of field-theory structures at all orders of the low-energy expansion including Berends–Giele recursions, color-kinematics duality and double copy.

The study of low-energy expansions in string perturbation theory has a long history. We focus on state-of-the-art techniques to expand the n-point disk integrals (6.51) or their Z-basis (6.62) using the Drinfeld associator [269], see section 8.5, or Berends–Giele recursions [225], see section 8.6. Based on these results for the disk integrals of open superstrings, the analogous expansions of the sphere integrals in tree amplitudes of type II and heterotic string theories will be obtained as corollaries under the so-called single valued map, see section 8.7. At $n \leq 7$ points, a variety of earlier calculations have been successfully carried out before the advent of the all-multiplicity methods in sections 8.5 and 8.6, often exploiting synergies with hypergeometric functions [213, 198, 214, 199, 216, 217, 270, 271, 166, 272, 273]. The loop-level extensions of the results in this section are under active investigation, and a short summary of the state of the art as of fall 2022 can be found in section 9.2.

Numerous developments related to the α'-expansion of string amplitudes have been crucially fueled by the recent number-theory and algebraic-geometry literature. As will be detailed below, the mathematical references underlying the tree-level results of this section include [274, 275, 276, 277, 278, 279, 280, 223]. Parts of the results of this section can also be found in the reviews [281, 282] from 2016, also see [283] for a helpful introductory reference on the Hopf-algebra structure of genus-zero integrals in the particle-physics literature.

8.1. Basics of α'-expansions

This subsection aims to set the stage for the main results of this section by reviewing four-point examples of α'-expansions and the connection with low-energy effective actions.

8.1.1. Four-point α'-expansions

Similar to the computation and simplification of the correlators, the main efforts in determining low-energy expansions kick in at five points. The four-point α'-expansion in turn has been known in closed form for decades from the simple expansion of the Gamma-function

$$\log\Gamma(1-z) = \gamma z + \sum_{n=2}^{\infty} \frac{z^n}{n} \zeta_n \, , \tag{8.1}$$

where the Euler-Mascheroni drops out from string tree-level computations and the Riemann zeta values are given by convergent infinite sums

$$\zeta_n = \sum_{k=1}^{\infty} k^{-n} \, , \qquad n \geq 2 \, . \tag{8.2}$$

The α'-dependence of the four-point open-superstring amplitude (6.13) can be written as

$$\begin{aligned} {F_2}^2 &= \frac{\Gamma(1-2\alpha' s_{12})\Gamma(1-2\alpha' s_{23})}{\Gamma(1-2\alpha' s_{12}-2\alpha' s_{23})} = \exp\bigg(\sum_{n=2}^{\infty} \frac{\zeta_n}{n} (2\alpha')^n \big[s_{12}^n + s_{23}^n - (s_{12}+s_{23})^n \big] \bigg) \\ &= 1 - (2\alpha')^2 \zeta_2 s_{12} s_{23} + (2\alpha')^3 \zeta_3 s_{12} s_{23} s_{13} - (2\alpha')^4 \zeta_4 s_{12} s_{23} \big(s_{12}^2 + \tfrac{1}{4} s_{12} s_{23} + s_{23}^2 \big) \\ &\quad - (2\alpha')^5 \zeta_2 \zeta_3 s_{12}^2 s_{23}^2 s_{13} + \frac{1}{2} (2\alpha')^5 \zeta_5 s_{12} s_{23} s_{13} (s_{12}^2 + s_{23}^2 + s_{13}^2) + \mathcal{O}(\alpha'^6) \, , \end{aligned} \tag{8.3}$$

where ${F_2}^2$ is the scalar four-point instance of the $(n-3)! \times (n-3)!$ matrix ${F_P}^Q$ of n-point disk integrals in (6.51). The Gamma functions in the numerators of ${F_2}^2$ introduce poles at $2\alpha' s_{12}, 2\alpha' s_{23} = 1, 2, \ldots$, i.e. at center-of-mass energies $(k_i+k_j)^2 \in \frac{\mathbb{N}}{\alpha'}$, that signal the exchange of massive string vibration modes. After α'-expansions, say in the exponential of (8.3), these poles are no longer manifest.

Also for closed superstrings, the α'-expansion of the four-point amplitude in the form (7.55) can be extracted from a scalar combination of Gamma functions,

$$\begin{aligned} \frac{\Gamma(1-\frac{\alpha'}{2}s_{12})\Gamma(1-\frac{\alpha'}{2}s_{23})\Gamma(1-\frac{\alpha'}{2}s_{13})}{\Gamma(1+\frac{\alpha'}{2}s_{12})\Gamma(1+\frac{\alpha'}{2}s_{23})\Gamma(1+\frac{\alpha'}{2}s_{13})} &= \exp\bigg(2 \sum_{k=1}^{\infty} \frac{\zeta_{2k+1}}{2k+1} \Big(\frac{\alpha'}{2}\Big)^{2k+1} \big[s_{12}^{2k+1} + s_{23}^{2k+1} + s_{13}^{2k+1} \big] \bigg) \\ &= 1 + 2\Big(\frac{\alpha'}{2}\Big)^3 \zeta_3 s_{12} s_{23} s_{13} + \Big(\frac{\alpha'}{2}\Big)^5 \zeta_5 s_{12} s_{23} s_{13} (s_{12}^2 + s_{23}^2 + s_{13}^2) + \mathcal{O}(\alpha'^6) \, . \end{aligned} \tag{8.4}$$

The coefficients in this series are still exclusively built from Riemann zeta values (8.2), but there is no more reference to the even zeta values ζ_{2k} seen in the open-string expansion (8.3). While these cancellations at four points can still be understood from Gamma-function expansions, their generalizations to $n \geq 5$ points are governed by an elaborate mathematical structure known as the *single-valued map*, see section 8.7 for details.

8.1.2. Low-energy effective actions

One of the traditional motivations for α'-expansions of massless string amplitudes is to determine the low-energy effective action of the gauge and gravity multiplets. An expansion around $\alpha' \to 0$ amounts to integrating out the massive vibration modes which become infinitely heavy in view of their mass-spectra $M^2 \in \mathbb{N}/\alpha'$ and $M^2 \in 4\mathbb{N}/\alpha'$ for open and closed strings, respectively. This can be anticipated from the fact that the poles of the Gamma functions due to massive-state exchange in (8.3) or (8.4) are no longer manifest in the exponentials encoding the respective α'-expansions, let alone the individual orders in α'.

The joint effort of all massive string modes leads to effective string interactions of schematic form $\alpha'^{m+k-2} D^{2k} \mathbb{F}^m$ and $\alpha'^{m+k-1} D^{2k} \mathbb{R}^m$ (with powers of the non-linear gluon field strength $\mathbb{F}$, Riemann curvature $\mathbb{R}$ and their respective gauge- and diffeomorphism-covariant derivatives D) and their supersymmetrizations. Low-energy effective operators $\alpha'^{m+k-2} \mathrm{Tr}\{D^{2k} \mathbb{F}^m\}$ or $\alpha'^{m+k-1} D^{2k} \mathbb{R}^m$ can be extracted from the

α'-expansion of massless string amplitudes, i.e. by reverse-engineering the Feynman rules that generate a given α'-order of the amplitude. This approach turns out to be cumbersome in practice since[48]

(i) field redefinitions and relations of the schematic form $D^2\mathbb{F} \cong \mathbb{F}^2$ or $D^2\mathbb{R} \cong \mathbb{R}^2$ introduce ambiguities,

(ii) extracting the new information on $D^{2k}\mathbb{F}^n$ or $D^{2k}\mathbb{R}^n$ interactions from n-point amplitudes necessitates the subtraction of reducible-diagram contributions with insertions of $D^{2k}\mathbb{F}^m$ or $D^{2k}\mathbb{R}^m$ at $m < n$,

(iii) operators with four or more field strengths $\mathbb{F}$ and in particular curvature tensors $\mathbb{R}$ admit a large number of Lorentz-index structures, which can be alleviated via manifestly supersymmetric approaches.

In the non-abelian gauge sector of the type I effective action, there are no explicit results beyond the order of α'^4 with all the tensor structures spelled out. The purely bosonic terms at leading orders are given by

$$\begin{aligned} S_{\rm eff}^{\rm open} = \int d^{10}x \, {\rm Tr}\Big\{ -\frac{1}{4}\mathbb{F}_{mn}\mathbb{F}^{mn} + \alpha'^2\zeta_2\Big[-2\mathbb{F}^m{}_p\mathbb{F}^p{}_n\mathbb{F}^q{}_m\mathbb{F}^n{}_q - \mathbb{F}^m{}_n\mathbb{F}^n{}_p\mathbb{F}^p{}_q\mathbb{F}^q{}_m \\ + \tfrac{1}{2}\mathbb{F}^{mn}\mathbb{F}_{mn}\mathbb{F}^{pq}\mathbb{F}_{pq} + \tfrac{1}{4}\mathbb{F}^{mn}\mathbb{F}^{pq}\mathbb{F}_{mn}\mathbb{F}_{pq}\Big] + \mathcal{O}(\alpha'^3)\Big\} + \text{fermions}\,, \end{aligned} \tag{8.5}$$

followed by 8-term expressions $\alpha'^3\zeta_3{\rm Tr}\{D^2\mathbb{F}^4{+}\mathbb{F}^5\}$ and 96-term-expressions $\alpha'^4\zeta_4{\rm Tr}\{D^4\mathbb{F}^4{+}D^2\mathbb{F}^5{+}\mathbb{F}^6\}$ at higher orders in α'. A state-of-the-art method to determine their tensor structure can be found in [285][49], also see [213, 214] and [287, 199, 288] for earlier results at the orders of $\alpha'^{\leq 4}$. The abelian gauge sector of type I superstrings incorporates supersymmetric Born–Infeld theory in its low-energy limit [238], see also [289, 290, 291], and the $\alpha'^3\mathbb{R}^4$ interaction of type II superstrings has been firstly investigated in [292].

8.1.3. On the scope of four-point amplitudes

The polarization dependence of the four-point open- and closed-string amplitudes (6.13) and (7.55) enjoys a simple tensor structure at all orders in α': for the bosonic components, the polarization vectors in the prefactors $A(1,2,3,4)$ and $M_4^{\rm grav}$ combine to linearized field strengths contracted by the famous t_8-tensor

$$t_8(f_1,f_2,f_3,f_4) = f_1^{mn}f_2^{np}f_3^{pq}f_4^{qm} - \frac{1}{4}f_1^{mn}f_2^{mn}f_3^{pq}f_4^{pq} + {\rm cyc}(2,3,4)\,, \tag{8.6}$$

namely

$$\begin{aligned} s_{12}s_{23}A(1,2,3,4) &= -\frac{1}{2}t_8(f_1,f_2,f_3,f_4) + \mathcal{O}(\chi_j)\,, \\ s_{12}s_{23}s_{13}M_4^{\rm grav} &= -\frac{1}{4}t_8(f_1,f_2,f_3,f_4)t_8(\tilde f_1,\tilde f_2,\tilde f_3,\tilde f_4) + \mathcal{O}(\chi_j,\tilde\chi_j)\,, \end{aligned} \tag{8.7}$$

where $A(1,2,3,4)$ and $M_4^{\rm grav}$ are given in (5.34) and (4.159). Accordingly, the closed-form expressions for the four-point α'-expansions (8.3) and (8.4) can be used to swiftly propose operators ${\rm Tr}\{D^{2k}\mathbb{F}^4\}$ or $D^{2k}\mathbb{R}^4$ which reproduce the four-point amplitudes. For instance, the first α'-correction in the open-superstring effective action (8.5) can be written as $-\alpha'^2\zeta_2{\rm Tr}\big(t_8(\mathbb{F},\mathbb{F},\mathbb{F},\mathbb{F})\big)$ and readily reflect the subleading order of

$$\mathcal{A}(1,2,3,4) = A(1,2,3,4)\Big(1 - (2\alpha')^2\zeta_2 s_{12}s_{23} + \mathcal{O}(\alpha'^3)\Big)\,. \tag{8.8}$$

However, the information from the four-point amplitudes does not fix any effective operators ${\rm Tr}\{D^{2k}\mathbb{F}^m\}$ or $D^{2k}\mathbb{R}^m$ with $m \geq 5$ required by non-linear supersymmetry. Moreover, the Mandelstam dependence of the four-point α'-expansion does not fix the order of the non-commutative covariant derivatives D_m acting

[48] For a brief review and for the practical struggles associated with (i) and (ii), see [284].

[49] The absence of tensor structures $(e_j \cdot k_i)^n$ in disk amplitudes was recognized as a valuable source of information on the open-string effective action [285] and properties of the amplitudes themselves [286].

on $\mathbb{F}^4$ or $\mathbb{R}^4$. At the time of writing, it is not clear whether one can find a tensor structure analogous to t_8 that governs the five-field operators $\mathrm{Tr}\{D^{2k}\mathbb{F}^5\}$ and $D^{2k}\mathbb{R}^5$ at all orders in α'.[50]

For closed-string effective actions, the complexity proliferates more drastically with the order in α'. As an additional complication as compared to open strings, already the simplest α'-correction $\alpha'^3\zeta_3\mathbb{R}^4$ to the ten-dimensional type II supergravity action goes beyond the t_8-tensor: The sixteen Lorentz indices of $\prod_{j=1}^4 \mathbb{R}^{m_j n_j p_j q_j}$ are contracted with a combination of $t_8 t_8$ and two ten-dimensional Levi–Civita tensors $\varepsilon_{10}^{abm_1n_1m_2n_2m_3n_3m_4n_4}\varepsilon_{10}^{abp_1q_1p_2q_2p_3q_3p_4q_4}$ in the tree-level effective action of both type IIA and type IIB superstrings [295, 296, 297], where the $\varepsilon_{10}\varepsilon_{10}$ terms do not contribute to four-point amplitudes.

Furthermore, the type II supergravity multiplets have a much richer structure than the gauge multiplet of ten-dimensional SYM. The multitude of superpartners of the type IIB tree-level interaction $\alpha'^3\zeta_3(t_8t_8 + \varepsilon_{10}\varepsilon_{10})\mathbb{R}^4$ for instance includes a sixteen-dilatino term [298]. More generally, the operators in the type IIB effective action are organized according to their charges w.r.t. the $U(1)$ R-symmetry of type IIB supergravity which is broken by the string corrections.

8.1.4. Manifestly supersymmetric approaches

The manifestly supersymmetric form of the n-point disk amplitude in (6.49) severely constrains the effective action of type I superstrings to all orders in α': The amplitudes computed from reducible and irreducible diagrams at various orders in α' must conspire to linear combinations of SYM trees. It is an open problem to translate this property together with the all-order results on the α'-expansion of disk integrals to be reviewed below into a new line of attack for the effective action.

By the success of pure spinor methods to obtain compact expressions for n-point amplitudes, one can expect that the open questions on effective actions will benefit from superspace methods. The supersymmetrization of the $\mathbb{F}^4$ interaction in (8.5) has a long history [299, 300, 301, 302] and manifestly supersymmetric formulations of more general effective string interactions have for instance been discussed in [303, 304, 305, 306, 307, 89, 288, 308, 309].

8.2. Multiple zeta values

The coefficients in the low-energy expansion of n-point string amplitudes and the associated low-energy effective action are rational linear combinations of multiple zeta values (MZVs) [274, 276]

$$\zeta_{n_1,n_2,\ldots,n_r} = \sum_{0<k_1<k_2<\ldots<k_r}^{\infty} k_1^{-n_1}k_2^{-n_2}\ldots k_r^{-n_r}\,, \qquad n_1,n_2,\ldots,n_r \in \mathbb{N}\,, \qquad n_r \geq 2 \tag{8.9}$$

that generalize the Riemann zeta values (8.2) to depend on multiple integers n_j. The infinite sum converges if $n_r \geq 2$, and we refer to r and $n_1{+}n_2{+}\ldots{+}n_r$ as the *depth* and the *weight* of the MZV, respectively. While even zeta values ζ_{2k} are rational multiples of π^{2k} (with B_{2k} denoting the Bernoulli numbers[51]),

$$\zeta_{2k} = -\frac{(2\pi i)^{2k}\mathrm{B}_{2k}}{2(2k)!}\,, \tag{8.10}$$

the numbers $\pi, \zeta_3, \zeta_5, \zeta_7, \ldots$ are conjectured to be algebraically independent over $\mathbb{Q}$. MZVs arise from iterated integrals over logarithmic forms $d\log(z_j{-}a_j)$ with $a_j \in \{0,1\}\ \forall\ j = 1,2,\ldots,w$ and $a_1 \neq 0$,

$$\begin{aligned} I(0;a_1a_2\ldots a_w;z) &= \int_{0<z_1<z_2<\ldots<z_w<z} \frac{dz_1}{z_1{-}a_1}\frac{dz_2}{z_2{-}a_2}\ldots\frac{dz_w}{z_w{-}a_w}\,, \\ \zeta_{n_1,n_2,\ldots,n_r} &= (-1)^r I(0;1\underbrace{0\ldots0}_{n_1-1}1\underbrace{0\ldots0}_{n_2-1}\ldots1\underbrace{0\ldots0}_{n_r-1};1)\,, \end{aligned} \tag{8.11}$$

[50]See for instance [293, 294] for explicit tensors t_r contracting $r \geq 16$ indices in eight-derivative interactions related to superpartners of $\mathbb{R}^4$.

[51]The generating function $\frac{t}{e^t-1} = \sum_{m=0}^{\infty}\frac{t^m}{m!}\mathrm{B}_m$ leads to even Bernoulli numbers such as $\mathrm{B}_2 = \frac{1}{6}$, $\mathrm{B}_4 = -\frac{1}{30}$ and $\mathrm{B}_6 = \frac{1}{42}$ whereas the odd ones vanish, $\mathrm{B}_{2k+1} = 0\ \forall\ k \in \mathbb{N}$, apart from $\mathrm{B}_1 = -\frac{1}{2}$.

i.e. multiple polylogarithms at unit argument. The combined set of relations following from the iterated-integral and nested-sum representations can be used to reduce any MZV of weight $w \leq 7$ to products of Riemann zeta values ζ_n and leave the conjectural bases over $\mathbb{Q}$ in table 1. The first instances of irreducible MZVs at depth 2 and 3 are believed to occur at weight 8 (e.g. $\zeta_{3,5}$) and weight 11 (e.g. $\zeta_{3,3,5}$), respectively.

w	0	1	2	3	4	5	6	7	8	9	10	11
MZV	1		ζ_2	ζ_3	ζ_2^2	ζ_5 $\zeta_2\zeta_3$	ζ_2^3 ζ_3^2	ζ_7 $\zeta_2\zeta_5$ $\zeta_2^2\zeta_3$	ζ_2^4 $\zeta_{3,5}$ $\zeta_2\zeta_3^2$ $\zeta_3\zeta_5$	ζ_9 $\zeta_2\zeta_7$ $\zeta_2^2\zeta_5$ $\zeta_2^3\zeta_3$ ζ_3^3	ζ_2^5 $\zeta_2\zeta_{3,5}$ $\zeta_{3,7}$ $\zeta_2^2\zeta_3^2$ ζ_5^2 $\zeta_3\zeta_7$ $\zeta_2\zeta_3\zeta_5$	ζ_{11} $\zeta_{3,3,5}$ $\zeta_2\zeta_9$ $\zeta_2^2\zeta_7$ $\zeta_2^3\zeta_5$ $\zeta_2^4\zeta_3$ $\zeta_3^2\zeta_5$ $\zeta_2\zeta_3^3$ $\zeta_3\zeta_{3,5}$
$\dim_w$	1	0	1	1	1	2	2	3	4	5	7	9

Table 1: Conjectural $\mathbb{Q}$-bases of MZVs at weights $w \leq 11$.

Comprehensive references on MZVs include [310, 311], and a datamine of $\mathbb{Q}$-relations with machine-readable ancillary files can be found in [312]. Any known relation among MZVs over $\mathbb{Q}$ preserves the weight, and the dimensions $\dim_w$ of the tentative $\mathbb{Q}$-bases at weight w are conjectured to obey the recursion $\dim_w = \dim_{w-2} + \dim_{w-3}$ with $\dim_0 = 1 = \dim_2$ and $\dim_1 = 0$ [313], see table 1 for possible representatives.

8.2.1. Motivic MZVs and the f alphabet

The conjectural counting of $\mathbb{Q}$-linearly independent MZVs through the above recursion for $\dim_w$ can be reproduced from a simple model, the so-called *f-alphabet* [277]: introduce non-commutative variables $f_3, f_5, f_7, \ldots$ for each odd integer ≥ 3, a single commutative variable f_2 and assign weight w to f_w. It is easy to show that the number of weight-w compositions (non-commutative words in f_{2m+1} along with non-negative powers of f_2) is counted by $\dim_w$ in table 1, e.g. $\{f_5,\ f_2 f_3\}$ at weight five or $\{f_2^4,\ f_2 f_3 f_3,\ f_3 f_5,\ f_5 f_3\}$ at weight eight. Note in particular that the first instance $f_3 f_5 \neq f_5 f_3$ of non-commutativity ties in with the first conjecturally irreducible MZV $\zeta_{3,5}$ beyond depth one.

It is tempting to map MZVs into the f-alphabet, i.e. the Hopf-algebra comodule $\mathcal{U} = \mathbb{Q}\langle f_3, f_5, \ldots\rangle \otimes_{\mathbb{Q}} \mathbb{Q}[f_2]$, in order to manifestly mod out by their $\mathbb{Q}$-relations. However, the unsettled transcendentality properties of MZVs currently obstruct a well-defined map to $\mathcal{U}$. As a workaround, one can consider *motivic MZVs* $\zeta^{\mathfrak{m}}_{n_1,n_2,\ldots,n_r}$ instead of the $\zeta_{n_1,n_2,\ldots,n_r} \in \mathbb{R}$ in (8.9). By definition, motivic MZVs obey the complete set of $\mathbb{Q}$ relations among $\zeta_{n_1,n_2,\ldots,n_r}$ known up to date, and their elaborate definition in the framework of algebraic geometry can be found in [275, 277, 314].

By passing to motivic MZVs, one can set up an invertible map ϕ to the f-alphabet starting from the normalization

$$\phi(\zeta^{\mathfrak{m}}_{2k+1}) = f_{2k+1} \, , \qquad \phi(\zeta^{\mathfrak{m}}_2) = f_2 \, . \tag{8.12}$$

For motivic MZVs $\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r}$ of weight w beyond depth one, the ϕ-image up to adding a $\mathbb{Q}$-multiple of $\phi(\zeta^{\mathfrak{m}}_w)$ can be determined from the shuffle product $\sqcup\!\sqcup$ and deconcatenation coaction Δ in $\mathcal{U}$:

$$\begin{aligned} f_A \sqcup\!\sqcup f_B &= \sum_{C \in A \sqcup\!\sqcup B} f_C \, , \\ \Delta(f_2^n f_A) &= f_2^n \sum_{A=BC} f_B \otimes f_C \, . \end{aligned} \tag{8.13}$$

We employ the shorthand $f_A = f_{a_1} f_{a_2} \ldots f_{a_{|A|}}$ for $A = a_1 a_2 \ldots a_{|A|}$ (with $f_\emptyset = 1$), and the sum over $\sum_{A=BC}$ includes the terms with $B = \emptyset$ or $C = \emptyset$, for instance

$$\begin{aligned} (f_3 f_5) \sqcup\!\sqcup f_7 &= f_3 f_5 f_7 + f_3 f_7 f_5 + f_7 f_3 f_5 \, , \\ \Delta(f_2 f_9 f_3) &= f_2 \otimes f_9 f_3 + f_2 f_9 \otimes f_3 + f_2 f_9 f_3 \otimes 1 \, . \end{aligned} \tag{8.14}$$

More specifically, the ϕ-images of motivic MZVs are required to preserve the product and coaction structure in the sense of

$$\begin{aligned} \phi(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r}\cdot\zeta^{\mathfrak{m}}_{p_1,\ldots,p_s}) &= \phi(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r}) \sqcup\!\sqcup\, \phi(\zeta^{\mathfrak{m}}_{p_1,\ldots,p_s})\,, \\ \Delta\big(\phi(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})\big) &= \phi\big(\Delta(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})\big)\,. \end{aligned} \tag{8.15}$$

The shuffle symbol in the first line is understood to act trivially on the commutative variable f_2, e.g. $\phi(\zeta^{\mathfrak{m}}_2\zeta^{\mathfrak{m}}_3) = f_2 f_3$ and $\phi\big((\zeta^{\mathfrak{m}}_2)^k\big) = f_2^k$ for and $k\in\mathbb{N}$. In the second line of (8.15), the coaction $\Delta(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})$ can be obtained from [275] and leaves the freedom to shift $\phi(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})$ at higher depth by the depth-one image $\phi(\zeta^{\mathfrak{m}}_{n_1+\ldots+n_r})$.[52] As the simplest examples of this procedure, the (conjecturally indecomposable) MZVs beyond depth one in table 1 are mapped to

$$\begin{aligned} \phi(\zeta^{\mathfrak{m}}_{3,5}) &= -5f_3f_5\,, \qquad \phi(\zeta^{\mathfrak{m}}_{3,7}) = -14f_3f_7 - 6f_5f_5\,, \\ \phi(\zeta^{\mathfrak{m}}_{3,3,5}) &= -5f_3f_3f_5 - 45f_9f_2 - \frac{6}{5}f_7f_2^2 + \frac{4}{7}f_5f_2^3\,, \end{aligned} \tag{8.16}$$

where we have chosen to exclude f_8, f_{10} and f_{11} from $\phi(\zeta^{\mathfrak{m}}_{3,5}), \phi(\zeta^{\mathfrak{m}}_{3,7})$ and $\phi(\zeta^{\mathfrak{m}}_{3,3,5})$, respectively. The coefficients of f_w in each other $\phi(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})$ at weight $w = 8, 10$ or 11 are determined by (8.12), (8.15), (8.16) and imposing that ϕ preserves the $\mathbb{Q}$-relations among motivic MZVs.

Higher-weight instances of $\phi(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})$ can be found in [271], where the MZVs beyond depth one in the conjectural $\mathbb{Q}$ bases of [312] are taken to have no f_w in their ϕ-images (also see section 8.2.4 for comments on the conventions). As will be reviewed in section 8.3, the α'-expansion of the disk integrals $F_P{}^Q$ in (6.51) takes a very compact form once the (motivic) MZVs in the coefficients are translated into the f-alphabet.

8.2.2. The Drinfeld associator

As will be described in section 8.5, the MZVs in the α'-expansion of n-point disk integrals can be derived from the *Drinfeld associator*, a generating series of MZVs. Besides the MZVs in (8.11) from convergent iterated integrals $I(0;1\ldots0;1)$, the Drinfeld associator also involves so-called shuffle-regularized MZVs which descend from formally divergent integrals. For the iterated integrals $I(0;a_1\ldots a_w;1)$ in (8.11), we assign regularized values

$$I(0;0;1) = I(0;1;1) = 0 \tag{8.17}$$

to the divergent cases at weight 1. At higher weight, the regularized values of integrals $I(0;0\ldots;1)$ or $I(0;\ldots1;1)$ with endpoint divergences are defined by (8.17) and by imposing them to obey the shuffle relations of convergent $I(0;1\ldots0;1)$,

$$I(0;A;1)I(0;B;1) = \sum_{C\in A\sqcup\!\sqcup B} I(0;C;1)\,, \qquad \text{also for } a_1, b_1 = 0 \text{ and } a_{|A|}, b_{|B|} = 1\,. \tag{8.18}$$

One can recursively remove leading zeros by relating $I(0;0^kD;1)$ with $d_1 = 1$ to $I(0;0^k;1)I(0;D;1)$ minus terms with $\leq k{-}1$ leading zeros, for instance $I(0;01;1) = I(0;0;1)I(0;1;1) - I(0;10;1) = \zeta_2$. Similarly, subtracting $I(0;D;1)I(0;1^k;1)$ from $I(0;D1^k;1)$ with $d_{|D|} = 0$ yields terms with $\leq k{-}1$ terminal ones, see for instance [316] for further details. The resulting regularized values of $I(0;A;1)$ with $a_1 = 0$ and/or $a_{|A|} = 1$ are known as shuffle-regularized MZVs.

The Drinfeld associator $\Phi(e_0,e_1)$ (not to be confused with the perturbiners for bi-adjoint scalars in section 6.4.4) is a generating series of shuffle-regularized MZVs, where the coefficient of $I(0;A;1)$ is a word $e_A = e_{a_1}e_{a_2}\ldots e_{a_w}$ of non-commutative variables e_0, e_1,

$$\Phi(e_0,e_1) = \sum_{A\in\{0,1\}^\times} (-1)^{\sum_{j=1}^{|A|} a_j} I(0;A;1)e_A$$

[52] Strictly speaking, the second entry of the coaction $\Delta(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})$ involves deRham periods $\zeta^{\mathfrak{dr}}_{n_1,\ldots,n_r}$, where the deRham version of ζ_2 vanishes, see for instance [315].

$$
\begin{aligned}
&= 1 + \zeta_2[e_0, e_1] + \zeta_3[e_0-e_1, [e_0, e_1]] \\
&\quad + \zeta_4\Big([e_0, [e_0, [e_0, e_1]]] + \tfrac{1}{4}[e_1, [e_0, [e_1, e_0]]] + [e_1, [e_1, [e_0, e_1]]] + \tfrac{5}{4}[e_0, e_1]^2\Big) + \dots ,
\end{aligned} \tag{8.19}
$$

and the summation range $\{0,1\}^\times$ denotes the set of words (of arbitrary length $0,1,2,\dots$) in letters $0,1$. In the first place, the Drinfeld associator has been introduced as the universal monodromy of the KZ equation

$$
\frac{dF(z)}{dz} = \Big(\frac{e_0}{z} + \frac{e_1}{1-z}\Big)F(z)\,, \tag{8.20}
$$

(with e_0, e_1 some non-commutative indeterminates) relating its regularized boundary values C_0, C_1 [317, 318]:

$$
C_0 = \lim_{z\to 0} z^{-e_0}F(z)\,, \quad C_1 = \lim_{z\to 1}(1-z)^{e_1}F(z) \quad \Rightarrow \quad C_1 = \Phi(e_0,e_1)C_0\;. \tag{8.21}
$$

The equivalence of this definition to the generating series (8.19) was then shown by Le and Murakami [319]. The relevance of the Drinfeld associator for open-string amplitudes will later on be illustrated by presenting $(n-2)!$-component vectors F subject to (8.20) and related to disk integrals, with matrix representations of e_0, e_1 linear in $\alpha' s_{ij}$.

8.2.3. Single-valued multiple zeta values

In comparing the α'-expansions of the four-point disk and sphere integrals (8.3) and (8.4), we already noted the dropout of even zeta values from the closed-string amplitude. The n-point systematics of dropouts in passing from open to closed strings is captured by the notion of *single-valued MZVs* to be reviewed in this section.

The terminology is borrowed from the polylogarithms that specialize to MZVs at unit argument, see (8.11): while the meromorphic polylogarithms are notoriously multivalued as the defining integration path is deformed by loops around $z=0$ or $z=1$, one can form single-valued combinations by adjoining complex conjugates. For instance, the real part of the multivalued $I(0;1;z) = \log(1-z)$ yields the single-valued $I^{\rm sv}(0;1;z) = I(0;1;z) + \overline{I(0;1;z)} = \log|1-z|^2$.

At higher weight, single-valued polylogarithms $I^{\rm sv}(0;A;z)$ can be systematically constructed from products of $I(0;B;z)\overline{I(0;C;z)}$ and MZVs as detailed in [320]. The guiding principle of the reference is to preserve the holomorphic derivatives

$$
\partial_z I(0;Ab;z) = \frac{I(0;A;z)}{z-b} \quad \leftrightarrow \quad \partial_z I^{\rm sv}(0;Ab;z) = \frac{I^{\rm sv}(0;A;z)}{z-b} \tag{8.22}
$$

on the expense of more complicated expressions for the antiholomorphic derivatives $\partial_{\bar z} I^{\rm sv}(0;Ab;z)$. The weight-two example $I^{\rm sv}(0;10;z) = I(0;10;z) + I(0;0;z)\overline{I(0;1;z)} + \overline{I(0;01;z)}$ illustrates that shuffle-regularized versions of polylogarithms (based on $I(0;0;z) = \log(z)$) are encountered even if the holomorphic part (in this case $I(0;10;z)$) is convergent.

In the same way as meromorphic polylogarithms yield MZVs at $z=1$, see (8.11), we define single-valued MZVs as single-valued polylogarithms at unit argument [278, 279],

$$
\zeta^{\rm sv}_{n_1,n_2,\dots,n_r} = (-1)^r I^{\rm sv}(0;1\underbrace{0\dots0}_{n_1-1}1\underbrace{0\dots0}_{n_2-1}\dots1\underbrace{0\dots0}_{n_r-1};1)\,. \tag{8.23}
$$

At depth one, this annihilates even zeta values and doubles odd ones,

$$
\zeta^{\rm sv}_{2k} = 0\,, \qquad \zeta^{\rm sv}_{2k+1} = 2\zeta_{2k+1}\,, \tag{8.24}
$$

and the expressions for single-valued MZVs at higher depth are usually less straightforward, e.g.

$$
\begin{aligned}
&\zeta^{\rm sv}_{3,5} = -10\zeta_3\zeta_5\,, \qquad \zeta^{\rm sv}_{3,7} = -28\zeta_3\zeta_7 - 12\zeta_5^2\,, \\
&\zeta^{\rm sv}_{3,3,5} = 2\zeta_{3,3,5} - 5\zeta_3^2\zeta_5 + 90\zeta_2\zeta_9 + \frac{12}{5}\zeta_2^2\zeta_7 - \frac{8}{7}\zeta_2^3\zeta_5\,.
\end{aligned} \tag{8.25}
$$

It is clear by the constituents of $I^{\rm sv}(0;A;z)$ that single-valued MZVs can be expressed in terms of $\mathbb{Q}$-linear combinations of MZVs.

The above constructions are formalized through the *single-valued map* that sends both meromorphic polylogarithms and arbitrary MZVs to their single-valued versions. However, the single-valued map sv of MZVs is only well-defined in a motivic setup, i.e. (8.24) and (8.25) are understood as

$$\mathrm{sv}(\zeta^{\mathfrak{m}}_{2k})=0\,, \qquad \mathrm{sv}(\zeta^{\mathfrak{m}}_{2k+1})=2\zeta^{\mathfrak{m}}_{2k+1}\,, \qquad \mathrm{sv}(\zeta^{\mathfrak{m}}_{3,5})=-10\zeta^{\mathfrak{m}}_{3}\zeta^{\mathfrak{m}}_{5}\,. \tag{8.26}$$

As a major advantage of adapting the single-valued map to motivic MZVs, one can employ the f-alphabet where the single-valued map follows a simple closed formula at arbitrary weight and depth [279],

$$\mathrm{sv}(f_2^n f_{i_1}f_{i_2}\ldots f_{i_r})=\delta_{n,0}\sum_{j=0}^{r} f_{i_j}\ldots f_{i_2}f_{i_1}\sqcup\!\sqcup\, f_{i_{j+1}}f_{i_{j+2}}\ldots f_{i_r}\,, \tag{8.27}$$

where $i_1,i_2,\ldots,i_r\in 2\mathbb{N}{+}1$, for instance

$$\begin{aligned}\mathrm{sv}(f_{i_1})=2f_{i_1}\,, \qquad \mathrm{sv}(f_{i_1}f_{i_2})&=2f_{i_1}\sqcup\!\sqcup\, f_{i_2}=2(f_{i_1}f_{i_2}+f_{i_2}f_{i_1})\,, \\ \mathrm{sv}(f_{i_1}f_{i_2}f_{i_3})=2(f_{i_1}f_{i_2}f_{i_3}&+f_{i_3}f_{i_2}f_{i_1}+f_{i_2}f_{i_1}f_{i_3}+f_{i_2}f_{i_3}f_{i_1})\,.\end{aligned} \tag{8.28}$$

In slight abuse of notation, we are employing the same notation sv for the single-valued map of motivic MZVs and the induced single-valued map $\phi\,\mathrm{sv}\,\phi^{-1}$ in the f alphabet. Since $f_{i_1}f_{i_2}$ and $f_{i_2}f_{i_1}$ (with i_1,i_2 odd) are indistinguishable under the single-valued map by (8.28), irreducible double zetas such as $\zeta^{\mathfrak{m}}_{3,5},\zeta^{\mathfrak{m}}_{3,7}$ in table 1 factorize into products of odd Riemann zeta values. Accordingly, $\zeta^{\rm sv}_{3,3,5}$ in (8.25) is the simplest indecomposable single-valued MZV beyond depth one.

Note that the single-valued map preserves the product structure,

$$\mathrm{sv}(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r}\cdot\zeta^{\mathfrak{m}}_{p_1,\ldots,p_s})=\mathrm{sv}(\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r})\cdot\mathrm{sv}(\zeta^{\mathfrak{m}}_{p_1,\ldots,p_s})\,, \tag{8.29}$$

as one can check from its f-alphabet representation (8.27). In section 8.7, we will apply the single-valued map to the α'-expansions of disk integrals which then acts on (motivic) MZVs at various weights.

8.2.4. Comments on conventions

In comparing the material of this section with the literature on MZVs and their f-alphabet description, the reader should be warned about two sources of mismatching conventions. First, many references including [312, 310, 311] define the nested sum on the right-hand side of (8.9) to be $\zeta_{n_r,\ldots,n_2,n_1}$ instead of $\zeta_{n_1,n_2,\ldots,n_r}$, and our ordering conventions for the arguments of MZVs agree with those of [313, 274, 276, 314, 277, 271, 321, 278, 279, 322, 137, 281, 282]. Second, our conventions for the motivic coaction are those of [283, 321, 281] but differ from [314, 277, 271, 278, 279, 322, 137, 311, 282] by the swap $A\otimes B\leftrightarrow B\otimes A$.

In the above references with opposite conventions for the motivic coaction, the order of the non-commutative f_{2k+1} will be reversed in comparison to the expressions in this work. In particular, the coaction of the commutative f_2 becomes $1\otimes f_2$ instead of $f_2\otimes 1$ in translating to those references. Moreover, the single-valued map of $f_2^n f_{i_1}f_{i_2}\ldots f_{i_r}$ becomes $\delta_{n,0}\sum_{j=0}^r f_{i_1}f_{i_2}\ldots f_{i_j}\sqcup\!\sqcup\, f_{i_r}f_{i_{r-1}}\ldots f_{i_{j+1}}$ in the place of (8.27) when changing the conventions to $A\otimes B\to B\otimes A$.

For instance, the (conjectural) $\mathbb{Q}$-bases of MZVs employed in the datamine [312] are related to those of this work and [271] by reversing $\zeta_{n_1,n_2,\ldots,n_r}\to\zeta_{n_r,\ldots,n_2,n_1}$. However, in order to import the ϕ-images at weight ≤ 16 from [271] into our present conventions, the order of the non-commutative letters f_{2k+1} requires a separate reversal, i.e. $f_{i_1}f_{i_2}\ldots f_{i_r}\to f_{i_r}\ldots f_{i_2}f_{i_1}$ with $i_j\in 2\mathbb{N}{+}1$.

8.3. Patterns in the α'-expansion

In this section, we review the structure of the α'-expansion of the n-point disk integrals $F_P{}^Q$ in (6.51) which has been firstly described in [271]. First of all, the MZVs contributing to the order of α'^w have total weight w, i.e. the α'-expansion of $F_P{}^Q$ is said to enjoy uniform transcendentality, see section 8.5 for a proof.

Once the MZVs at given weight w are organized in the conjectural $\mathbb{Q}$-bases of table 1, the coefficients of the Riemann zeta values ζ_w are claimed to determine all other coefficients, say those of indecomposable higher-depth zetas $\zeta_{3,5}$ or products such as $\zeta_a\zeta_b$ or $\zeta_a\zeta_{b,c}$. These intriguing patterns are checked for a variety of weights & multiplicities and most conveniently described in the f-alphabet of section 8.2.1. They imply a remarkably simple formula for the coaction of the integrals $F_P{}^Q$ [321] that resonates with recent studies of Feynman integrals [323, 324, 325, 326] and Lauricella hypergeometric functions [327].

8.3.1. The pattern in terms of MZVs

For the four-point instance of the disk integrals $F_P{}^Q$ in (6.51), the α'-expansion is given in closed form by the exponential in (8.3). This expression manifests that the coefficients,

$$M_{2k+1}\big|_{n=4} = (2\alpha')^{2k+1}\frac{s_{12}^{2k+1}+s_{23}^{2k+1}+s_{13}^{2k+1}}{2k+1}\,, \qquad P_{2k}\big|_{n=4} = (2\alpha')^{2k}\frac{\zeta_{2k}(s_{12}^{2k}+s_{23}^{2k}-s_{13}^{2k})}{2k(\zeta_2)^k}\,, \tag{8.30}$$

of ζ_{2k+1} and ζ_{2k} determine those of products $\zeta_{a_1}\zeta_{a_2}\ldots$ by expanding the exponential. In order to generalize this observation to $n\geq 5$ points, we shall consider the matrix-valued coefficients of Riemann zeta values

$$(M_{2k+1})_P{}^Q = F_P{}^Q\big|_{\zeta_{2k+1}}\,, \qquad (P_{2k})_P{}^Q = F_P{}^Q\big|_{\zeta_2^k}\,, \tag{8.31}$$

where the entries of the $(n{-}3)!\times(n{-}3)!$ matrices P_w and M_w are homogeneous degree-w polynomials in $2\alpha' s_{ij}$ with rational coefficients. A variety of examples at $n=5,6,7$ are available for download from [328], where the conventions in this review are matched after rescaling $s_{ij}\to -2\alpha' s_{ij}$ in the dataset of the website. At the leading orders in α', the expansion of n-point disk integrals is found to exhibit the following multiplicity-agnostic pattern [271]

$$\begin{aligned} F = \mathbb{1} &+ \zeta_2 P_2 + \zeta_3 M_3 + \zeta_2^2 P_4 + \zeta_5 M_5 + \zeta_2\zeta_3 P_2 M_3 \\ &+ \zeta_2^3 P_6 + \frac{1}{2}\zeta_3^2 M_3^2 + \zeta_7 M_7 + \zeta_2\zeta_5 P_2 M_5 + \zeta_2^2\zeta_3 P_4 M_3 + \mathcal{O}(\alpha'^8)\,, \end{aligned} \tag{8.32}$$

where we have suppressed the row and column indices of the $F_P{}^Q$ in (6.51). The coefficients P_w, M_w of ζ_w defined in (8.31) turn out to determine those of $\zeta_2\zeta_3$ or ζ_3^2 via matrix products P_2M_3 or M_3^2. In other words, there is only one piece of independent information P_w or M_w at each order $\alpha'^{w\leq 7}$ in (8.32).

Starting from weights $w=8,10,11,\ldots$, the $\mathbb{Q}$-bases of MZVs are believed to contain indecomposable elements of depth ≥ 2 which can be chosen as $\zeta_{3,5},\zeta_{3,7},\zeta_{3,3,5},\ldots$ [312]. In the conjectural bases of table 1, the simplest instances of depth-two and depth-three MZVs are accompanied by the following matrix commutators [271]

$$\begin{aligned} F\big|_{(\alpha')^8} &= \zeta_2^4 P_8 + \frac{1}{2}\zeta_2\zeta_3^2 P_2 M_3^2 + \zeta_3\zeta_5 M_5 M_3 + \frac{1}{5}\zeta_{3,5}[M_5,M_3]\,, \\ F\big|_{(\alpha')^9} &= \zeta_9 M_9 + \zeta_2\zeta_7 P_2 M_7 + \zeta_2^2\zeta_5 P_4 M_5 + \zeta_2^3\zeta_3 P_6 M_3 + \frac{1}{6}\zeta_3^3 M_3^3\,, \\ F\big|_{(\alpha')^{10}} &= \zeta_2^5 P_{10} + \zeta_2\zeta_3\zeta_5 P_2 M_5 M_3 + \frac{1}{5}\zeta_2\zeta_{3,5} P_2[M_5,M_3] + \frac{1}{2}\zeta_2^2\zeta_3^2 P_4 M_3^2 \\ &\quad + \frac{1}{2}\zeta_5^2 M_5^2 + \zeta_3\zeta_7 M_7 M_3 + \Big(\frac{1}{14}\zeta_{3,7} + \frac{3}{14}\zeta_5^2\Big)[M_7,M_3]\,, \\ F\big|_{(\alpha')^{11}} &= \zeta_{11} M_{11} + \zeta_2\zeta_9 P_2 M_9 + \zeta_2^2\zeta_7 P_4 M_7 + \zeta_2^3\zeta_5 P_6 M_5 + \zeta_2^4\zeta_3 P_8 M_3 + \frac{1}{6}\zeta_2\zeta_3^3 P_2 M_3^3 \\ &\quad + \frac{1}{2}\zeta_3^2\zeta_5 M_5 M_3^2 + \frac{1}{5}\zeta_{3,5}\zeta_3[M_5,M_3]M_3 + \Big(\frac{1}{5}\zeta_{3,3,5} + 9\zeta_2\zeta_9 + \frac{6}{25}\zeta_2^2\zeta_7 - \frac{4}{35}\zeta_2^3\zeta_5\Big)[M_3,[M_5,M_3]]\,. \end{aligned} \tag{8.33}$$

These expressions are consistent with the dropout of $\zeta_{3,5},\zeta_{3,7}$ and $\zeta_{3,3,5}$ at four points, see (8.3), since the $(n{-}3)!\times(n{-}3)!$ matrices M_{2k+1} then reduce to the scalars (8.30) with vanishing commutators. However, rational prefactors such as $\frac{1}{5}$ of $\zeta_{3,5}[M_5,M_3]$ or $-\frac{4}{35}$ of $\zeta_2^3\zeta_5[M_3,[M_5,M_3]]$ may appear surprising at first glance. As we will see in the next section, these rational numbers conspire to unit coefficients once the MZVs in (8.33) are taken to be motivic ones and mapped into the f-alphabet reviewed in section 8.2.1.

8.3.2. The pattern in the f alphabet

In preparation for a well-defined map into the f-alphabet, we promote the MZVs in the α'-expansion of the matrix F to their motivic versions,

$$(F^{\mathfrak{m}})_P{}^Q = F_P{}^Q\,\big|_{\zeta_{n_1,\ldots,n_r}\to\zeta^{\mathfrak{m}}_{n_1,\ldots,n_r}} \,. \tag{8.34}$$

The image of (8.33) in the f-alphabet can be assembled from the action (8.12), (8.15) and (8.16) of the ϕ-isomorphism on $\zeta^{\mathfrak{m}}_w, \zeta^{\mathfrak{m}}_{3,5}, \zeta^{\mathfrak{m}}_{3,3,5}$ and products thereof,

$$\begin{aligned}
\phi(F^{\mathfrak{m}})\,\big|_{(\alpha')^8} &= f_2^4 P_8 + f_2 f_3 f_3 P_2 M_3^2 + f_3 f_5 M_3 M_5 + f_5 f_3 M_5 M_3 \,,\\
\phi(F^{\mathfrak{m}})\,\big|_{(\alpha')^9} &= f_9 M_9 + f_2 f_7 P_2 M_7 + f_2^2 f_5 P_4 M_5 + f_2^3 f_3 P_6 M_3 + f_3 f_3 f_3 M_3^3 \,,\\
\phi(F^{\mathfrak{m}})\,\big|_{(\alpha')^{10}} &= f_2^5 P_{10} + f_2 f_3 f_5 P_2 M_3 M_5 + f_2 f_5 f_3 P_2 M_5 M_3 + f_2^2 f_3 f_3 P_4 M_3^2 \\
&\quad + f_5 f_5 M_5^2 + f_3 f_7 M_3 M_7 + f_7 f_3 M_7 M_3 \,,\\
\phi(F^{\mathfrak{m}})\,\big|_{(\alpha')^{11}} &= f_{11} M_{11} + f_2 f_9 P_2 M_9 + f_2^2 f_7 P_4 M_7 + f_2^3 f_5 P_6 M_5 + f_2^4 f_3 P_8 M_3 + f_2 f_3^3 P_2 M_3^3 \\
&\quad + f_3 f_3 f_5 M_3^2 M_5 + f_3 f_5 f_3 M_3 M_5 M_3 + f_5 f_3 f_3 M_5 M_3^2 \,.
\end{aligned} \tag{8.35}$$

Each word in the non-commutative generators f_{2k+1} is accompanied by a matrix product of M_{2k+1} with a matching multiplication order, and powers of the commutative ϕ-image f_2 of $\zeta_2^{\mathfrak{m}}$ occur with left-multiplicative $f_2^k P_{2k}$. Moreover, all the unwieldy rational prefactors of $\frac{1}{5}$ or $-\frac{4}{35}$ in (8.33) have conspired to unit coefficients in passing to (8.35)! By reinstating the lower-weight results in the f-alphabet[53], the orders of $\alpha'^{\leq 11}$ can be reconstructed from

$$\begin{aligned}
\phi(F^{\mathfrak{m}}) &= \big(\mathbb{1} + f_2 P_2 + f_2^2 P_4 + f_2^3 P_6 + f_2^4 P_8 + f_2^5 P_{10}\big)\\
&\quad \times \big(\mathbb{1} + f_3 M_3 + f_5 M_5 + f_3 f_3 M_3 M_3 + f_7 M_7 + f_3 f_5 M_3 M_5 + f_5 f_3 M_5 M_3 \\
&\qquad + f_9 M_9 + f_3 f_3 f_3 M_3 M_3 M_3 + f_5 f_5 M_5 M_5 + f_3 f_7 M_3 M_7 + f_7 f_3 M_7 M_3 \\
&\qquad + f_{11} M_{11} + f_3 f_3 f_5 M_3^2 M_5 + f_3 f_5 f_3 M_3 M_5 M_3 + f_5 f_3 f_3 M_5 M_3^2\big) + \mathcal{O}(\alpha'^{12}) \,.
\end{aligned} \tag{8.36}$$

This suggests the following all-order formula for the α'-expansion of n-point disk integrals [271],

$$\begin{aligned}
\phi(F^{\mathfrak{m}}) &= \bigg(\sum_{k=0}^{\infty} f_2^k P_{2k}\bigg) \sum_{r=0}^{\infty} \sum_{i_1,i_2,\ldots,i_r\in 2\mathbb{N}+1} f_{i_1} f_{i_2} \ldots f_{i_r} M_{i_1} M_{i_2} \ldots M_{i_r} \\
&= \bigg(\sum_{k=0}^{\infty} f_2^k P_{2k}\bigg) \frac{1}{1-\sum_{n=1}^{\infty} f_{2n+1} M_{2n+1}} \,,
\end{aligned} \tag{8.37}$$

where the fraction in the last line is understood as a geometric series $\frac{1}{1-x} = \sum_{m=0}^{\infty} x^m$. The only independent pieces of information in (8.37) are the matrices M_{2n+1} and P_{2k} along with f_{2n+1} and f_2^k. The coefficients of any product $f_2^k f_{2n+1}$ or higher-depth terms $f_{2n_1+1} f_{2n_2+1} \ldots$ are determined by (8.37) in terms of matrix multiplications among the P_{2k} and M_{2n+1}.

At multiplicities $n = 5, 6, 7$, (8.37) has been checked up to and including weight $21, 9, 7$ [166] and is conjectural beyond this.

8.3.3. Coaction

The coefficients P_w, M_w of ζ_w in (8.31) are defined w.r.t. a prescribed $\mathbb{Q}$-basis of MZVs at weight w. At weights $w \leq 7$, we have employed the unique bases in (8.32) that are expressible in terms of Riemann

[53] At the orders of $\alpha'^{w\leq 7}$, the rational prefactors in (8.32) remain unchanged in passing to $\phi(F^{\mathfrak{m}})$ apart from $\phi((\zeta_3^{\mathfrak{m}})^2) = f_3 ⧢ f_3 = 2 f_3 f_3$. Similarly, contributions of $\frac{1}{n!}(\zeta^{\mathfrak{m}}_{2k+1} M_{2k+1})^n$ at higher orders that resemble the expansion of a matrix-valued exponential are mapped to n-fold concatenation products $f^n_{2k+1} M^n_{2k+1}$ under ϕ.

zeta values. Starting from weight $w=8$, however, the choices of basis elements $\zeta_{3,5},\zeta_{3,7},\zeta_{3,3,5},\ldots$ beyond depth one in table 1 is somewhat arbitrary and leaves various equally natural alternatives. One could for instance change the basis to include $\zeta_{5,3}=\zeta_3\zeta_5-\zeta_{3,5}-\zeta_8$ instead of $\zeta_{3,5}$ which would add commutator terms $\sim[M_3,M_5]$ to the coefficient P_8 of $\zeta_2^4=\frac{175}{24}\zeta_8$ in the new basis. Similarly, trading $\zeta_{3,3,5}$ for a different basis elements at depth ≥ 3 leads to a shift of M_{11} by a rational multiple of $[M_3,[M_5,M_3]]$.

We shall illustrate the basis dependence of M_{11} by rewriting the α'^{11}-order in (8.33) in terms of $\zeta_{3,5,3}=-2\zeta_{3,3,5}+\frac{299}{2}\zeta_{11}+\zeta_3\zeta_{3,5}+\frac{8}{7}\zeta_2^3\zeta_5-\frac{12}{5}\zeta_2^2\zeta_7-90\zeta_2\zeta_9$ rather than $\zeta_{3,3,5}$,

$$\begin{aligned}F\,\big|_{(\alpha')^{11}}&=\zeta_{11}\bigg(M_{11}+\frac{299}{20}[M_3,[M_5,M_3]]\bigg)+\zeta_2\zeta_9P_2M_9+\zeta_2^2\zeta_7P_4M_7+\zeta_2^3\zeta_5P_6M_5+\zeta_2^4\zeta_3P_8M_3\\&\quad+\frac{1}{6}\zeta_2\zeta_3^3P_2M_3^3+\frac{1}{2}\zeta_3^2\zeta_5M_5M_3^2+\frac{1}{10}\zeta_{3,5}\zeta_3[M_5,M_3^2]-\frac{1}{10}\zeta_{3,5,3}[M_3,[M_5,M_3]]\,.\end{aligned}\tag{8.38}$$

The coefficient of ζ_{11} became $M_{11}+\frac{299}{20}[M_3,[M_5,M_3]]$ in the place of M_{11} in (8.33). Hence, the definition (8.31) of M_{2k+1} requires the specification of a (conjectural) $\mathbb{Q}$-basis of MZVs at weight $2k+1$, and we will follow the choices of the datamine [312] as done in [271]. It would be interesting if alternative choices of basis MZVs at higher weight may lead to similar shortenings as seen in the more compact form (8.38) of the α'^{11} order with $\zeta_{3,5,3}$ in the place of $\zeta_{3,3,5}$ in (8.33).

These ambiguities in the definition of $P_{\geq 8}$ and $M_{\geq 11}$ are also reflected by the freedom to add f_8 to $\phi(\zeta^{\mathfrak{m}}_{3,5})$ in (8.16) and more generally f_w to the ϕ-image of indecomposable weight-w MZVs of depth ≥ 2 in a $\mathbb{Q}$-basis. In other words, the isomorphism ϕ is non-canonical starting from weight 8. Nevertheless, the form of the all-weight result (8.37) is unaffected by the above choices.

The information of the all-order result (8.37) can be encoded in the following coaction formula without any reference to basis dependent quantities P_w,M_w [321],

$$\Delta(F^{\mathfrak{m}})_P{}^Q=\sum_{R\in S_{n-3}}(F^{\mathfrak{m}})_P{}^R\otimes(F^{\mathfrak{dr}})_R{}^Q\,.\tag{8.39}$$

Following the coaction of MZVs in [275], the MZVs in the second entry $F^{\mathfrak{dr}}$ are promoted to deRham periods $\zeta^{\mathfrak{m}}_{n_1,n_2,\ldots,n_r}\rightarrow\zeta^{\mathfrak{dr}}_{n_1,n_2,\ldots,n_r}$ (see for instance [315]) with a net effect of modding out by ζ_2 since $\zeta_2^{\mathfrak{dr}}=0$. One can straightforwardly verify (8.39) by insertion of (8.37) and using the simple form (8.13) of the deconcatenation coaction in the f-alphabet.

Based on $Z(1,P,n{-}1,n|1,Q,n,n{-}1)=-\sum_{R\in S_{n-3}}S^{-1}(Q|R)_1F_P{}^R$ with the KLT matrix $S(A|B)_1$ in (4.160), the coaction formula (8.39) can be readily translated to motivic and deRham versions $Z^{\mathfrak{m}},Z^{\mathfrak{dr}}$ of the Z-integrals (6.62). In the first place, one arrives at the coaction of the $(n{-}3)!\times(n{-}3)!$ basis of $Z(1,P,n{-}1,n|1,Q,n,n{-}1)$, but one can generalize to arbitrary $A,B\in S_n$ in

$$\Delta Z^{\mathfrak{m}}(A|B)=-\sum_{P,Q\in S_{n-3}}Z^{\mathfrak{m}}(A|1,P,n,n{-}1)S(P|Q)_1\otimes Z^{\mathfrak{dr}}(1,Q,n{-}1,n|B)\tag{8.40}$$

by noting that both sides of the equation obey the same monodromy relations in A and IBP relations in B.

Note that (8.39) and (8.40) are special cases of more general coaction formulae

$$\Delta\bigg(\int_\gamma\omega\bigg)^{\mathfrak{m}}=\sum_{j=1}^{d}\bigg(\int_\gamma\omega_j\bigg)^{\mathfrak{m}}\otimes\bigg(\int_{\gamma_j}\omega\bigg)^{\mathfrak{dr}}\tag{8.41}$$

for wider classes of integration cycles γ and differential forms ω that were studied in the context of Feynman integrals [323, 324, 325, 326] and Lauricella hypergeometric functions [327]. The sum over j in (8.41) runs over d-dimensional bases of twisted homologies $\{\gamma_j\}$ and cohomologies $\{\omega_j\}$, respectively. Moreover, these bases are understood to be chosen as orthonormal in the sense that the zero-transcendentality part of $\int_{\gamma_i}\omega_j$ is given by δ_{ij}. In our setting, the orthonormality condition is met by the Kronecker delta in the field-theory limit (6.52) of $F_P{}^Q$. The superscripts $\mathfrak{m}$ and $\mathfrak{dr}$ in (8.41) again refer to the motivic and deRham periods with $\zeta_2^{\mathfrak{dr}}=0$ in the second entry.

8.4. KK-like and BCJ relations within the α'-expansion

We have seen in section 7.4 that the monodromy relations obeyed by color-ordered open-string amplitudes deform the KK and BCJ relations of field-theory amplitudes by trigonometric functions in $\alpha' s_{ij}$. It will now be shown that certain sectors in the α'-expansions of disk integrals in section 8.3 preserve the field-theory BCJ relations of the SYM amplitudes they multiply. Other sectors in the α'-expansion of open-superstring amplitudes will be reviewed to obey analogues of KK relations where the coefficients are still integers independent on s_{ij}.

8.4.1. BCJ relations at all orders in α'

The organization (8.37) of the α'-expansion of ${F_P}^Q$ can be used to generate solutions of the BCJ relations (5.55) or (5.62) at arbitrary mass dimensions. This can be seen by inserting permutations of the string-amplitude formula (6.50) into the monodromy relations (7.75) and expanding in α'. It is crucial to note that the trigonometric factors yield series in even zeta values (8.10), i.e. exclusively the commutative generator f_2 in the f-alphabet upon passing to motivic MZVs and taking the ϕ-image,

$$\sin(\pi x) = \pi x \exp\left(- \sum_{k=1}^{\infty} \frac{\zeta_{2k}}{k} x^{2k} \right) . \tag{8.42}$$

Hence, the only departures of the monodromy relations from the BCJ relations occur for non-zero powers of f_2 – the appearance of the odd generators f_{2k+1} is unaffected by the sine functions.

In order to identify independent solutions of the BCJ relations, we impose (the motivic version of) the monodromy relations separately for the coefficient of any $f_2^k f_{i_1} f_{i_2} \ldots f_{i_r}$ with $k, r \in \mathbb{N}_0$ and $i_j \in 2\mathbb{N}{+}1$. The separation of different transcendentality structures has been firstly applied in [136] to demonstrate the KK and BCJ relations of the tree-level matrix elements of the $\alpha'^3 \zeta_3 \mathrm{Tr}\{D^2\mathbb{F}^4 + \mathbb{F}^5\}$ operator in the superstring effective action and the $\alpha' \mathrm{Tr}\{\mathbb{F}^3\}$ operator of the open bosonic string. By focusing on the $f_2 \to 0$ part of the monodromy relations, the coefficient of any $f_{i_1} f_{i_2} \ldots f_{i_r}$ in color-ordered open-string amplitude is found to obey KK and BCJ relations,

$$\begin{aligned} 0 &= \mathcal{A}_{i_1,i_2,\ldots,i_r}(P \sqcup\!\sqcup Q, n) \, , \qquad P, Q \neq \emptyset \, , \\ 0 &= \sum_{j=2}^{n-1} (k_{p_1} \cdot k_{p_2 p_3 \ldots p_j}) \mathcal{A}_{i_1,i_2,\ldots,i_r}(p_2 p_3 \ldots p_j p_1 p_{j+1} \ldots p_n) \, , \end{aligned} \tag{8.43}$$

where $\mathcal{A}_{i_1,i_2,\ldots,i_r}(P)$ is a shorthand for the coefficient of $f_{i_1} f_{i_2} \ldots f_{i_r}$ in the ϕ-image of the motivic versions $\mathcal{A}^{\mathfrak{m}}$ of superstring amplitudes $\mathcal{A}$, i.e.

$$\begin{aligned} \mathcal{A}_{i_1,i_2,\ldots,i_r}(1, Q, n{-}1, n) &:= \phi\big(\mathcal{A}^{\mathfrak{m}}(1, Q, n{-}1, n)\big) \big|_{f_{i_1} f_{i_2} \ldots f_{i_r}} \\ &= \sum_{R \in S_{n-3}} (M_{i_1} M_{i_2} \ldots M_{i_r})_Q{}^R A(1, R, n{-}1, n) \, . \end{aligned} \tag{8.44}$$

By isolating the coefficients of $f_{i_1} f_{i_2} \ldots f_{i_r}$ in last three lines of (8.36), we for instance arrive at the simplest independent solutions to the BCJ relations at the orders of $\alpha'^{\leq 11}$ in table 2. The table only tracks the solutions of BCJ relations that are realized in the α'-expansion of superstring disk amplitudes – a variety of further solutions multilinear in polarization vectors can be systematically generated from tree-level amplitudes of bosonic or heterotic strings [210, 211] or from building blocks of loop-level string amplitudes [168]. Similarly, as will be detailed in section 8.4.4, the α'-expansion of $Z(P|Q)$ integrals can be used to generate rational functions in s_{ij} at various mass dimensions that obey BCJ relations in Q.

Note that permutations $\mathcal{A}_{i_1,i_2,\ldots,i_r}(P)$ outside the $(n{-}3)!$-element basis of $\mathcal{A}_{i_1,i_2,\ldots,i_r}(1, Q, n{-}1, n)$ in (8.44) can be expanded via

$$\mathcal{A}_{i_1,i_2,\ldots,i_r}(P) = - \sum_{Q, R \in S_{n-3}} m(P|1, R, n, n{-}1) S(R|Q)_1 \mathcal{A}_{i_1,i_2,\ldots,i_r}(1, Q, n{-}1, n) \, , \tag{8.45}$$

i.e. by adapting the solutions (7.48) of BCJ relations to $\mathcal{A}_{i_1,i_2,\ldots,i_r}(P)$ in the place of SYM amplitudes.

w	#(solutions)	w	#(solutions)
0	$A = \mathcal{A}_\emptyset$	6	$\mathcal{A}_{3,3}$
1	$\times$	7	$\mathcal{A}_7$
2	$\times$	8	$\mathcal{A}_{3,5},\ \mathcal{A}_{5,3}$
3	$\mathcal{A}_3$	9	$\mathcal{A}_9,\ \mathcal{A}_{3,3,3}$
4	$\times$	10	$\mathcal{A}_{3,7},\ \mathcal{A}_{7,3},\ \mathcal{A}_{5,5}$
5	$\mathcal{A}_5$	11	$\mathcal{A}_{11},\ \mathcal{A}_{3,3,5},\ \mathcal{A}_{3,5,3},\ \mathcal{A}_{5,3,3}$

Table 2: The solutions $\mathcal{A}_{i_1,i_2,\ldots,i_r}$ of BCJ relations at the order of $\alpha'^{\leq 11}$ that can be read off from the coefficient of $f_{i_1} f_{i_2} \ldots f_{i_r}$ in the α'-expansion of open-superstring amplitudes.

8.4.2. KK-like relations

Inspired by the Kleiss–Kuijf (KK) relations (5.8) among tree-level amplitudes in field theories, we shall now investigate the α'-expansion of disk amplitudes in (8.37) for identities with constant coefficients. More precisely, we shall go beyond the coefficients (8.44) of the $f_B = f_{b_1} f_{b_2} \ldots f_{b_r}$ with odd b_j and identify KK-like relations among

$$\begin{aligned} \mathcal{A}_{\ell|B}(1,Q,n{-}1,n) &= \phi\big(\mathcal{A}^{\mathfrak{m}}(1,Q,n{-}1,n)\big)\,\big|_{f_2^\ell f_{b_1} f_{b_2} \ldots f_{b_r}} \\ &= \sum_{R \in S_{n-3}} (P_{2\ell} M_{b_1} M_{b_2} \ldots M_{b_r})_Q{}^R A(1,R,n{-}1,n) \end{aligned} \tag{8.46}$$

associated with arbitrary powers $\ell \geq 0$ of f_2, following version 3 of [138]. For instance, the well-known cyclic and reflection properties of disk amplitudes hold separately along with each $f_2^\ell f_B$,

$$\begin{aligned} \mathcal{A}_{\ell|B}(1,2,\ldots,n) &= \mathcal{A}_{\ell|B}(2,3,\ldots,n,1)\,, \\ \mathcal{A}_{\ell|B}(1,2,\ldots,n) &= (-1)^n \mathcal{A}_{\ell|B}(n,n{-}1,\ldots,2,1)\,, \end{aligned} \tag{8.47}$$

leaving at most $\frac{1}{2}(n{-}1)!$ independent permutations. However, the coefficients (8.46) of $f_2^\ell f_B$ at different values of ℓ obey different additional relations, so we will focus on the individual components. To this effect, following [329], we will refer to relations of the form

$$\sum_\sigma c_\sigma \mathcal{A}_{\ell|B}(\sigma) = 0\,, \tag{8.48}$$

with constant coefficients $c_\sigma \in \mathbb{Q}$ as *KK-like* and study them separately at each $\ell \geq 0$. For simple examples of KK-like relations, we have the permutation symmetry $\mathcal{A}_{1|B}(1,2,3,4) = \mathcal{A}_{1|B}(1,2,4,3)$ as well as the six-term identity [163]

$$\mathcal{A}_{1|B}(1,2,3,4,5) + \mathrm{perm}(2,4,5) = 0 \tag{8.49}$$

universal to the coefficients of $f_2 f_B$. These KK-like relations of $\mathcal{A}_{1|B}$ clearly differ from the KK relations (8.43) of $\mathcal{A}_{0|B}$ such as the three-term identity $\mathcal{A}_{0|B}(1,2,3,4)+\mathrm{cyc}(2,3,4) = 0$ at four points or the four-term identity $\mathcal{A}_{0|B}(1,2 \sqcup\!\sqcup 345) = 0$ at five points.

Note that the simplest examples of $\mathcal{A}_{1|B}$ up to and including the order of α'^{10} do not require an f-alphabet description and can be equivalently obtained from the coefficient of ζ_2 or products $\zeta_2\zeta_{2k+1}$, $\zeta_2\zeta_{2k_1+1}\zeta_{2k_2+1}$, $\zeta_2\zeta_{3,5}$ in (8.32) and (8.33). Starting from the order of α'^{11} with the MZV basis choice of (8.33), the coefficients of $\zeta_2\zeta_{2k+1}$ in the matrix F generically receive admixtures of products $M_{i_1} M_{i_2} \ldots$ with odd i_j on top of the expected $P_2 M_{2k+1}$ in (8.46). As illustrated by the coefficient $F\,|_{\zeta_2\zeta_9} = P_2 M_9 + 9[M_3,[M_5,M_3]]$ in (8.33), passing to the f-alphabet is necessary to isolate the matrix product $\phi(F^{\mathfrak{m}})\,|_{f_2 f_9} = P_2 M_9$ in (8.35). The five-point KK-like relation (8.49) only holds if $\mathcal{A}_{1|9}$ is constructed from $P_2 M_{2k+1}$ in (8.46), i.e. defined by $\mathcal{A}_{1|9} = \phi(\mathcal{A}^{\mathfrak{m}})\,|_{f_2 f_9}$, but fails in presence of extra terms $\sim [M_3,[M_5,M_3]]$ that would arise from $\mathcal{A}\,|_{\zeta_2\zeta_9}$.[54]

[54]We would like to thank Ricardo Medina for email correspondence on this point.

Note, however, that different MZV basis choices may push this issue to higher orders ${\alpha'}^{>11}$, as evidenced by the expansion (8.38) in which $F|_{\zeta_2\zeta_9} = P_2 M_9$.

8.4.3. Berends–Giele idempotents and BRST-invariant permutations

In order to write down the explicit form of KK-like amplitude relations, we need to specify a way to generate permutations with the correct properties. As discussed in [138], the relevant permutations achieving this are related to the *descent algebra* of permutations via the so-called *BRST-invariant permutations*[55] $\gamma_{1|P_1,P_2,\ldots,P_k}$ depending on a number k of words $P_1,\ldots,P_k$

$$\gamma_{1|P_1,\ldots,P_k} = 1\big(\mathcal{E}(P_1) \sqcup\!\sqcup\, \mathcal{E}(P_2) \sqcup\!\sqcup \ldots \sqcup\!\sqcup\, \mathcal{E}(P_k)\big)\,, \tag{8.50}$$

where $\mathcal{E}(P)$ is the *Berends–Giele idempotent* defined in terms of the right-action multiplication $P \circ Q$ of permutations by [138]

$$\mathcal{E}(P) := P \circ \mathcal{E}_n\,, \quad |P| = n\,, \tag{8.51}$$

where[56]

$$\mathcal{E}_n = \sum_{\sigma\in S_n} \kappa_{\sigma^{-1}}\sigma\,, \qquad \kappa_\sigma = \frac{(-1)^{d_\sigma}}{|\sigma|\binom{|\sigma|-1}{d_\sigma}}\,, \tag{8.52}$$

and d_σ denotes the descent number of the permutation σ. Moreover, they were shown to satisfy the shuffle relations

$$\mathcal{E}(R \sqcup\!\sqcup\, S) = 0\,, \quad R, S \neq \emptyset\,. \tag{8.53}$$

This implies that the number of linearly independent BRST-invariant permutations at n points is given by

$$\#(\gamma_{1|P_1,\ldots,P_k}) = \begin{bmatrix} n-1 \\ k \end{bmatrix}, \quad \sum_{i=1}^{k} |P_i| = n-1\,, \tag{8.54}$$

where $\left[{p \atop q}\right]$ denotes the Stirling cycle numbers [333, 334] (traditionally called Stirling numbers of the first kind) that count the number of ways to arrange p objects into q cycles. For example, $\left[{5 \atop q}\right] = 24, 50, 35, 10, 1$ for $q = 1, 2, 3, 4, 5$. For example permutations of the above definitions, see the Appendix I.

KK-like amplitude relations. The KK-like relations among the component amplitudes of (8.46) were observed to satisfy the following decomposition according to the number of parts k in the partitions of $n-1$ legs [138]

$$\begin{aligned}
&\mathcal{A}_{0|B}(\gamma_{1|P_1,\ldots,P_k}) = 0\,, \quad k \neq 1\,,\\
&\mathcal{A}_{1|B}(\gamma_{1|P_1,\ldots,P_k}) = 0\,, \quad k \neq 3\,,\\
&\mathcal{A}_{\ell|B}(\gamma_{1|P_1,\ldots,P_k}) = 0\,, \quad k \neq 1, 3, 5, \ldots, 2\ell{+}1\,, \quad \ell \geq 2\,.
\end{aligned} \tag{8.55}$$

In addition, it was demonstrated in [138] that the even cases when $k = 2m$ encode the parity and cyclicity relations (8.47). In this sense, the KK-like relations for even k are equivalent to (8.47).

For example, the case $k = 3$ for $n = 5$ with $\gamma_{1|23,4,5}$ given in (I.3) leads to the 12-term relation after using (8.47):

$$\begin{aligned}
&\mathcal{A}_{0|B}(1,2,3,4,5) + \mathcal{A}_{0|B}(1,2,3,5,4) + \mathcal{A}_{0|B}(1,2,4,3,5) + \mathcal{A}_{0|B}(1,2,4,5,3)\\
&+ \mathcal{A}_{0|B}(1,2,5,3,4) + \mathcal{A}_{0|B}(1,2,5,4,3) - \mathcal{A}_{0|B}(1,3,2,4,5) - \mathcal{A}_{0|B}(1,3,2,5,4)\\
&- \mathcal{A}_{0|B}(1,3,4,2,5) - \mathcal{A}_{0|B}(1,3,5,2,4) + \mathcal{A}_{0|B}(1,4,2,3,5) - \mathcal{A}_{0|B}(1,4,3,2,5) = 0\,,
\end{aligned} \tag{8.56}$$

[55] The terminology of "BRST-invariant permutations" was coined in [138] by the analogy of $\gamma_{1|P_1,\ldots,P_k}$ in (8.50) with certain BRST invariants in the pure spinor computation of one-loop amplitudes, but that we are not claiming that BRST transformations act on permutations.

[56] In fact, $\mathcal{E}_n$ is given by the inverse permutations of the Eulerian or Solomon idempotent [330, 331, 332].

which can be reduced to linear combinations of the KK relations (8.43).

In addition, given that the BRST-invariant permutations constitute a basis for permutations in the descent algebra[57], any other KK-like relation can be written as a linear combination of $\gamma_{1|P_1,\ldots,P_k}$. For instance, using the decomposition

$$W_{12345} + \mathrm{perm}(2,4,5) = 3\gamma_{1|2345} - \frac{1}{2}\gamma_{1|345,2} + \frac{1}{2}\gamma_{1|4,235} + \frac{1}{8}\gamma_{1|5,4,3,2} + (4\leftrightarrow 5)\,, \tag{8.57}$$

where a permutation σ is written as W_σ for typographical convenience, it follows that the KK-like relation (8.49) can be rewritten as

$$3\mathcal{A}_{1|B}(\gamma_{1|2345}) - \frac{1}{2}\mathcal{A}_{1|B}(\gamma_{1|345,2}) + \frac{1}{2}\mathcal{A}_{1|B}(\gamma_{1|4,235}) + \frac{1}{8}\mathcal{A}_{1|B}(\gamma_{1|5,4,3,2}) + (4\leftrightarrow 5) = 0\,, \tag{8.58}$$

or, equivalently after using the reflection relation (8.47), $3\mathcal{A}_{1|B}(\gamma_{1|2345}) + 3\mathcal{A}_{1|B}(\gamma_{1|2354}) = 0$. Note the crucial absence of $\gamma_{1|P_1,\ldots,P_k}$ with $k=3$ in the decomposition (8.57), which provides a consistency check of (8.55). For another example, one can check that the following 720-term (or 360 after using (8.47)) linear combination vanishes, $\mathcal{A}_{1|B}(\gamma_{1|2,3,4,5,67}) = 0$, in agreement with the second line of (8.55) with $k=5$.

Basis dimensions. Using the counting (8.54) one can show that the number of linearly independent amplitudes under the KK-like relations is given by [138]

$$\begin{aligned}
\#\big(\mathcal{A}_{0|B}(1,2,\ldots,n)\big) &= \begin{bmatrix} n{-}1 \\ 1 \end{bmatrix} = (n{-}2)!\,, \\
\#\big(\mathcal{A}_{1|B}(1,2,\ldots,n)\big) &= \begin{bmatrix} n{-}1 \\ 3 \end{bmatrix}, \\
\#\big(\mathcal{A}_{\ell|B}(1,2,\ldots,n)\big) &= \begin{bmatrix} n{-}1 \\ 1 \end{bmatrix} + \begin{bmatrix} n{-}1 \\ 3 \end{bmatrix} + \cdots + \begin{bmatrix} n{-}1 \\ 2\ell{+}1 \end{bmatrix}, \quad \ell \geq 2\,,
\end{aligned} \tag{8.59}$$

with the implicit assumption that $\left[\begin{smallmatrix} p \\ q \end{smallmatrix}\right] = 0$ for $q{>}p$. The counting of independent permutations of the amplitudes $\mathcal{A}_{1|B}(1,2,\ldots,n)$ associated with $f_2 f_B$ yields $\left[\begin{smallmatrix} n-1 \\ 3 \end{smallmatrix}\right] = 1, 6, 35, 225, \ldots$ at $n = 4,5,6,7,\ldots$ and has been studied in [329, 163].

A consistency check on the claim that the cyclicity and reflection symmetries (8.47) are encoded in the even-k BRST-invariant permutations follows from the counting (8.54) as $\sum_{k \text{ even}}^{n-1} \#\big(\gamma_{1|P_1,\ldots,P_k}\big) = \frac{1}{2}(n{-}1)!$. To see this, we note the elementary identity $\sum_{k \text{ even}}^{n-1} \left[\begin{smallmatrix} n-1 \\ k \end{smallmatrix}\right] = \frac{1}{2}(n{-}1)!$ of Stirling cycle numbers. It furthermore follows from (8.59) that those coefficients $\mathcal{A}_{\ell|B}(1,\ldots,n)$ with $\ell \geq 2$ and $n \leq 2\ell{+}3$ obey no additional KK-like relations other than the cyclicity and reflection symmetry (8.47): the counting of (8.59) yields $\sum_{k \text{ odd}}^{n-1} \left[\begin{smallmatrix} n-1 \\ k \end{smallmatrix}\right] = \frac{1}{2}(n{-}1)!$ independent permutations in these cases. For instance, when $\ell = 2$ the number of linearly independent permutations of $\mathcal{A}_{\ell|B}(1,2,\ldots,n)$ w.r.t. KK-like relations is $\frac{1}{2}(n{-}1)! = 360$ for $n = 7$ but $2519 = \frac{1}{2}(n{-}1)!{-}1$ for $n = 8$. This last prediction has been confirmed by a brute-force search using [152].

8.4.4. BCJ and KK relations of Z-theory amplitudes

The BCJ and KK-like relations in specific sectors of the α'-expansion of string amplitudes can be traced back to analogous relations for the disk integrals $Z(P|Q)$ in (6.62). As before, the discussion hinges on the f-alphabet description of the α'-expansion and the underlying motivic MZVs, and we employ the notation

$$\mathfrak{Z}(P|Q) = \phi\big(Z^{\mathfrak{m}}(P|Q)\big)\,, \qquad \mathfrak{Z}_\times(Q) = \phi\big(Z^{\mathfrak{m}}_\times(Q)\big) = \sum_{P \in S_{n-1}} \mathfrak{Z}(P,n|Q) \tag{8.60}$$

for the ϕ-image of motivic Z-theory amplitudes $Z^{\mathfrak{m}}(P|Q) = Z(P|Q)\,|_{\zeta_{n_1,\ldots} \to \zeta^{\mathfrak{m}}_{n_1,\ldots}}$ to avoid cluttering.

[57] This claim follows from the conjectural relation between the BRST-invariant permutations and the inverse of the *idempotent basis* [140] of the descent algebra. See [138] for more details.

As shown in section 7.4, tree-level amplitudes of the NLSM of Goldstone bosons can be obtained from the symmetrized versions $Z_\times(Q)$ of the $Z(P|Q)$-integrals defined in (7.91). The color-ordering Q of the NLSM amplitude (7.92) is encoded in the Parke–Taylor integrand $\mathrm{PT}(Q)$ of the symmetrized integral $Z_\times(Q)$. Hence, KK and BCJ relations of the NLSM are a simple consequence of partial-fraction and IBP relations of Parke–Taylor integrals, see section 6.4.3.

This worldsheet derivation of amplitudes relations of the NLSM is actually not tied to the low-energy limit in (7.92) since the KK and BCJ relations of $Z(P|Q)$ w.r.t. the Parke–Taylor orderings Q are valid at all orders in α'. In particular, KK and BCJ relations apply to every $\mathbb{Q}$-independent combination of MZVs in the α'-expansion of abelian Z-integrals. After peeling off the leading power of $(\pi\alpha')^{n-2}$ exposed by the sine functions in (7.95), we expect all combinations $f_2^\ell f_{i_1} f_{i_2}\ldots$ with $\ell \in \mathbb{N}_0$ and odd letters i_j to appear in the α'-expansion of $Z_\times(Q)$ at sufficiently large multiplicity $|Q|$.[58]

Any combination of $f_2^\ell f_{i_1} f_{i_2}\ldots$ in the α'-expansion of $Z_\times(Q)$ can be interpreted as an effective interaction among scalars with one color degree of freedom that preserves the KK and BCJ relation of the NLSM. By uniform transcendentality of disk integrals, the w^{th} subleading order of α' features MZVs of weight w each of which signals scalar interactions with $2w$ additional derivatives beyond the NLSM. The subleading order $\sim \zeta_2(\pi\alpha')^{n-2}$ of $Z_\times(Q)$ for instance defines a four-derivative deformation of the NLSM that preserves its amplitude relations and can also be described through the Lagrangian in section 3.3 of [335].

More generally, the f-alphabet images $\mathfrak{Z}_\times(Q)$ of symmetrized (motivic) disk integrals in (8.60) can be taken as generating functions of scalar effective-field-theory amplitudes subject to KK and BCJ relations,

$$\mathfrak{Z}_\times(A \shuffle B, n)\big|_{f_2^\ell f_{i_1} f_{i_2}\ldots} = \mathfrak{Z}_\times(\{A,B\},n)\big|_{f_2^\ell f_{i_1} f_{i_2}\ldots} = 0 \quad \forall\, A,B \neq \emptyset\,, \quad \ell \geq 0\,, \quad i_j \in 2\mathbb{N}{+}1\,. \tag{8.61}$$

The same type of reasoning applies to the Parke–Taylor orderings Q of *non-abelian* Z-integrals and their ϕ-images $\mathfrak{Z}(P|Q)$ in (8.60): each combination of MZVs in the α'-expansion corresponds to effective interactions of *bi-colored* scalars that preserve the KK- and BCJ relations of bi-adjoint scalars in Q,

$$\mathfrak{Z}(P|A \shuffle B, n)\big|_{f_2^\ell f_{i_1} f_{i_2}\ldots} = \mathfrak{Z}(P|\{A,B\},n)\big|_{f_2^\ell f_{i_1} f_{i_2}\ldots} = 0 \quad \forall\, A,B \neq \emptyset\,, \quad \ell \geq 0\,, \quad i_j \in 2\mathbb{N}{+}1\,. \tag{8.62}$$

For the amplitude relations that vary P at fixed Q in turn, the reasoning in section 8.4.1 implies that only the $f_2 \to 0$ sector of the α'-expansion preserves KK- and BCJ relations. These field-theory relations of $\mathfrak{Z}(P|Q)$ at fixed Q then hold independently for the coefficients of any $f_{i_1} f_{i_2}\ldots$ with odd i_j,

$$\mathfrak{Z}(A \shuffle B, n|Q)\big|_{f_{i_1} f_{i_2}\ldots} = \mathfrak{Z}(\{A,B\},n|Q)\big|_{f_{i_1} f_{i_2}\ldots} = 0 \quad \forall\, A,B \neq \emptyset\,, \quad i_j \in 2\mathbb{N}{+}1\,. \tag{8.63}$$

For the coefficient of $f_2^\ell f_{i_1} f_{i_2}\ldots$ in $\mathfrak{Z}(P|Q)$ at $\ell \geq 1$, we obtain bi-colored scalar amplitudes subject to the KK-like relations of section 8.4.2 in P.

Based on a Berends–Giele recursion for the α'-expansion of Z-integrals, a proposal for the non-linear equations of motion of the underlying non-abelian Z-theory can be found in [225] and section 8.6. Explicit results up to and including the order of α'^7 are publicly available from the website [336].

8.5. String corrections from the Drinfeld associator

We shall now review a recursive all-multiplicity method to determine the polynomial structure of the α'-expansion of the disk integrals $F_P{}^Q$ in (6.51). This methods generates all the MZVs from the Drinfeld associator (see section 8.2.2) whose non-commutative variables e_0, e_1 are identified with specific matrices whose entries are linear in $\alpha' s_{ij}$. More specifically, the recursive step in passing from $n{-}1$ to n points [269],

$$F^{\sigma_i} = \sum_{j=1}^{(n-3)!} \big[\Phi(e_0,e_1)\big]_{ij} \big(F^{\sigma_j}\big|_{k_{n-1}=0}\big)\,, \tag{8.64}$$

[58]Since the four-point α'-expansion (8.3) is expressible in terms of Riemann zeta values only, the onset of irreducible MZVs $\zeta_{3,5}, \zeta_{3,7}, \zeta_{3,3,5}, \ldots$ at higher depth is relegated to $Z_\times(Q)$ at multiplicities $|Q| \geq 6$.

is based on $(n{-}2)!\times(n{-}2)!$ matrices e_0, e_1 whose derivation will be described below. The F^{σ_i} are understood to be the $F_P{}^Q$ with $P=23\ldots n{-}2$ the canonical ordering and Q the $i^{\rm th}$ permutation σ_i of $2,3,\ldots,n{-}2$ in lexicographical ordering. The Drinfeld associator Φ is understood to be expanded in terms of shuffle-regularized MZVs as in (8.19). The soft limit on the right-hand side of (8.64) acts recursively in the sense that

$$F^{\sigma(23\ldots n-2)}\big|_{k_{n-1}=0} = \left\{ \begin{array}{cl} F^{\sigma(23\ldots n-3)} & : \ \sigma(n{-}2)=n{-}2\,, \\ 0 & : \ \text{otherwise}\,, \end{array}\right. \tag{8.65}$$

which terminates with the three-point integral $F^{\emptyset}=1$. The relevance of the Drinfeld associator for string amplitudes was firstly pointed out in [321], among other things by relating its coaction properties with those of the $F_P{}^Q$ in (8.39). Nevertheless, it is an open problem to deduce (8.39) from the results of this section. The specific construction towards the α'-expansion of the $F_P{}^Q$ was given in [269] and is based on an expansion method for Selberg integrals from the mathematics literature [274]. Its description in terms of twisted deRham theory and intersection numbers of twisted forms can be found in [337], where e_0, e_1 are identified as braid matrices.

8.5.1. Construction of the matrices e_0, e_1

The recursion (8.64) can be derived from the deformation

$$\begin{aligned} \hat F^\sigma_\nu &:= (2\alpha')^{n-3} \int\limits_{0<z_2<z_3<\ldots<z_{n-2}<z_0} dz_2\, dz_3\, \ldots\, dz_{n-2} \prod_{1\leq p<q}^{n-1} |z_{pq}|^{-2\alpha' s_{pq}} \prod_{r=2}^{n-2} |z_{0r}|^{-2\alpha' s_{0r}} \omega^\sigma_\nu\,, \\ \omega^\sigma_\nu &= \sigma\Big\{ \prod_{k=2}^{\nu} \sum_{j=1}^{k-1} \frac{s_{jk}}{z_{jk}} \prod_{m=\nu+1}^{n-2} \sum_{n=m+1}^{n-1} \frac{s_{mn}}{z_{mn}} \Big\} \end{aligned} \tag{8.66}$$

of the disk integrals $F^\sigma = F_{23\ldots n-2}{}^{\sigma(23\ldots n-2)}$ in (6.51) by additional Mandelstam invariants s_{0j} and an auxiliary puncture $z_0\in(0,1)$ on the disk boundary. The permutation σ acts on the labels of both s_{ij} and z_{ij} enclosed in the curly brackets of the second line while leaving $\sigma(1)=1$ and $\sigma(n{-}1)=n{-}1$ invariant. One can recover (8.66) from a basis of disk integrals in the $(n{+}1)$-point open-string amplitudes (6.50) after removing the integration over $z_0\in(0,1)$ and the associated dz_0/z_{0j}.

The integer $\nu=1,2,\ldots,n{-}2$ in (8.66) labels different classes of integrands ω^σ_ν that were related by the IBP identities from the Koba–Nielsen factor in the undeformed case (6.51). By the contributions $|z_{0r}|^{-2\alpha' s_{0r}}$ to the Koba–Nielsen factor in (8.66), the $n{-}2$ values of ν together with the $(n{-}3)!$ permutations σ of $2,3,\ldots,n{-}2$ yield a total of $(n{-}2)!$ different integrals $\hat F^\sigma_\nu$. The components of this $(n{-}2)!$-vector will be ordered as $\hat F=(\hat F^\sigma_{n-2},\hat F^\sigma_{n-3},\ldots,\hat F^\sigma_2,\hat F^\sigma_1)$ with lexicographic ordering for the permutations σ indexing the $(n{-}3)!$-component subvectors $\hat F^\sigma_1,\ldots,\hat F^\sigma_{n-2}$. The examples of (8.66) at $n=4$ and 5 points are the two- and six-component vectors

$$\begin{aligned} \hat F\,\big|_{n=4} &= 2\alpha' \int_0^{z_0} dz_2\, |z_{12}|^{-2\alpha' s_{12}} |z_{23}|^{-2\alpha' s_{23}} |z_{02}|^{-2\alpha' s_{02}} \begin{pmatrix} X_{12} \\ X_{23} \end{pmatrix}\,, \\ \hat F\,\big|_{n=5} &= (2\alpha')^2 \int_0^{z_0} dz_3 \int_0^{z_3} dz_2\, |z_{23}|^{-2\alpha' s_{23}} \prod_{j=2}^{3} |z_{1j}|^{-2\alpha' s_{1j}} |z_{j4}|^{-2\alpha' s_{j4}} |z_{0j}|^{-2\alpha' s_{0j}} \begin{pmatrix} X_{12}(X_{13}{+}X_{23}) \\ X_{13}(X_{12}{+}X_{32}) \\ X_{12}X_{34} \\ X_{13}X_{24} \\ (X_{23}{+}X_{24})X_{34} \\ (X_{32}{+}X_{34})X_{24} \end{pmatrix}\,, \end{aligned} \tag{8.67}$$

with the shorthand $X_{ij}=\frac{s_{ij}}{z_{ij}}$ as in section 6.3. Since the entries of the n-point vectors $\hat F$ form IBP bases, their z_0-derivatives are bound to yield homogeneous Knizhnik–Zamolodchikov (KZ) equations of the following form

$$\frac{d}{dz_0}\hat F = \left(\frac{\hat e_0}{z_0}+\frac{\hat e_1}{1{-}z_0}\right)\hat F\,. \tag{8.68}$$

The entries of the $(n{-}2)!\times(n{-}2)!$ braid matrices $\hat e_0,\hat e_1$ are linear in $\alpha' s_{ij}$ with $0\leq i<j\leq n{-}1$ as can be seen from the z_0-derivative of the deformed Koba–Nielsen factor in (8.66). The same factors of $|z_{0r}|^{-2\alpha' s_{0r}}$

suppress the boundary terms $z_{n-2}=z_0$ from $\frac{d}{dz_0}$-action on the integration limits for $z_{n-2}\in(z_{n-3},z_0)$. The explicit form of $\hat{e}_0,\hat{e}_1$ follows from reducing the contributions $\sum_{r=2}^{n-2}\frac{s_{0r}}{z_{0r}}\omega_\nu^\sigma$ of the Koba–Nielsen derivatives to a basis of ω_ν^σ/z_0 and $\omega_\nu^\sigma/(1-z_0)$.

In fact, the recursion (8.64) only requires the kinematic limit

$$e_0=\hat{e}_0\,\big|_{s_{0j}=0}\,,\qquad e_1=\hat{e}_1\,\big|_{s_{0j}=0} \tag{8.69}$$

of the braid matrices $\hat{e}_0,\hat{e}_1$ in (8.68). The four- and five-point integrals (8.67) give rise to the following 2×2 and 6×6 examples:

$$e_0\,\big|_{n=4}=2\alpha'\begin{pmatrix}-s_{12}&s_{12}\\0&0\end{pmatrix}\,,\qquad e_1\,\big|_{n=4}=2\alpha'\begin{pmatrix}0&0\\-s_{23}&s_{23}\end{pmatrix}\,, \tag{8.70}$$

$$e_0\,\big|_{n=5}=2\alpha'\begin{pmatrix}-s_{123}&0&s_{13}{+}s_{23}&s_{12}&s_{12}&-s_{12}\\0&-s_{123}&s_{13}&s_{12}{+}s_{23}&-s_{13}&s_{13}\\0&0&-s_{12}&0&s_{12}&0\\0&0&0&-s_{13}&0&s_{13}\\0&0&0&0&0&0\\0&0&0&0&0&0\end{pmatrix}\,,$$

$$e_1\,\big|_{n=5}=2\alpha'\begin{pmatrix}0&0&0&0&0&0\\0&0&0&0&0&0\\-s_{34}&0&s_{34}&0&0&0\\0&-s_{24}&0&s_{24}&0&0\\-s_{34}&s_{34}&-s_{23}{-}s_{24}&-s_{34}&s_{234}&0\\s_{24}&-s_{24}&-s_{24}&-s_{23}{-}s_{34}&0&s_{234}\end{pmatrix}\,.$$

The explicit form of the braid matrices e_0,e_1 at $n\leq 9$ points is available in machine-readable form [328] (where our conventions are matched after rescaling $s_{ij}\rightarrow -2\alpha' s_{ij}$ in the dataset of the website), and a graphical all-multiplicity description can be found in [337].

8.5.2. Uniform transcendentality

The factorization of α' in the braid matrices e_0,e_1 as exemplified by (8.70) persists to any multiplicity n. On these grounds, the $(n{-}2)!\times(n{-}2)!$ matrix representations $\Phi(e_0,e_1)$ of the Drinfeld associator enjoy uniform transcendentality: the words e_A in e_0,e_1 of length $|A|=w$ are of the order $(\alpha')^w$ and accompanied by MZVs $I(0;A;1)$ of transcendental weight w in (8.19). It is then easy to show by induction that the recursion (8.64) propagates uniform transcendentality from the $(n{-}1)$-point integrals $F^{\sigma_j}|_{k_{n-1}=0}$ on the right-hand side to the n-point integrals F^{σ_i} on the left-hand side.

Together with the α'-independent SYM amplitudes $A(\ldots)$ in (6.50), we conclude that n-point open-superstring amplitudes are uniformly transcendental. The KLT relations (7.58) together with the uniform transcendentality of the KLT kernel in (7.56) imply that also type II amplitudes are uniformly transcendental. However, the α'-dependence of the kinematic factors $A_{(DF)^2+\mathrm{YM}+\phi^3}(\ldots)$ in (7.119) obstructs uniform transcendentality of massless heterotic-string amplitudes (except for single-trace gauge amplitudes with no or one graviton in (7.133) and (7.135) [137, 259]). Massless amplitudes of open or closed bosonic strings are in general non-uniformly transcendental for the same reason [210, 211].

The factorization of α' on the right-hand side of the KZ equation (8.68) can be viewed as a string-theory analogue of the so-called ϵ-form of differential equations of Feynman integrals [338, 339]: The dimensional-regularization parameter ϵ of Feynman integrals in $D_0{-}2\epsilon$ spacetime dimensions with $D_0\in\mathbb{N}$ serves as an expansion variable similar to α' in string amplitudes. Various families of Feynman integrals admit uniformly transcendental bases under IBP relations [340, 341, 342], also see [343, 344] and [345, 346, 347] for white papers and recent reviews. In a growing number of examples, uniform transcendentality can be manifested by casting the differential equations of vectors I of Feynman integrals into ϵ-form $dI=\epsilon AI$, where the matrix A of one forms no longer depends on ϵ [338, 339].

8.5.3. *Regularized boundary values*

Given the braid matrices e_0, e_1 derived from the KZ equation (8.68) at $s_{0j} = 0$, we shall now review the origin of the recursion (8.64) for the α'-expansion of disk integrals. The key idea is to use the relation (8.21) between the regularized boundary values C_0, C_1 of the solutions to a general KZ equation. For the $(n-2)!$-component vector $\hat{F}$ in (8.66), the regularized boundary value as $z_0 \to 0$ is given by [269]

$$C_0 \big|_{s_{0j}=0} \to \big(F^\sigma \big|_{s_{j,n-1}=0}, \underbrace{0, 0, \ldots, 0}_{(n-3)(n-3)!} \big) \, , \tag{8.71}$$

where the $(n-3)(n-3)!$ vanishing entries stem from the subvectors with $\nu = 1, 2, \ldots, n-3$. The $(n-3)!$ undeformed integrals $F^\sigma|_{s_{j,n-1}=0}$ realize the soft limit $k_{n-1} \to 0$ in (8.64) and can be understood from a rescaling of the integration variables $z_j = x_j z_0$ that transforms the integration domain in (8.66) to $0 < x_2 < x_3 < \ldots < x_{n-2} < 1$. The appearance of lower-point disk integrals from the soft limit in (8.65) is easiest to see from the following IBP rewriting of (6.51)

$$F^{\sigma(23\ldots n-2)} = (2\alpha')^{n-3} \int\limits_{0<z_2<z_3<\ldots<z_{n-2}<1} dz_2\, dz_3 \, \ldots \, dz_{n-2} \prod_{1\leq p<q}^{n-1} |z_{pq}|^{-2\alpha' s_{pq}} \sigma \bigg\{ \frac{s_{n-2,n-1}}{z_{n-2,n-1}} \prod_{k=2}^{n-3} \sum_{j=1}^{k-1} \frac{s_{jk}}{z_{jk}} \bigg\} . \tag{8.72}$$

If $\sigma(n-2) \neq n-2$, then the denominator of $\sigma\{\frac{s_{n-2,n-1}}{z_{n-2,n-1}}\}$ involves non-adjacent variables $z_{\sigma(n-2)}, z_{n-1}$ in the integration domain of (8.72). One can set $s_{j,n-1} = 0$ at the level of the integrand and reproduce the zeros on the right-hand side of (8.65). If $\sigma(n-2) = n-2$ in turn, the Koba–Nielsen integral over $|z_{n-2,n-1}|^{-2\alpha' s_{n-1,n-2}-1}$ results in a kinematic pole $s_{n-2,n-1}^{-1}$ whose residue is obtained from setting $z_{n-2} = 1$ in the integrand. This residue is given by the $(n-1)$-point integral $F^{\sigma(23\ldots n-3)}$ on the right-hand side of (8.65), and the soft limit $s_{j,n-1} = 0$ suppresses the regular terms in $s_{n-2,n-1}$ beyond the residue.

The second regularized boundary value (8.21) obtained from the $(n-2)!$ integrals in (8.66) reproduces the undeformed n-point disk integrals in its first $(n-3)!$ components [269]

$$C_1 \big|_{s_{0j}=0} \to \big(F^\sigma, \ldots \big) \, . \tag{8.73}$$

This can be intuitively understood from the fact that $z_0 \to 1$ restores the original integration domain $0 < z_2 < \ldots < z_{n-2} < 1$ of F^σ, and we are setting $s_{0j} = 0$ in (8.73) to remove the deformation of the Koba–Nielsen factor. However, the components in the ellipsis of (8.73) involve lower-multiplicity contributions from the difference between the regions $z_{n-2} \in (z_{n-3}, z_0)$ and $z_{n-2} \in (z_{n-3}, 1)$: For some the components of the integrands ω^σ_ν in (8.66) with $\nu \leq n-3$, the difference $z_{n-2} \in (z_0, 1)$ between the above regions contributes to the α'-expansion even though it shrinks to zero size as $z_0 \to 1$. The detailed evaluation of these $\nu \leq n-3$ components of C_1 is subtle and fortunately not needed to derive the recursion (8.64).

Note that the field-theory limit $[\Phi(e_0, e_1)]_{ij} = \delta_{ij} + \mathcal{O}(\alpha'^2)$ of the associator in (8.64) together with the three-point integral $F^\emptyset = 1$ imply by induction in n that $F^{\sigma_j} = \delta_{j,1} + \mathcal{O}(\alpha'^2)$, i.e. that the α'-expansions of all the F^{σ_j} with $j \neq 1$ start at order α'^2. This is one way of deriving the orthonormal field-theory limits (6.52) of the ${F_P}^Q$.

In summary, the relation (8.21) between regularized boundary values and their representations (8.71), (8.73) for the specific solution $\hat{F}$ of the KZ equation in (8.66) implies the recursion (8.64) for n-point disk integrals.

8.5.4. *Connection with twisted deRham theory and outlook*

The z_0-deformed integrals (8.66) are special cases of more general Koba–Nielsen or Selberg integrals over the disk boundary with an arbitrary number of integrated and unintegrated punctures [348, 209, 274]. They obey KZ equations in multiple variables, and a recursion for the braid matrices in their differential operator has been given in [190]. The discussion in the reference is tailored to specific fibration bases w.r.t. IBP, and the transformation matrices to the bases (8.66) in the case of four unintegrated punctures can be found in [337]. This is how the all-multiplicity results for braid matrices in [190] translate into the n-point instances of e_0, e_1 in [337].

The coaction properties (8.39) of the n-point disk integrals in string amplitudes generalize to the case of more than three unintegrated punctures at $(z_i, z_j, z_k) \to (0, 1, \infty)$, for instance to the family of Selberg integrals (8.66) with an $(n-2)!$ basis of integration contours. The α'-expansions of Selberg integrals with an arbitrary number of integrated and unintegrated punctures were investigated in [349]. Their coactions in the basis choice of the reference line up with the master formula (8.41) that initially arose from studies of dimensionally regulated Feynman integrals [323, 324, 325, 326]. It is striking to see that the coaction formula (8.41) manipulating contours γ_j and differential forms ω_j in twisted-(co-)homology bases is compatible with that of the polylogarithms in the respective ϵ- or α'-expansions. A mathematical proof for Lauricella hypergeometric functions can be found in [327].

Selberg integrals with arbitrary numbers of integrated and unintegrated punctures on a disk boundary have been generalized to genus one and investigated from a multitude of perspectives in the mathematics [350, 351] and physics [352, 353, 354, 355, 356] literature. These references offer several lines of attack to expand the configuration-space integrals of one-loop open-string amplitudes in α'. In particular, the construction of [354] can be viewed as a direct genus-one analogue of the Drinfeld-associator method of this section.

8.6. Berends–Giele recursion for disk integrals

In this section we review the construction [225] of a Berends–Giele formula to compute the α'-expansion of $Z(P|Q)$ disk integrals (6.62) recursively in the length $|P|$, or alternatively in the number of points of the associated disk amplitude (6.105). Given the interpretation of Berends–Giele currents as coefficients in the perturbiner solution of an equation of motion, this method adds support to the introduction of Z theory [187, 225, 253]; the scalar theory whose amplitudes computed by the standard Berends–Giele method [27] are given by the integrals $Z(P|Q)$.

8.6.1. Extending the field-theory limit

The starting point behind the Berends–Giele method to evaluate disk integrals is the assumption that the Berends–Giele method to evaluate their field-theory limit $\alpha' \to 0$ as [224]

$$\lim_{\alpha' \to 0} Z(P, n|Q, n) = \lim_{s_P \to 0} s_P \phi(P|Q) \, , \tag{8.74}$$

can be lifted to arbitrary α' orders

$$Z(P, n|Q, n) = \lim_{s_P \to 0} s_P \phi^{\alpha'}(P|Q) \tag{8.75}$$

via the introduction of an α'-corrected Berends–Giele current[59] $\phi^{\alpha'}(P|Q)$. In the field-theory limit case of (8.74), the Berends–Giele current $\phi(P|Q)$ is the coefficient of perturbiner solution $\Phi(X)$ in (6.83) of the equation of motion $\Box\Phi = [\![\Phi, \Phi]\!]$ of the bi-adjoint scalar theory as reviewed in section 6.4.4. Interpreting $\phi(P|Q) = \lim_{\alpha' \to 0} \phi^{\alpha'}(P|Q)$, the required step to evaluate (8.75) is to obtain the α'-corrections to the equation of motion of the bi-adjoint theory and to recursively generate α'-dependent Berends–Giele currents from its perturbiner solution

$$\Phi(X) := \sum_{P,Q} \phi^{\alpha'}(P|Q) \, t^P \otimes \tilde{t}^Q \, e^{k_P \cdot X} , \qquad t^P := t^{p_1} t^{p_2} \dots t^{p_{|P|}} \tag{8.76}$$

with initial condition $\phi^{\alpha'}(i|j) = \delta_{ij}$ in the single-particle case. The α'-corrected equation of motion found in [225] is written in the following compact way

$$\frac{1}{2} \Box \Phi = \sum_{p=2}^{\infty} (2\alpha')^{p-2} \int^{\text{eom}} \prod_{i<j}^{p} |z_{ij}|^{-2\alpha' \partial_{ij}} \tag{8.77}$$

[59]We adopt the notation $\phi(P|Q) = \phi_{P|Q}$ whenever convenient.

$$
\begin{aligned}
&\times \Big(\sum_{l=1}^{p-1} \frac{[\Phi_{12\ldots l}, \Phi_{p,p-1\ldots l+1}]}{(z_{12}z_{23}\ldots z_{l-1,l})(z_{p,p-1}z_{p-1,p-2}\ldots z_{l+2,l+1})} + \mathrm{perm}(2,3,\ldots,p-1) \Big) \\
&= [\Phi_1, \Phi_2] + 2\alpha' \int^{\mathrm{eom}} |z_{12}|^{-2\alpha'\partial_{12}} |z_{23}|^{-2\alpha'\partial_{23}} \Big(\frac{[\Phi_{12}, \Phi_3]}{z_{12}} + \frac{[\Phi_1, \Phi_{32}]}{z_{32}} \Big) \\
&+ (2\alpha')^2 \int^{\mathrm{eom}} |z_{23}|^{-2\alpha'\partial_{23}} \prod_{j=2}^{3} |z_{1j}|^{-2\alpha'\partial_{1j}} |z_{j4}|^{-2\alpha'\partial_{j4}} \Big(\frac{[\Phi_{123}, \Phi_4]}{z_{12}z_{23}} + \frac{[\Phi_{12}, \Phi_{43}]}{z_{12}z_{43}} + \frac{[\Phi_1, \Phi_{432}]}{z_{43}z_{32}} + (2 \leftrightarrow 3) \Big) \\
&+ \ldots ,
\end{aligned}
$$

which is obtained directly from the local representation of the disk amplitude (6.8) under the following mappings discussed at length in [225]. The unintegrated vertices in (6.8) are replaced as

$$
\langle V_P V_Q V_n \rangle \longrightarrow [\Phi_P, \Phi_Q] \,, \tag{8.78}
$$

where Φ_P is a shorthand for various linear combinations of $\phi^{\alpha'}(R|S)$ as explained below. The contributions spelled out at the end of (8.77) descend from $\mathcal{A}(1,2,3) = \langle V_1 V_2 V_3 \rangle$ as well as the four- and five-point amplitudes in (6.13) and (6.18). The ellipsis refers to permutations of $[\Phi_{12\ldots l}, \Phi_{p\ldots l+1}]$ with $p \geq 5$ following the form of disk amplitudes (6.8) at six points and beyond.

The mapping denoted by $\int^{\mathrm{eom}}$ encodes a series of rules meant to compute the regularized integrals over $z_2, \ldots, z_{p-1}$ appearing in (6.8) after fixing $(z_1, z_p) = (0, 1)$ and expanding the Koba–Nielsen factor in a series of α'. The replacement $s_{ij} \to \partial_{ij}$ in the Koba–Nielsen exponents will be defined in (8.88) below. The technical details involving manipulations of shuffle-regularized polylogarithms can be found in [225] and lead to rational combinations of MZVs at each order in the α'-expansion of (8.77).

Even though the origin of (8.77) from a Lagrangian is unsettled, we interpret it as the non-linear equation of motion of Z-theory.

The shorthand Φ_P. The shorthand Φ_P in the equation of motion (8.77) denotes an expansion of several factors of $\phi^{\alpha'}(A|B)$ according to the following rules. First, define

$$
T^{\mathrm{dom}}_{A_1,A_2,\ldots,A_n} \otimes T^{\mathrm{int}}_{B_1,B_2,\ldots,B_n} \equiv \phi^{\alpha'}(A_1|B_1)\phi^{\alpha'}(A_2|B_2)\cdots\phi^{\alpha'}(A_n|B_n) \tag{8.79}
$$

for arbitrary words A_i and B_j. Next, define linear combinations

$$
T^{B_1,B_2,\ldots,B_n}_{A_1,A_2,\ldots,A_n} := T^{\mathrm{dom}}_{A_1,A_2,\ldots,A_n} \otimes T^{\mathrm{int}}_{\rho(B_1,B_2,\ldots,B_n)} \,, \tag{8.80}
$$

where the map ρ on words is given in (C.3), and it is understood here to act on the labels i of the words B_i. It is straightforward to see that $T^{B_1,B_2,\ldots,B_n}_{A_1,A_2,\ldots,A_n}$ satisfies the recursion

$$
T^{B_1,B_2,\ldots,B_n}_{A_1,A_2,\ldots,A_n} = T^{B_1,B_2,\ldots,B_{n-1}}_{A_1,A_2,\ldots,A_{n-1}} \phi^{\alpha'}(A_n|B_n) - T^{B_2,B_3,\ldots,B_n}_{A_1,A_2,\ldots A_{n-1}} \phi^{\alpha'}(A_n|B_1) \,, \tag{8.81}
$$

which can be taken as its alternative definition. The simplest examples of (8.81) are,

$$
\begin{aligned}
T^{B_1,B_2}_{A_1,A_2} &= \phi^{\alpha'}(A_1|B_1)\phi^{\alpha'}(A_2|B_2) - \phi^{\alpha'}(A_1|B_2)\phi^{\alpha'}(A_2|B_1) \,, \\
T^{B_1,B_2,B_3}_{A_1,A_2,A_3} &= \phi^{\alpha'}(A_1|B_1)\phi^{\alpha'}(A_2|B_2)\phi^{\alpha'}(A_3|B_3) - \phi^{\alpha'}(A_1|B_2)\phi^{\alpha'}(A_2|B_3)\phi^{\alpha'}(A_3|B_1) \\
&\quad - \phi^{\alpha'}(A_1|B_2)\phi^{\alpha'}(A_2|B_1)\phi^{\alpha'}(A_3|B_3) + \phi^{\alpha'}(A_1|B_3)\phi^{\alpha'}(A_2|B_2)\phi^{\alpha'}(A_3|B_1) \,, \\
T^{B_1,B_2,B_3,B_4}_{A_1,A_2,A_3,A_4} &= \phi^{\alpha'}(A_1|B_1)\phi^{\alpha'}(A_2|B_2)\phi^{\alpha'}(A_3|B_3)\phi^{\alpha'}(A_4|B_4) - \phi^{\alpha'}(A_1|B_2)\phi^{\alpha'}(A_2|B_1)\phi^{\alpha'}(A_3|B_3)\phi^{\alpha'}(A_4|B_4) \\
&\quad - \phi^{\alpha'}(A_1|B_2)\phi^{\alpha'}(A_2|B_3)\phi^{\alpha'}(A_3|B_1)\phi^{\alpha'}(A_4|B_4) + \phi^{\alpha'}(A_1|B_3)\phi^{\alpha'}(A_2|B_2)\phi^{\alpha'}(A_3|B_1)\phi^{\alpha'}(A_4|B_4) \\
&\quad - \phi^{\alpha'}(A_1|B_2)\phi^{\alpha'}(A_2|B_3)\phi^{\alpha'}(A_3|B_4)\phi^{\alpha'}(A_4|B_1) + \phi^{\alpha'}(A_1|B_3)\phi^{\alpha'}(A_2|B_2)\phi^{\alpha'}(A_3|B_4)\phi^{\alpha'}(A_4|B_1) \\
&\quad + \phi^{\alpha'}(A_1|B_3)\phi^{\alpha'}(A_2|B_4)\phi^{\alpha'}(A_3|B_2)\phi^{\alpha'}(A_4|B_1) - \phi^{\alpha'}(A_1|B_4)\phi^{\alpha'}(A_2|B_3)\phi^{\alpha'}(A_3|B_2)\phi^{\alpha'}(A_4|B_1) \,.
\end{aligned} \tag{8.82}
$$

By construction, the above satisfy the shuffle symmetries on the B_i slots

$$T_{A_1,A_2,\ldots,A_n}^{(B_1,B_2,\ldots,B_j)\shuffle(B_{j+1},\ldots,B_n)} = 0\,, \qquad j = 1,2,\ldots,n{-}1\,, \tag{8.83}$$

where the shuffle product is understood to act on the labels i of B_i. Surprisingly, the definition (8.81) satisfies the generalized Jacobi identities (4.48) on its A_j slots; i.e. $T^{B_1,B_2,\ldots,B_n}_{A_1,A_2,\ldots,A_n}$ at fixed order of the B_i slots satisfy the same symmetries as the nested commutator $[[\ldots[[A_1,A_2],A_3]\ldots],A_n]$ such as $T^{B_1,B_2}_{A_2,A_1} = -T^{B_1,B_2}_{A_1,A_2}$ and $T^{B_1,B_2,B_3}_{A_1,A_2,A_3} + T^{B_1,B_2,B_3}_{A_2,A_3,A_1} + T^{B_1,B_2,B_3}_{A_3,A_1,A_2} = 0$.

Finally, the shorthand Φ_P for a word $P = p_1p_2\ldots p_n$ is defined as

$$\Phi_P := T^{B_1,B_2,\ldots,B_n}_{A_{p_1},A_{p_2},\ldots,A_{p_n}}\,, \tag{8.84}$$

that is, the word P captures the ordering of the labels i of the words A_i, while the labels of words B_j are in the canonical order. For example, $\Phi_2 = T^{B_1}_{A_2}$ as well as

$$\begin{aligned}
\Phi_{21} &= T^{B_1,B_2}_{A_2,A_1}\,, \\
\Phi_{231} &= T^{B_1,B_2,B_3}_{A_2,A_3,A_1}\,, \\
\Phi_{4213} &= T^{B_1,B_2,B_3,B_4}_{A_4,A_2,A_1,A_3}\,.
\end{aligned} \tag{8.85}$$

In this way, the equation of motion (8.77) leads to a recursion for the Berends–Giele currents $\phi^{\alpha'}(P|Q)$ that can be used to obtain the α'-expansion of the $Z(P|Q)$ integrals using the Berends–Giele formula (8.75).

The equation of motion up to ${\alpha'}^3$ order. Applying the integration rules discussed in [225] to (8.77) one obtains expansions such as

$$\begin{aligned}
2\alpha' \int^{\mathrm{eom}} |z_{12}|^{-2\alpha'\partial_{12}} |z_{23}|^{-2\alpha'\partial_{23}} \frac{1}{z_{12}} &= \frac{1}{\partial_{12}} \biggl(\frac{\Gamma(1-2\alpha'\partial_{12})\Gamma(1-2\alpha'\partial_{23})}{\Gamma(1-2\alpha'\partial_{12}-2\alpha'\partial_{23})} - 1 \biggr) \\
&= -(2\alpha')^2\zeta_2\partial_{23} - (2\alpha')^3\zeta_3\partial_{23}(\partial_{12}{+}\partial_{23}) \\
&\quad - (2\alpha')^4\zeta_4\partial_{23}\bigl(\partial_{12}^2 + \tfrac{1}{4}\partial_{12}\partial_{23} + \partial_{23}^2\bigr) + \mathcal{O}({\alpha'}^5)
\end{aligned} \tag{8.86}$$

and thereby the following equation of motion including α'-corrections

$$\begin{aligned}
\frac{1}{2}\Box\Phi = {}& [\Phi_1,\Phi_2] + \Bigl((2\alpha')^2\zeta_2\partial_{12} + (2\alpha')^3\zeta_3\partial_{12}(\partial_{12}+\partial_{23})\Bigr)[\Phi_1,\Phi_{32}] \\
&- \Bigl((2\alpha')^2\zeta_2\partial_{23} + (2\alpha')^3\zeta_3\partial_{23}(\partial_{12}+\partial_{23})\Bigr)[\Phi_{12},\Phi_3] \\
&+ \Bigl((2\alpha')^2\zeta_2 + (2\alpha')^3\zeta_3\bigl(\partial_{21}+2\partial_{31}+2\partial_{32}+2\partial_{42}+\partial_{43}\bigr)\Bigr)[\Phi_{12},\Phi_{43}] \\
&- \Bigl((2\alpha')^2\zeta_2 + (2\alpha')^3\zeta_3\bigl(2\partial_{21}+\partial_{31}+3\partial_{32}+\partial_{42}+2\partial_{43}\bigr)\Bigr)[\Phi_{13},\Phi_{42}] \\
&- 2(2\alpha')^3\zeta_3\bigl(\partial_{42}+\partial_{43}\bigr)[\Phi_{123},\Phi_4] + (2\alpha')^3\zeta_3\bigl(3\partial_{42}+\partial_{43}\bigr)[\Phi_{132},\Phi_4] \\
&- 2(2\alpha')^3\zeta_3\bigl(\partial_{31}+\partial_{21}\bigr)[\Phi_1,\Phi_{432}] + (2\alpha')^3\zeta_3\bigl(3\partial_{31}+\partial_{21}\bigr)[\Phi_1,\Phi_{423}] \\
&+ (2\alpha')^3\zeta_3\Bigl([\Phi_{12},\Phi_{534}] - 2[\Phi_{12},\Phi_{543}] + 2[\Phi_{123},\Phi_{54}] + 2[\Phi_{13},\Phi_{524}] - [\Phi_{132},\Phi_{54}] \\
&\quad - 2[\Phi_{134},\Phi_{52}] - 3[\Phi_{14},\Phi_{523}] + 2[\Phi_{14},\Phi_{532}] - 2[\Phi_{142},\Phi_{53}] + 3[\Phi_{143},\Phi_{52}]\Bigr) + \mathcal{O}({\alpha'}^4)\,,
\end{aligned} \tag{8.87}$$

where

$$\partial_{ij}\Phi_P := (k_{A_i}\cdot k_{A_j})\Phi_P\,. \tag{8.88}$$

For a simple example of a practical calculation using these definitions, the low-energy expansion of disk integrals up to ${\alpha'}^2$ at any multiplicity is determined from $s_A\phi^{\alpha'}(A|B)$ in (8.75) as follows (here $\phi^{\alpha'}_{A|B} :=$

$\phi^{\alpha'}(A|B)$, and the initial conditions are $\phi^{\alpha'}_{i|j} = \delta_{ij}$)

$$
\begin{aligned}
s_A\phi^{\alpha'}_{A|B} = &\sum_{\substack{A_1A_2=A\\B_1B_2=B}} (\phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_2} - \phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_1}) \\
&+ (2\alpha')^2\zeta_2 \sum_{\substack{A_1A_2A_3=A\\B_1B_2B_3=B}} \Big[(k_{A_1}\cdot k_{A_2})\big(\phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_3}\phi^{\alpha'}_{A_3|B_2} - \phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_3} \\
&\qquad + \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_1}\phi^{\alpha'}_{A_3|B_2} - \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_1}\big) \\
&\qquad + (k_{A_2}\cdot k_{A_3})\big(\phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_1}\phi^{\alpha'}_{A_3|B_3} - \phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_3} \\
&\qquad + \phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_3}\phi^{\alpha'}_{A_3|B_1} - \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_1}\big)\Big] \\
&+ (2\alpha')^2\zeta_2 \sum_{\substack{A_1A_2A_3A_4=A\\B_1B_2B_3B_4=B}} \Big[\phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_4}\phi^{\alpha'}_{A_4|B_3} - \phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_3}\phi^{\alpha'}_{A_4|B_4} \\
&\qquad + \phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_3}\phi^{\alpha'}_{A_3|B_2}\phi^{\alpha'}_{A_4|B_4} - \phi^{\alpha'}_{A_1|B_1}\phi^{\alpha'}_{A_2|B_4}\phi^{\alpha'}_{A_3|B_2}\phi^{\alpha'}_{A_4|B_3} \\
&\qquad + \phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_1}\phi^{\alpha'}_{A_3|B_3}\phi^{\alpha'}_{A_4|B_4} - \phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_1}\phi^{\alpha'}_{A_3|B_4}\phi^{\alpha'}_{A_4|B_3} \\
&\qquad - \phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_3}\phi^{\alpha'}_{A_3|B_1}\phi^{\alpha'}_{A_4|B_4} + \phi^{\alpha'}_{A_1|B_2}\phi^{\alpha'}_{A_2|B_4}\phi^{\alpha'}_{A_3|B_1}\phi^{\alpha'}_{A_4|B_3} \\
&\qquad - \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_1}\phi^{\alpha'}_{A_3|B_4}\phi^{\alpha'}_{A_4|B_2} + \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_4}\phi^{\alpha'}_{A_4|B_1} \\
&\qquad + \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_4}\phi^{\alpha'}_{A_3|B_1}\phi^{\alpha'}_{A_4|B_2} - \phi^{\alpha'}_{A_1|B_3}\phi^{\alpha'}_{A_2|B_4}\phi^{\alpha'}_{A_3|B_2}\phi^{\alpha'}_{A_4|B_1} \\
&\qquad + \phi^{\alpha'}_{A_1|B_4}\phi^{\alpha'}_{A_2|B_1}\phi^{\alpha'}_{A_3|B_3}\phi^{\alpha'}_{A_4|B_2} - \phi^{\alpha'}_{A_1|B_4}\phi^{\alpha'}_{A_2|B_2}\phi^{\alpha'}_{A_3|B_3}\phi^{\alpha'}_{A_4|B_1} \\
&\qquad - \phi^{\alpha'}_{A_1|B_4}\phi^{\alpha'}_{A_2|B_3}\phi^{\alpha'}_{A_3|B_1}\phi^{\alpha'}_{A_4|B_2} + \phi^{\alpha'}_{A_1|B_4}\phi^{\alpha'}_{A_2|B_3}\phi^{\alpha'}_{A_3|B_2}\phi^{\alpha'}_{A_4|B_1}\Big] + \mathcal{O}(\alpha'^3)\,.
\end{aligned}
\tag{8.89}
$$

For example, one can show from the above recursion that

$$
Z(13524|32451) = -\frac{1}{s_{13}s_{135}} + (2\alpha')^2\zeta_2\Big(\frac{s_{35}}{s_{135}} + \frac{s_{25}}{s_{13}} - 1\Big) + \mathcal{O}(\alpha'^3)\,. \tag{8.90}
$$

From the above example, it is not hard to imagine that these calculations, despite systematic, are long and tedious to perform by hand. A `FORM` program that computes the α'-expansion of integrals of arbitrary multiplicity up to α'^7 can be found in the git repository [336].

8.6.2. Planar binary trees and α'-corrections

From the discussion above, the α'-expansion of string disk integrals is determined by the Berends–Giele formula (8.75) whose currents $\phi^{\alpha'}(P|Q)$ are recursively generated by the equations of motion (8.77) of the non-abelian Z-theory [225].

As discussed in [157], one can promote this setup to the theory of free Lie algebras by assuming the existence of α'-corrections to the binary-tree expansion (4.125) as $b^{\alpha'}(P)$ by defining

$$
\phi^{\alpha'}(P|Q) := \langle b^{\alpha'}(P), Q\rangle\,, \tag{8.91}
$$

where $\langle A, B\rangle = \delta_{A,B}$ denotes the scalar product of words defined in (C.11). Using the explicit expressions of $\phi^{\alpha'}(P|Q)$ up to α'^7 order one can show that the Lie-polynomial form of the binary tree expansion with α'-corrections becomes

$$
\begin{aligned}
s_P b^{\alpha'}(P) = &\sum_{XY=P} [b^{\alpha'}(X), b^{\alpha'}(Y)] \\
&+ (2\alpha')^2\zeta_2 \sum_{XYZ=P} k_X\cdot k_Y [b^{\alpha'}(X), [b^{\alpha'}(Z), b^{\alpha'}(Y)]]
\end{aligned}
\tag{8.92}
$$

$$
\begin{aligned}
&-(2\alpha')^2\zeta_2 \sum_{XYZ=P} k_Y\cdot k_Z[[b^{\alpha'}(X),b^{\alpha'}(Y)],b^{\alpha'}(Z)]\\
&+(2\alpha')^2\zeta_2 \sum_{XYZW=P} [[b^{\alpha'}(X),b^{\alpha'}(Y)],[b^{\alpha'}(W),b^{\alpha'}(Z)]]\\
&-(2\alpha')^2\zeta_2 \sum_{XYZW=P} [[b^{\alpha'}(X),b^{\alpha'}(Z)],[b^{\alpha'}(W),b^{\alpha'}(Y)]]+\mathcal{O}(\alpha'^3)\,.
\end{aligned}
$$

It is important to emphasize that the symmetries of the "domain" P and "integrand" Q are different, in particular $\phi^{\alpha'}(P|Q)\neq\phi^{\alpha'}(Q|P)$, unlike its field-theory version (6.86). The integrand Q satisfies shuffle symmetry as $\phi^{\alpha'}(P|R\sqcup S)=\langle b^{\alpha'}(P),R\sqcup S\rangle=0\ \forall\ R,S\neq\emptyset$ because $b^{\alpha'}(P)$ is a Lie polynomial, as a consequence of Ree's theorem 3.1 (iv) in [143]. The shuffle symmetries of the domain P are spoiled by the monodromy properties of the disk integrals.

The Lie polynomial (8.92) begs for a combinatorial understanding via free-Lie-algebra methods in combination with the properties of MZVs following from the Drinfeld associator, whose logarithm is known to be a Lie series.

8.7. Closed strings as single-valued open strings

In this section, we review the relation between open- and closed-string α'-expansions through the single-valued map of MZVs (see section 8.2.3). In most parts of this section, we shall set $\alpha'=\frac{1}{2}$ for open-string quantities and $\alpha'=2$ for closed strings in order to implement the rescaling of $\alpha'\to4\alpha'$ in (7.58) or (7.86).

On the one hand, one can already reduce (integrated) closed-string tree-level amplitudes to open-string computations by means of the KLT formula (7.58) or (7.62). On the other hand, the KLT formula does not manifest if some of the MZVs in open-string α'-expansions (8.37) cancel in between the amplitude factors and the sine functions. As will be reviewed below, the single-valued map reduces closed-string α'-expansions to those of open strings while exposing all the dropouts of MZVs (including powers of ζ_2 at low weights and certain indecomposable MZVs such as $\zeta_{3,5}$).

Some of the selection rules on MZVs can already be illustrated from the α'-expansions (8.3) and (8.4) of four-point open- and closed-string amplitudes. It is easy to see at the level of the exponents of (8.3) and (8.4) that the expansions of the disk and sphere integrals are related by

$$
\mathrm{sv}\bigg(\frac{\Gamma(1-s_{12})\Gamma(1-s_{23})}{\Gamma(1-s_{12}-s_{23})}\bigg)=\frac{\Gamma(1-s_{12})\Gamma(1-s_{23})\Gamma(1+s_{12}+s_{23})}{\Gamma(1+s_{12})\Gamma(1+s_{23})\Gamma(1-s_{12}-s_{23})}\,, \tag{8.93}
$$

where the single-valued map is applied order by order in α' and acts trivially on the s_{ij}. According to (8.26), sv annihilates even zeta values ζ_{2k} while doubling the odd ones ζ_{2k+1}. Note that we have eliminated $s_{13}=-s_{12}-s_{23}$ in (8.4) to expose the independent variables. From the perspective of the four-point KLT formula (7.54), the trigonometric expansion of the KLT kernel $\mathcal{S}_{\alpha'}$ in terms of even zeta values via (8.42) cancels all the ζ_{2k} in the α'-expansion of $\mathcal{A}(1,2,3,4)$ or $\frac{\Gamma(1-s_{12})\Gamma(1-s_{23})}{\Gamma(1-s_{12}-s_{23})}$ and leads to the relation (8.93). In the rest of this section, we will study the n-point generalization of this observation from several perspectives and describe the cancellation of certain indecomposable MZVs at higher depth from closed-string amplitudes in terms of the single-valued map.

8.7.1. From the KLT formula to the single-valued map

The selection rules on MZVs in n-point closed-string amplitudes were firstly identified by combining the KLT relations with the structure (8.37) of the open-string α'-expansion and exploiting conjectural properties of the M_w,P_w matrices [271]. The construction in the reference starts from the general form (7.83) of the KLT relations with a symmetric choice of bases $\mathcal{B}_1,\mathcal{B}_2\to(1,P,n-1,n)$ of permutations. With the expansion (6.50) of open-string amplitudes $\mathcal{A}(\ldots)$ in terms of SYM tree amplitudes $A(\ldots)$, we obtain

$$
\mathcal{M}_n^{\mathrm{closed}}=\sum_{P,Q\in S_{n-3}} A(1,P,n-1,n)G^{PQ}\tilde{A}(1,Q,n-1,n)\,, \tag{8.94}
$$

$$G^{PQ} = \sum_{A,B\in S_{n-3}} (F^t)^P{}_A m^{-1}_{\alpha'}(1,A,n{-}1,n|1,B,n{-}1,n) F_B{}^Q \,.$$

The next step is to insert the α'-expansion (8.37) for both $(n{-}3)!\times(n{-}3)!$ matrices of disk integrals $(F^t)^P{}_A$ and $F_B{}^Q$ as well as the observation [271]

$$(\mathbb{P}^t) m^{-1}_{\alpha'} \mathbb{P} = m^{-1} \,, \qquad M^t_{2k+1} m^{-1} = m^{-1} M_{2k+1} \tag{8.95}$$

for $\mathbb{P} = \sum_{k=0}^\infty \zeta_2^k P_{2k}$ in order to move all the P_w, M_w matrices to the right of $m^{-1}_{\alpha'}$. We have suppressed the permutations indexing the $(n{-}3)!\times(n{-}3)!$ matrices $m^{-1}_{\alpha'}$ and m^{-1} which are simply the KLT kernels of string and field theory for the symmetric choice of bases $\mathcal{B}_1,\mathcal{B}_2 \to (1,P,n{-}1,n)$ in (7.83), for instance

$$m^{-1}_{\alpha'}(1,2,3,4|1,2,3,4) = \frac{\sin(\pi s_{12})\sin(\pi s_{23})}{\pi \sin(\pi(s_{12}{+}s_{23}))} \,, \qquad m^{-1}(1,2,3,4|1,2,3,4) = \frac{s_{12}s_{23}}{s_{12}+s_{23}} \tag{8.96}$$

at four points. By passing to motivic MZVs and applying the isomorphism ϕ to the f-alphabet, (8.95) leads to the following simplified form of (8.94) [271],

$$\begin{aligned}
\phi(G^{\mathfrak{m}}) &= m^{-1} \sum_{r,s=0}^\infty \sum_{\substack{a_1,a_2,\ldots,a_r \\ \in 2\mathbb{N}+1}} \sum_{\substack{b_1,b_2,\ldots,b_s \\ \in 2\mathbb{N}+1}} M_{a_r}\ldots M_{a_2}M_{a_1}M_{b_1}M_{b_2}\ldots M_{b_s} (f_{a_1}f_{a_2}\ldots f_{a_r} \sqcup\!\sqcup f_{b_1}f_{b_2}\ldots f_{b_s}) \\
&= m^{-1} \sum_{r=0}^\infty \sum_{i_1,i_2,\ldots,i_r\in 2\mathbb{N}+1} M_{i_1}M_{i_2}\ldots M_{i_r} \sum_{j=0}^r (f_{i_j}\ldots f_{i_2}f_{i_1} \sqcup\!\sqcup f_{i_{j+1}}f_{i_{j+2}}\ldots f_{i_r}) \,,
\end{aligned} \tag{8.97}$$

where the sums over words in odd a_j, b_j have been rearranged to expose the coefficients of a given matrix product in the last line. At this point, one can recognize the form (8.27) of the single-valued map in the f-alphabet and obtain [322]

$$\begin{aligned}
\phi(\mathcal{M}_n^{\rm closed}) &= -\sum_{P,Q,R\in S_{n-3}} A(1,P,n,n{-}1) S(P|Q)_1 \\
&\quad \times \sum_{r=0}^\infty \sum_{i_1,\ldots,i_r\in 2\mathbb{N}+1} {\rm sv}(f_{i_1}f_{i_2}\ldots f_{i_r}) (M_{i_1}M_{i_2}\ldots M_{i_r})_Q{}^R \tilde A(1,R,n{-}1,n) \\
&= -\sum_{P,Q,R\in S_{n-3}} A(1,P,n,n{-}1) S(P|Q)_1 \big[{\rm sv}\,\phi(F^{\mathfrak{m}})\big]_Q{}^R \tilde A(1,R,n{-}1,n)
\end{aligned} \tag{8.98}$$

upon insertion into (8.94) and changing bases of left-moving SYM amplitudes to $A(1,P,n,n{-}1)$.[60] In passing to the last line, we have identified the series over words in ${\rm sv}(f_{i_j})$ as the single-valued map of $\phi(F^{\mathfrak{m}})$ in (8.37), where ${\rm sv}(f_2)=0$ removes all contributions from the P_{2k}.

8.7.2. Closed-string amplitudes as a field-theory double copy

Since the ϕ-map retains the complete information on the MZVs in its preimage, we can rewrite (8.98) as an amplitude relation [322]

$$\mathcal{M}_n^{\rm closed} = -\sum_{P,Q\in S_{n-3}} A(1,P,n,n{-}1) S(P|Q)_1 \,{\rm sv}\, \tilde{\mathcal{A}}(1,Q,n{-}1,n) \,, \tag{8.99}$$

[60] This has been done via

$$\sum_{P\in S_{n-3}} A(1,P,n{-}1,n) m^{-1}(1,P,n{-}1,n|Q)_1 = -\sum_{P\in S_{n-3}} A(1,P,n,n{-}1) S(P|Q)_1 \,.$$

where the single-valued image of the entire open-superstring amplitude (6.50) can be presented in one of the following forms:

$$
\begin{aligned}
\mathrm{sv}\,\tilde{\mathcal{A}}(1,P,n{-}1,n) &= \sum_{Q\in S_{n-3}} (\mathrm{sv}\,F)_P{}^Q A(1,Q,n{-}1,n) \\
&= \sum_{r=0}^{\infty} \sum_{i_1,\ldots,i_r\in 2\mathbb{N}+1} \phi^{-1}\big[\mathrm{sv}(f_{i_1}f_{i_2}\ldots f_{i_r})\big] \sum_{Q\in S_{n-3}} (M_{i_1}M_{i_2}\ldots M_{i_r})_P{}^Q \tilde{A}(1,Q,n{-}1,n) \\
&= \sum_{Q\in S_{n-3}} \big(1+2\zeta_3M_3+2\zeta_5M_5+2\zeta_3^2M_3^2+2\zeta_7M_7+2\zeta_3\zeta_5\{M_3,M_5\}+2\zeta_9M_9+\tfrac{4}{3}\zeta_3^3M_3^3 \\
&\quad + 2\zeta_5^2M_5^2+2\zeta_3\zeta_7\{M_3,M_7\}+2\zeta_{11}M_{11}+\zeta_3^2\zeta_5(M_3^2M_5+2M_3M_5M_3+M_5M_3^2) \\
&\quad + 2\big(\tfrac{1}{5}\zeta_{3,3,5}-\tfrac{4}{35}\zeta_2^3\zeta_5+\tfrac{6}{25}\zeta_2^2\zeta_7+9\zeta_2\zeta_9\big)[M_3,[M_3,M_5]]+\ldots\big)_P{}^Q \tilde{A}(1,Q,n{-}1,n)\,,
\end{aligned}
\tag{8.100}
$$

with weight or α'-orders ≥ 12 in the ellipsis. The matrix anticommutators $\{M_a,M_b\}=M_aM_b{+}M_bM_a$ along with $\zeta_3\zeta_5$ and $\zeta_3\zeta_7$ are the remnant of applying the single-valued map to the contributions from $\zeta_{3,5}$ and $\zeta_{3,7}$, where the relevant terms of (8.33) are mapped to

$$
\mathrm{sv}\bigg(\zeta_3\zeta_5M_5M_3+\frac{1}{5}\zeta_{3,5}[M_5,M_3]\bigg) = 4\zeta_3\zeta_5M_5M_3-2\zeta_3\zeta_5[M_5,M_3] = 2\zeta_3\zeta_5\{M_3,M_5\} \tag{8.101}
$$
$$
\mathrm{sv}\bigg(\zeta_3\zeta_7M_7M_3+\frac{1}{14}(\zeta_{3,7}+3\zeta_5^2)[M_7,M_3]\bigg) = 4\zeta_3\zeta_7M_7M_3-2\zeta_3\zeta_7[M_7,M_3] = 2\zeta_3\zeta_7\{M_3,M_7\}\,,
$$

see (8.25) for $\mathrm{sv}\,\zeta_{3,5}$ and $\mathrm{sv}\,\zeta_{3,7}$. While $\zeta_{3,5}$, $\zeta_{3,7}$ and in fact all higher-depth MZVs with at most two odd letters f_af_b in their f-alphabet image drop out from closed-string amplitudes, the α'-expansion (8.100) retains $\zeta_{3,3,5}$ as the simplest conjecturally irreducible higher-depth MZV in the image of the single-valued map.

Even though the derivation started out from the string-theory KLT formula (7.83), we brought the closed-string amplitude into the form (8.99) of a *field-theory* KLT formula with the α'-independent kernel $S(P|Q)_1$ in (4.160). At tree level, the closed superstring is said to be a field-theory double copy of SYM with the single-valued open superstring. A similar type of field-theory double copy was found for the open superstring in (6.69) with the scalar Z-integrals in the place of the $\mathrm{sv}\,\tilde{\mathcal{A}}$: both double-copy formulae (6.69) and (8.99) for open and closed superstrings involve SYM trees as a field-theory building block and carry the entire α'-dependence in a string-theoretic double-copy constituent Z or $\mathrm{sv}\,\tilde{\mathcal{A}}$.

Permutation invariance of the field-theory KLT formula hinges on the BCJ relations of the double-copy constituents which are certainly satisfied for the SYM amplitudes $A(\ldots)$ on the left of the KLT matrix in (8.99). The single-valued open-superstring amplitudes in turn obey BCJ relations [137] by the reasoning in section 8.4.1 – the matrix products at each order in the α'-expansion of (8.100) preserve the BCJ relations of the SYM amplitudes.

8.7.3. Sphere integrals as single-valued disk integrals

The relation between closed and single-valued open superstrings as well as the associated KLT relation (8.99) can be rewritten at the level of Parke–Taylor-type sphere integrals $J(P|Q)$ defined in (7.61). This can be seen by inserting the single-valued map of (6.69),

$$
\mathrm{sv}\,\widetilde{\mathcal{A}}(P) = -\sum_{Q,R\in S_{n-3}} \mathrm{sv}\,Z(P|1,Q,n,n{-}1)S(Q|R)_1\widetilde{A}(1,R,n{-}1,n)\,, \tag{8.102}
$$

into (8.99) and comparing with the representation (7.70) of $\mathcal{M}_n^{\mathrm{closed}}$. The coefficients of the $(n{-}3)!^2$ independent bilinears $A(1,P,n,n{-}1)\widetilde{A}(1,Q,n{-}1,n)$ have to agree in both representations of $\mathcal{M}_n^{\mathrm{closed}}$, and we conclude that [137]

$$
\mathrm{sv}\,Z(P|Q) = J(P|Q)\,. \tag{8.103}
$$

In comparing the definitions (6.62) and (7.61) of the disk and sphere integrals, the single-valued map is seen to effectively trade a disk integration over the domain $D(P)$ in (3.77) for a sphere integration with an insertion of the antiholomorphic Parke–Taylor factor $\overline{\mathrm{PT}(P)}$. This is natural from the connection between disk orderings and Parke–Taylor factors via Betti-deRham duality [357, 358], relating simple poles of $\overline{\mathrm{PT}(P)}$ in $\overline{z}_i-\overline{z}_j$ to inequalities $z_i < z_j$ characterizing the integration domain $D(P)$.

Instead of relying on the α'-expansion (8.37) of open-superstring amplitudes and the properties (8.95) of the matrices P_w, M_w, one can prove (8.103) at all multiplicities and orders in α' via single-valued integration [278]. A simple "physicists' proof" on the basis of the Betti-deRham duality between $D(P)$ and $\overline{\mathrm{PT}(P)}$ as well as standard transcendentality conjectures on MZVs can be found in [359], and the reader is referred to [280, 223] for a mathematically rigorous proof. Moreover, the fact that the expansion coefficients of $J(P|Q)$ are single-valued MZVs can be explained from the study of single-valued correlation functions [360].

As an important plausibility check of (8.103), we note that both sides obey BCJ relations in both P and Q. First, IBP relations among Parke–Taylor factors readily imply the BCJ relations of Z and J in Q and those of J in P. Second, the sv-action $\frac{1}{\pi}\sin(\pi x) \rightarrow x$ on the trigonometric factors of section 7.3 maps the monodromy relations of Z in P into BCJ relations, see section 8.4 and [137].

8.7.4. The web of field-theory double copies for string amplitudes

Single-trace amplitudes $\mathcal{A}^{\mathrm{het}}$ of gauge multiplets in heterotic string theories have been expressed in terms of SYM trees and the sphere integrals J in (7.133). By the relation (8.103) between sphere and single-valued disk integrals, one can identify [137],

$$\mathcal{A}^{\mathrm{het}}(P) = \mathrm{sv}\,\mathcal{A}(P)\,, \tag{8.104}$$

i.e. single-trace amplitudes of the gauge multiplet in type I and heterotic string theories are related by the single-valued map. Moreover, the field-theory double copy (7.119) together with (8.102) and (8.103) imply that all massless tree amplitudes for the heterotic string reduce to single-valued type I amplitudes [211],

$$\mathcal{M}_n^{\mathrm{het}} = -\sum_{P,Q\in S_{n-3}} A_{(DF)^2+\mathrm{YM}+\phi^3}(1,P,n,n{-}1)S(P|Q)_1\,\mathrm{sv}\,\mathcal{A}(1,Q,n{-}1,n)\,, \tag{8.105}$$

where the amplitudes $A_{(DF)^2+\mathrm{YM}+\phi^3}$ of the $(DF)^2{+}\mathrm{YM}{+}\phi^3$ field theory [262] (see section 7.5.3) are rational functions of α'. Just like the expression (8.99) for type II amplitudes, (8.105) double-copies single-valued open superstrings with a field theory (with $A_{(DF)^2+\mathrm{YM}+\phi^3}$ in the place of SYM amplitudes in case of the heterotic string).

Similar double-copy formulae apply to bosonic strings: removing the bi-adjoint scalars from the $(DF)^2+\mathrm{YM}+\phi^3$ theory leaves a simpler field theory $(DF)^2+\mathrm{YM}$ of massless and massive vectors [262] which casts n-point tree-level amplitudes $\mathcal{A}_n^{\mathrm{bos}}$ and $\mathcal{M}_n^{\mathrm{bos}}$ of open and closed bosonic strings into the compact form [211]

$$\begin{aligned}\mathcal{A}_n^{\mathrm{bos}}(R) &= -\sum_{P,Q\in S_{n-3}} A_{(DF)^2+\mathrm{YM}}(1,P,n,n{-}1)S(P|Q)_1 Z(R|1,Q,n{-}1,n)\,, \\ \mathcal{M}_n^{\mathrm{bos}} &= -\sum_{P,Q\in S_{n-3}} \tilde{A}_{(DF)^2+\mathrm{YM}}(1,P,n,n{-}1)S(P|Q)_1\,\mathrm{sv}\,\mathcal{A}^{\mathrm{bos}}(1,Q,n{-}1,n)\,.\end{aligned} \tag{8.106}$$

Moreover, the gravity sector of heterotic-string amplitudes admits an alternative form [211]

$$\mathcal{M}_n^{\mathrm{het}}\big|_{\mathrm{grav}} = -\sum_{P,Q\in S_{n-3}} \tilde{A}(1,P,n,n{-}1)S(P|Q)_1\,\mathrm{sv}\,\mathcal{A}^{\mathrm{bos}}(1,Q,n{-}1,n)\,, \tag{8.107}$$

where the supersymmetries arise from the opposite double-copy constituent as compared to (8.105) – from the SYM field-theory amplitudes $\tilde{A}$ instead of single-valued superstring disk amplitudes.

A summary of the field-theory double-copy formulae (6.69), (8.99) and (8.104) to (8.107) for tree amplitudes in various string theories can be found in table 3. In all cases, the double copy refers to the KLT formula with α'-independent kernel $S(P|Q)_1$ and features a field-theory building block (SYM, $(DF)^2+\mathrm{YM}$ or $(DF)^2+\mathrm{YM}+\phi^3$) without any transcendentality. The infinite tower of massive poles characteristic

string $\otimes$ QFT	SYM	$(DF)^2 + \text{YM}$	$(DF)^2 + \text{YM} + \phi^3$
Z-theory	open superstring	open bosonic string	comp. open bosonic string
sv(open superstring)	closed superstring	heterotic (gravity)	heterotic (gauge & gravity)
sv(open bosonic string)	heterotic (gravity)	closed bosonic string	comp. closed bosonic string

Table 3: Double copy constructions of tree-level amplitudes in various string theories as presented in [211].

to string amplitudes occurs through the other double-copy constituent – either the universal basis of disk integrals Z for open strings or single-valued open-string amplitudes in case of type II or heterotic strings.

Note that the compactified versions of open and closed bosonic strings in the rightmost column of table 3 refer to the geometric realization of the gauge sector of the heterotic string: the Kac–Moody currents $\mathcal{J}^a(z)$ in the vertex operators of the gauge multiplet in section 7.5.1 can be obtained from compactifying free bosons $\partial_z X^I(z)$ on a torus [97].

8.7.5. Twisted KLT relations

An interesting variant of the sphere integrals $J(P|Q)$ in (7.61) arises in so-called chiral or twisted string theories [361, 362, 363]. These theories are characterized by finite spectra due to a flipped level-matching condition that can be informally identified with a relative sign flip of α' between left- and right-moving worldsheet degrees of freedom. In particular, the spectrum of twisted type II superstrings reduces to the associated supergravity multiplets.

At the level of the sphere integrals in the tree-level amplitudes of twisted string theories, the sign flip between left- and right movers applies to the antiholomorphic part of the Koba–Nielsen factor,

$$\widehat{J}(P|Q) := \Big(-\frac{\alpha'}{2\pi}\Big)^{n-3} \int\limits_{\mathbb{C}^{n-3}} \frac{d^2 z_1 \ d^2 z_2 \ \cdots \ d^2 z_n}{\text{vol}(\text{SL}(2,\mathbb{C}))} \prod_{i<j}^{n} \Big(\frac{\overline{z}_{ij}}{z_{ij}}\Big)^{\frac{\alpha'}{2} s_{ij}} \ \text{PT}(Q)\overline{\text{PT}(P)}\,, \tag{8.108}$$

where the single-valued factors of $|z_{ij}|^{-\alpha' s_{ij}}$ in (7.61) from the correlators of conventional strings are replaced by $(\frac{\overline{z}_{ij}}{z_{ij}})^{\alpha' s_{ij}/2}$. Apart from these modifications of the Koba–Nielsen factors, the chiral correlators $\mathcal{K}_n$ among closed-string vertex operators can be freely interchanged between the type II versions of twisted and conventional strings [363, 364]. Hence, the supergravity n-point function, computed from twisted type II strings, takes the form of (7.70) with the modified sphere integrals $\widehat{J}(P|Q)$ in the place of $J(P|Q)$. In order to arrive at the field-theory KLT formula (4.159) for supergravity, the sphere integrals of the twisted strings have to directly match the doubly-partial amplitudes $m(P|Q)$,

$$\widehat{J}(P|Q) = m(P|Q)\,. \tag{8.109}$$

However, the sphere integrals (8.108) are ill-defined due of the multivalued factors of $(\frac{\overline{z}_{ij}}{z_{ij}})^{\alpha' s_{ij}/2}$ in the integrand. Still, one can formally define $\widehat{J}(P|Q)$ by a KLT formula, where the reversal of α' along with the antiholomorphic $\overline{z}_{ij}$ leads to a sign-flipped version of standard Z-integrals

$$\widehat{Z}(P|Q) = Z(P|Q)\,\big|_{\alpha' \to -\alpha'}\,, \tag{8.110}$$

namely

$$\widehat{J}(P|Q) = -\sum_{P,Q \in S_{n-3}} Z(1,P,n{-}1,n|A)\mathcal{S}_{\alpha'}(P|Q)_1 \widehat{Z}(1,Q,n,n{-}1|B)\,. \tag{8.111}$$

Upon comparison with the requirement (8.109), the KLT formula for the twisted sphere integral needs to reproduce the α'-independent doubly-partial amplitude,

$$m(A|B) = -\sum_{P,Q \in S_{n-3}} Z(1,P,n{-}1,n|A)\mathcal{S}_{\alpha'}(P|Q)_1 \widehat{Z}(1,Q,n,n{-}1|B)\,. \tag{8.112}$$

Hence, the conclusion is that a sign-flip in one of the Z-integrals in the conventional KLT formula (7.62) is enough to cancel the entire tower of α'-corrections. In fact, (8.112) can be deduced from the twisted period relations [236] as introduced into the physics literature in [189].

Another way of understanding the dropout of α'-corrections from (8.108) is to revisit the simplification of the sphere integrals in section 8.7.1. The sign flip effectively reverses $M_{2k+1} \to -M_{2k+1}$ in one of the F-factors in the matrix G^{PQ} of sphere integrals in (8.94) and turns (8.97) into

$$\phi(G^{\mathfrak{m}}) \to m^{-1} \sum_{r=0}^{\infty} \sum_{i_1,i_2,\ldots,i_r \in 2\mathbb{N}+1} M_{i_1} M_{i_2} \ldots M_{i_r} \sum_{j=0}^{r} (-1)^j (f_{i_j} \ldots f_{i_2} f_{i_1} \shuffle f_{i_{j+1}} f_{i_{j+2}} \ldots f_{i_r}) \,. \tag{8.113}$$

By the alternating signs $(-1)^j$ on the right-hand side, the coefficient of each non-trivial matrix product $M_{i_1} \ldots M_{i_r}$ with $r \neq 0$ cancels [363]. Hence, the matrix G^{PQ} in (8.94) is mapped to the KLT kernel in passing to the twisted string, and we obtain supergravity amplitudes as expected.

One can also turn the logic around and impose that the expression for G^{PQ} in (8.94) reduces to

$$m^{-1}(1,P,n{-}1,n|1,Q,n{-}1,n) = \sum_{A,B \in S_{n-3}} (F^t)^P{}_A m^{-1}_{\alpha'}(1,A,n{-}1,n|1,B,n{-}1,n) \widehat{F}_B{}^Q \tag{8.114}$$

with $\widehat{F}_B{}^Q = F_B{}^Q\big|_{\alpha' \to -\alpha'}$ as in (8.110). The coefficients of $f_2^{\ell=0,1,2,3,\ldots}$ or f_{2k+1} in (8.114) then imply the properties (8.95) of the matrices P_w, M_w.

Finally, we note an amusing variant of (8.112): instead of obtaining doubly-partial amplitudes $m(P|Q)$ by contracting $Z(A|P)\widehat{Z}(B|Q)$ with the string-theory KLT kernel $\mathcal{S}_{\alpha'}(A|B)_1$, one can instead contract the free indices P,Q with the field-theory kernel $S(P|Q)_1$. In this way, one arrives at the inverse $m_{\alpha'}$ of the *string-theory* KLT kernel [235]

$$m_{\alpha'}(A|B) = -\sum_{P,Q \in S_{n-3}} Z(A|1,P,n{-}1,n) S(P|Q)_1 \widehat{Z}(B|1,Q,n,n{-}1) \,. \tag{8.115}$$

Note that the heterotic and bosonic versions of twisted strings provide a worldsheet realization of the $(DF)^2 + \mathrm{YM}$ and $(DF)^2 + \mathrm{YM} + \phi^3$ theories [365]: up to the sign flip $\alpha' \to -\alpha'$ between left- and right-movers, the Parke–Taylor expansion (7.115) of bosonic correlators takes the identical form in twisted string theories. The massive modes in the $(DF)^2 + \mathrm{YM}$ and $(DF)^2 + \mathrm{YM} + \phi^3$ theories arise from asymmetric double copies of vertex operators for an open-string tachyon and the first mass level of compactified open bosonic strings. Similarly, the spectrum of heterotic twisted strings contains a colorless spin-two multiplet from a double copy of tachyons with the first mass level of the open superstring.

The generalizations of the correlators $\mathcal{K}_n, \mathcal{K}_n^{\mathrm{bos}}$ to massive states also take a universal form for conventional and twisted strings up to $\alpha' \to -\alpha'$. This was exploited in [366] to pioneer field-theory double-copy structures in tree-level amplitudes involving massive open- and closed-superstring states based on tools from the heterotic twisted string.

9. Conclusion and outlook

This work aims to give a comprehensive review of string tree-level amplitudes in the pure spinor formalism. The manifestly spacetime supersymmetric worldsheet description of the pure spinor superstring reviewed in section 3 introduces massless open-string excitations in the framework of ten-dimensional SYM, see section 2. The OPEs of these superspace vertex operators give rise to the multiparticle formalism whose rich combinatorial structure has been presented from different perspectives in section 4. The multiparticle formalism connects conformal-field-theory techniques with recursive organizations of Feynman diagrams and led to compact formulae for n-point tree amplitudes of SYM, see section 5.

The setup of the first sections is the key to find the decomposition (6.49) of n-point superstring disk amplitudes into a basis of color-ordered SYM trees. This is the main result of this review whose derivation and interplay with disk integrals and their field-theory limit is presented in section 6. The structure of the disk amplitude together with its corollaries for type II superstrings and heterotic strings have profound implications on field-theory amplitudes reviewed in section 7 – the color-kinematics duality of gauge theories and Goldstone bosons as well as double-copy descriptions of gravity, Born–Infeld and Einstein–Yang–Mills. The last section 8 is dedicated to the low-energy expansion of superstring tree-level amplitudes and the elegant mathematical structures of the multiple zeta values therein.

From the material in this review, both the moduli-space integrand for n-point string tree-level amplitudes and the α'-expansion of the integrated expressions are available to any desired order. We have presented the strong connectivity of both the integrands and the integrated results with the web of double-copies among field-theory amplitudes and various areas of pure mathematics including combinatorics, number theory and algebraic geometry. The detailed control over string tree amplitudes cross-fertilizes with ambitious questions on string dualities (say through the multiple zeta values in multiparticle type IIB amplitudes) but also offers new connections between perturbative string theories beyond any known duality (e.g. gauge amplitudes of heterotic strings as single-valued type I amplitudes).

The diverse insights unlocked by the results on string tree-level amplitudes in this review motivate a similar investigation of loop amplitudes, where already the last years witnessed progress on several frontiers. We shall now give an overview of recent loop-level developments that generalize selected aspects of this review beyond tree level. The subsequent path through the literature is far from complete and may quickly become outdated after the time of writing this review. The reader is referred to [367] for an overview of loop-level amplitude computations in the pure spinor formalism as of October 2022 and to the white paper [77] for a status report on a broader selection of topics in string perturbation theory as of March 2022.

9.1. Loop amplitudes in the pure spinor formalism

By the manifest spacetime supersymmetry of the pure spinor formalism, it automatically incorporates a variety of cancellations in loop amplitudes between internal bosons and fermions. In the pure spinor prescription for loop-level string amplitudes, many of these cancellations can be traced back to the saturation of fermionic zero modes. The loop-amplitude prescription in the "minimal" worldsheet variables of section 3 dates back to 2004 [2], followed by its extension to "non-minimal" variables in 2005 [3]. A central ingredient in loop amplitudes of the pure spinor superstring is a composite b-ghost whose explicit form in the non-minimal variables [3] involves poles in the pure spinor ghosts.

Just like at tree level, the loop amplitudes computed from these prescriptions automatically involve kinematic factors in pure spinor superspace after integrating out the non-zero modes of the worldsheet variables. Moreover, the pure spinor formalism is readily compatible with the chiral-splitting procedure [5, 368] to express closed-string correlators at arbitrary genus as a holomorphic square of chiral amplitudes that integrate to open-string amplitudes after specifying boundary conditions for the endpoints. In fact, the so-called homology invariance of chiral amplitudes – their single-valuedness on higher-genus surfaces under suitable shifts of the loop momenta – provided crucial input for recent loop-amplitude computations in the pure spinor formalism.

The constraints from of zero-mode counting facilitated the derivation of non-renormalization theorems in string theory [369, 370] and led to multiloop results on the ultraviolet structure of maximal supergravity through a worldline version of the pure spinor formalism [174]. The computation of non-vanishing string loop amplitudes with the pure spinor formalism was initiated with the one-loop four-point amplitude in 2004 [2] and the two-loop four-point amplitude in 2005 [91]. The bosonic components of the two-loop result were later on confirmed [371] to reproduce the earlier two-loop four-point computation in the RNS formalism [372].

The non-minimal pure spinor formalism has been used to compute one-loop five-point amplitudes [373], the exactly normalized four-point amplitudes at one loop [110] and two loops [111] as well as the low-energy limits of the three-loop four-point [86] and two-loop five-point [374] amplitudes. In all of these cases, the b-ghosts only contribute through their zero modes. However, the non-zero modes of the b-ghost and the complexity of its multiparticle correlators currently cause a bottleneck in performing higher-order computations directly from the prescription. Still, consistency conditions on loop amplitudes and in particular the multiparticle formalism of section 4 often allowed to circumvent the most daunting challenges from the b-ghost and led to many recent advances on higher-point amplitudes.

The multiparticle formalism spawned simplified expressions for one-loop open- and closed-string amplitudes in an integral basis at five points [168] and at six points [375]. The latter reference also reconciles the hexagon gauge anomaly of individual worldsheet diagrams in type I theories [376, 377] with BRST cohomology techniques and derives the anomaly kinematic factor [378] from an explicit amplitude representation. Based on additional input from locality, chiral splitting and the associated homology invariance, a systematic procedure to derive one-loop correlators is described in [379, 380, 171]. The resulting chiral amplitudes enjoy a double-copy structure [381] similar to the KLT-type formula for the open-superstring correlator (6.73) at genus zero. However, the coefficients of holomorphic Eisenstein series in ($n \geq 8$)-point correlators have so far resisted a computation from this method.

Similarly, two-loop five-point amplitudes were constructed beyond their low-energy limit by a confluence of BRST invariance, locality and chiral-splitting techniques [172]. Their parity-even bosonic components were later on verified from a first-principles computation in the RNS formalism [382]. Finally, an exact-in-α' expression for the three-loop four-point amplitude was proposed in [383] based on input from the field-theory limit, ambitwistor strings and modular invariance. It would be interesting to analyze this three-loop result from a pure spinor perspective.

The combined power of the multiparticle formalism, BRST invariance and locality has also been used to directly propose loop integrands for ten-dimensional SYM, see [169] for one-loop integrands up to six points and [170] for two-loop five points. The five-point results at one and two loops readily manifested the color-kinematics duality and induced the corresponding loop integrands for type II supergravity in pure spinor superspace via double copy as in [384]. Based on the tropical-geometry methods of [385], these five-point field-theory amplitudes were independently derived from the $\alpha' \to 0$ limit of the corresponding string amplitudes at one loop [169] and at two loops [172].

However, the present pure spinor methods leave open questions on the loop-level realization of the color-kinematics duality and double copy at $n \geq 6$ points. The first superspace construction of one-loop six-point SYM numerators in [169] violated certain kinematic Jacobi identities. These violations disappear[61] in passing to the linearized variant of Feynman propagators [387] that typically arise from ambitwistor strings [388]. The resulting supergravity integrands on linearized propagators are available in KLT- and cubic-diagram form [389, 387] obtained from the forward limits of (4.159) and (7.67). A solution of all the one-loop kinematic Jacobi identities on quadratic propagators was offered in [390] by taking the field-theory limit of the corresponding string amplitudes in different color orderings. However, the conventional cubic-diagram double copy [228] of these color-kinematics dual SYM numerators conflicts with BRST invariance, so it is an open problem to construct supergravity loop integrands on quadratic propagators at $n \geq 6$ points.

Both the above subtleties in finding string-theory realizations of the gravitational double copy and the non-zero mode contributions of the b-ghosts kick in at the one-loop six-point level. One may speculate about a connection between the two kinds of challenges, for instance whether an incorporation of the b-ghost into the multiparticle formalism is the missing puzzle piece for n-point one-loop string amplitudes with manifest double-copy structure in the field-theory limit. This scenario is supported by the role of the b-ghost for a kinematic algebra and its connection with tree-level multiparticle superfields identified in [391]. Also higher-loop string amplitudes call for further investigations of the b-ghost since its poles in the pure spinor ghosts necessitate regularization techniques such as [392, 393] at genus $g \geq 3$.

9.2. Worldsheet integrals in loop-level string amplitudes

A central line of tree-level results in this review is driven by the Parke–Taylor bases of disk integrals Z in (6.62) and sphere integrals J in (7.61). First, their integration-by-parts relations and $(n-3)!$ bases fruitfully resonate with the BRST properties and BCJ relations of the accompanying kinematic factors in pure spinor superspace. Second, their logarithmic singularities (including the absence of double poles) ensure that the α'-expansion is uniformly transcendental and in fact realizes field-theory amplitude relations along with infinite families of multiple zeta values. These properties of Parke–Taylor integrals at tree level motivate the goal of constructing similar kinds of integral bases at higher genus.

[61] The violations of kinematic Jacobi identities in the one-loop six-point results of [169] also disappear in MHV helicity configurations upon dimensional reduction to $D = 4$ [386].

At genus one, generating functions Z_η of open-string integrals in different string theories furnish conjectural $(n{-}1)!$-element bases with uniformly transcendental α'-expansions [352, 353]. At suitable orders in the bookkeeping variables $\eta_2, \eta_3, \ldots, \eta_n$, one can read off the combinations of theta functions for one-loop correlators of the pure spinor superstring that share the logarithmic singularities of the Parke–Taylor factors. More precisely, the function space to assemble the one-loop analogues of the tree-level correlators $\mathcal{K}_n$ in (6.73) is controlled by the loop momenta of the chiral-splitting procedure [5, 368] and the coefficients $g^{(k)}(z,\tau)$ of the Kronecker-Eisenstein series [394],

$$\frac{\theta_1'(0,\tau)\theta_1(z{+}\eta,\tau)}{\theta_1(z,\tau)\theta_1(\eta,\tau)} = \frac{1}{\eta} + \sum_{k=1}^{\infty} \eta^{k-1} g^{(k)}(z,\tau)\,. \tag{9.1}$$

After integration over the loop momenta, the moduli-space integrand of open- and closed-string amplitudes is expressed in terms of doubly-periodic versions $f^{(k)}(z,\tau)$ of the Kronecker-Eisenstein coefficients $g^{(k)}(z,\tau)$ which also manifest the modular properties in the closed-string case.

In the same way as Parke–Taylor factors share the BCJ relations of gauge-theory tree amplitudes, the combinations of $g^{(k)}(z_{ij},\tau)$ and loop momenta seen in one-loop correlators [381, 171] are observed to obey the same identities as the BRST-invariant kinematic factors [380]. Up to and including seven points, this duality between kinematics and worldsheet functions is fully established in the references and underpins a double-copy structure in the chiral amplitudes. Starting from eight points, the chiral amplitudes involve holomorphic Eisenstein series $\mathrm{G}_k(\tau) = -g^{(k)}(0,\tau)$ with $k \geq 4$, and it is an open problem to accommodate them into the duality between kinematics and worldsheet functions.

The loop-level analogue of the string-theory KLT relation of section 7.2.1 is uncharted terrain at the time of writing. However, the tree-level monodromy relations whose vibrant interplay with KLT relations was illustrated in section 7.3 were generalized to loop level, investigated from several perspectives [395, 396, 397, 398, 399] and extended to one-loop amplitude relations between mixed open-and-closed-string amplitudes and pure open-string amplitudes [400]. It would be very interesting to relate properties of the Kronecker-Eisenstein-type functions in the chiral correlators to one-loop monodromy relations as done for Parke–Taylor factors and disk orderings through the relations (6.76), (6.77) and (7.77) of the $Z(P|Q)$ integrals.

The dependence of integrated string loop amplitudes on α' and the kinematic variables is more involved than at tree level and features branch cuts in addition to the poles for the infinite tower of massive string modes. At one loop for instance, the analytic continuations in the external momenta required by the integral representations and compatible with the poles and branch cuts are discussed in [401, 402, 403]. Already the α'-expansion of one-loop string amplitudes contains logarithms in Mandelstam invariants on top of the Laurent series in $s_{ij\ldots k}$ seen in the tree-level α'-expansions of section 8. The logarithms in one-loop four-point closed-string amplitudes due to effective tree-level interactions $D^{2k}\mathbb{R}^m$ were pioneered in [404, 405] and computed to all orders in α' in [406]. Two recent lines of attack to determine the non-analytic sector of higher-point one-loop string amplitudes are based on one-loop matrix elements of tree-level effective interactions [407] and an implementation of Witten's $i\epsilon$ prescription [408].

A prominent motivation for the computation and low-energy expansion of string loop amplitudes stems from their implications for string dualities. In this context, the main interest is in the analytic contributions to the α'-expansion which reflect new interactions in the loop-level effective actions. For type IIB superstrings, the $SL(2,\mathbb{Z})$-invariance w.r.t. the axio-dilaton field [409] must be realized in the coefficients of all independent $D^{2k}\mathbb{R}^m$ interactions. Perturbative string amplitudes carry important information on these modular invariant functions of the string coupling. Four-point string amplitudes up to and including three loops were successfully matched with $SL(2,\mathbb{Z})$-invariant coefficients of the $\mathbb{R}^4$, $D^4\mathbb{R}^4$ and $D^6\mathbb{R}^4$ interactions [410, 411, 412, 413, 86, 414].

Pure spinor methods gave rise to compact representations of loop amplitudes also beyond four points. Together with a low-energy expansion of the worldsheet integrals, the five-point amplitude computations at one loop [168] and two loops [415] made a duality analysis of $D^{2k}\mathbb{R}^5$ interactions tractable. Both references confirmed the duality properties of supermultiplet components that violate the $U(1)$ R-symmetry of type IIB supergravity [416, 272, 417] and cannot arise in four-point string amplitudes [418]. Moreover, the loop-level effective actions will involve new superinvariants starting with $D^6\mathbb{R}^5$ that are absent in the tree-level effective action [168, 415]. The classification of independent interactions and $SL(2,\mathbb{Z})$-invariant type IIB couplings necessitates precise control over the tensor structure of multiparticle loop amplitudes as provided by pure spinor superspace.

A frequently used strategy towards low-energy expansions of string loop amplitudes is to first integrate over vertex insertion points prior to the complex-structure moduli τ of the genus-g surface. In this way, the α'-expansions generate infinite families of modular invariant functions of τ w.r.t. $Sp(2g,\mathbb{Z})$ on the worldsheet (rather than the $SL(2,\mathbb{Z})$ acting on the axio-dilaton field in case of type IIB). These functions generalize the (single-valued) MZVs of section 8 to higher genus and were dubbed modular graph forms in [419, 420] after earlier case studies in [421, 404, 405]. Already at genus one, modular graph forms stimulated interdisciplinary research lines at the interface of string theory, algebraic geometry and number theory, see e.g. [422] for an overview as of November 2020, [423] for lecture notes and [424] for the connection with Brown's equivariant iterated Eisenstein integrals [425, 426]. The study of higher-genus modular graph forms started with [427, 414, 428, 429, 430, 431] and suggests a generalization to modular graph tensors [432].

From a string-theory perspective, a major appeal of modular graph forms is to investigate the loop-level generalization of the tree-level relation (8.103) between closed-string and single-valued open-string integrals. In one-loop amplitudes of open superstrings, the iterated integrals over vertex-operator insertions on a cylinder- or Möbius-strip boundary were shown in [394, 433] to yield elliptic MZVs [434] and elliptic polylogarithms [435]. There is a variety of evidence [436, 419, 437, 438] that modular graph forms may be viewed as single-valued elliptic MZVs. In particular, the closed-string counterparts [439, 440] of the conjectural $(n-1)!$ basis Z_η of one-loop open-string integrals led to an explicit all-order proposal [441] how to relate modular graph forms to single-valued elliptic MZVs in the respective α'-expansions. On the one hand, this line of reasoning aims to extract the more challenging configuration-space integrals over punctured tori from the simpler iterated integrals over cylinder- and Möbius-strip boundaries. On the other hand, this research direction may reveal loop-level manifestations of a deeper relation between closed and open strings beyond any known string duality.

Acknowledgements

We are grateful to Maor Ben-Shahar, Nathan Berkovits, Lorenz Eberhardt, Max Guillen and Callum Hunter for helpful discussions and comments on a draft version of this work. CRM is supported by a University Research Fellowship from the Royal Society. OS is supported by the European Research Council under ERC-STG-804286 UNISCAMP. CRM would like to thank the engineers and the construction workers responsible for good quality of the electric poles alongside the roads in Waterloo, Canada. Without them, this work would not have been possible in its current form. OS dedicates this work to the 18th birthday of his brother Robin Schlotterer – finally you are old enough to have fun with pure spinors, Robin.

Appendix A. Gamma matrices

Pure spinor calculations in ten dimensions often involve the handling of gamma matrices. In this appendix we review some of the most common manipulations involving ten-dimensional gamma matrices (for a computer implementation, see [442]). Most of this material can also be found in [443]. In particular, the book [444] contains a variety of discussions in general dimensions and should be consulted for further reading.

Appendix A.1. The Clifford Algebra in $\mathbb{R}^{1,9}$

Lorentzian signature. The 32×32 Dirac matrices Γ^m in ten-dimensional Minkowski space $\mathbb{R}^{1,9}$ with $m = 0,\ldots,9$ satisfy the Clifford algebra

$$\{\Gamma^m,\Gamma^n\} = 2\eta^{mn}1_{32\times 32}\,. \tag{A.1}$$

The signature of the metric is the mostly plus $(-++\cdots+)$. In the Weyl representation of Γ^m only the off-diagonal 16×16 blocks are non-vanishing, parameterized as

$$\Gamma^m = \begin{pmatrix} 0 & (\gamma^m)^{\alpha\beta} \\ (\gamma^m)_{\alpha\beta} & 0 \end{pmatrix}\,, \tag{A.2}$$

in terms of chiral gamma 16×16 matrices γ^m subject to

$$\gamma^m_{\alpha\beta}(\gamma^n)^{\beta\delta} + \gamma^n_{\alpha\beta}(\gamma^m)^{\beta\delta} = 2\eta^{mn}\delta^\delta_\alpha\,. \tag{A.3}$$

Numerical representation. An explicit representation of the 16×16 gamma matrices (A.2) is given by

$$\begin{aligned} (\gamma^0)^{\alpha\beta} &= \begin{pmatrix} 1_{8\times8} & 0 \\ 0 & 1_{8\times8} \end{pmatrix}\,, & (\gamma^0)_{\alpha\beta} &= \begin{pmatrix} -1_{8\times8} & 0 \\ 0 & -1_{8\times8} \end{pmatrix}\,, \\ (\gamma^i)^{\alpha\beta} &= \begin{pmatrix} 0 & \sigma^i_{a\dot a} \\ \sigma^i_{\dot b b} & 0 \end{pmatrix}\,, & (\gamma^i)_{\alpha\beta} &= \begin{pmatrix} 0 & \sigma^i_{a\dot a} \\ \sigma^i_{\dot b b} & 0 \end{pmatrix}\,, \\ (\gamma^9)^{\alpha\beta} &= \begin{pmatrix} 1_{8\times8} & 0 \\ 0 & -1_{8\times8} \end{pmatrix}\,, & (\gamma^9)_{\alpha\beta} &= \begin{pmatrix} 1_{8\times8} & 0 \\ 0 & -1_{8\times8} \end{pmatrix}\,, \end{aligned} \tag{A.4}$$

where σ^i with $i=1,2,\ldots,8$ are 8×8 matrices ($\mathbb{1} := 1_{2\times2}$)

$$\begin{aligned} \sigma^1_{a\dot a} &= \varepsilon\otimes\varepsilon\otimes\varepsilon\,, & \sigma^2_{a\dot a} &= \mathbb{1}\otimes\sigma^1\otimes\varepsilon\,, \\ \sigma^3_{a\dot a} &= \mathbb{1}\otimes\sigma^3\otimes\varepsilon\,, & \sigma^4_{a\dot a} &= \sigma^1\otimes\varepsilon\otimes\mathbb{1}\,, \\ \sigma^5_{a\dot a} &= \sigma^3\otimes\varepsilon\otimes\mathbb{1}\,, & \sigma^6_{a\dot a} &= \varepsilon\otimes\mathbb{1}\otimes\sigma^1\,, \\ \sigma^7_{a\dot a} &= \varepsilon\otimes\mathbb{1}\otimes\sigma^3\,, & \sigma^8_{a\dot a} &= \mathbb{1}\otimes\mathbb{1}\otimes\mathbb{1}\,, \end{aligned} \tag{A.5}$$

and σ^i with $i=1,2,3$ are the Pauli matrices with $\varepsilon = i\sigma^2$

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \quad \varepsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\,, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,. \tag{A.6}$$

While the gamma-matrix representations (A.4) are tailored to Minkowski-space $\mathbb{R}^{1,9}$, their Euclidean analogues can be found in (B.3) below.

Charge conjugation and chirality matrices. The chirality matrix in ten spacetime dimensions is given by

$$\Gamma = \Gamma^0\Gamma^1\ldots\Gamma^9 = \begin{pmatrix} 1_{16\times16} & 0 \\ 0 & -1_{16\times16} \end{pmatrix} \tag{A.7}$$

which splits a 32-component Dirac spinor into two 16-component spinors of opposite chiralities

$$\lambda = \begin{pmatrix} \lambda^\alpha \\ \lambda_\alpha \end{pmatrix} \tag{A.8}$$

called Weyl (λ^α) and anti-Weyl (λ_α). The charge conjugation matrix satisfying $C\Gamma^m = -(\Gamma^m)^T C$ is $C = \Gamma^0$. Since it is off-diagonal, the Weyl and anti-Weyl are inequivalent representations in ten dimensions (unlike in four).

Generalized Kronecker delta. It is convenient to define the generalized Kronecker delta as

$$\delta^{a_1a_2\ldots a_n}_{b_1b_2\ldots b_n} = \frac{1}{n!}\delta^{[a_1}_{b_1}\delta^{a_2}_{b_2}\cdots\delta^{a_n]}_{b_n} \tag{A.9}$$

which is totally antisymmetric in both sets of indices, e.g. $\delta^{ab}_{mn} = \frac{1}{2}(\delta^a_m\delta^b_n - \delta^b_m\delta^a_n)$. Using the notation of words, in D dimensions we have

$$\delta^{PA}_{QA} = \frac{\binom{D-p}{a}}{\binom{p+a}{a}}\delta^P_Q\,, \qquad p := |P|\,, \quad a := |A|\,, \quad p+a \leq D\,. \tag{A.10}$$

For example in $D = 10$ we have $\delta^{mna}_{pqa} = \frac{8}{3}\delta^{mn}_{pq}$. In particular, when $D = 10$ where $\delta^m_m = 10$, the identity (A.10) gives the full contraction when $P = Q = \emptyset$ as

$$\delta^{m_1\ldots m_n}_{m_1\ldots m_n} = \binom{10}{n} = 10, 45, 120, 210, 252 \quad \text{for } n = 1, 2, 3, 4, 5\,. \tag{A.11}$$

Appendix A.2. Fierz decompositions

Antisymmetric products of gamma matrices are defined by the *n-forms*

$$\gamma^{m_1m_2\ldots m_n} = \frac{1}{n!}\gamma^{[m_1}\gamma^{m_2}\ldots\gamma^{m_n]}\,, \tag{A.12}$$

and they lead to four possible configurations of Weyl-spinor indices, namely $(\gamma^{\cdots})_{\alpha\beta}$, $(\gamma^{\cdots})^{\alpha\beta}$, $(\gamma^{\cdots})_\alpha{}^\beta$ and $(\gamma^{\cdots})^\beta{}_\alpha$. The mixed combinations occur at ranks $2, 4, 6, 8, 10$,

$$(\gamma^{m_1m_2})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_4})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_6})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_8})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_{10}})_\alpha{}^\beta\,, \tag{A.13}$$

$$(\gamma^{m_1m_2})^\beta{}_\alpha\,, \quad (\gamma^{m_1\ldots m_4})^\beta{}_\alpha\,, \quad (\gamma^{m_1\ldots m_6})^\beta{}_\alpha\,, \quad (\gamma^{m_1\ldots m_8})^\beta{}_\alpha\,, \quad (\gamma^{m_1\ldots m_{10}})^\beta{}_\alpha\,, \tag{A.14}$$

and are related by (anti)symmetry, i.e. the matrices in (A.14) can be rewritten in terms of the matrices in (A.13), see below. Ranks $1, 3, 5, 7, 9$ give rise to spinor indices of alike chiralities

$$\begin{gathered}(\gamma^{m_1})_{\alpha\beta}\,, \quad (\gamma^{m_1m_2m_3})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_5})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_7})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_9})_{\alpha\beta}\,, \\ (\gamma^{m_1})^{\alpha\beta}, \quad (\gamma^{m_1m_2m_3})^{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_5})^{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_7})^{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_9})^{\alpha\beta}\,,\end{gathered} \tag{A.15}$$

and their symmetry properties do not mix the two lines.

The symmetry properties of the four types of matrices with respect to their spinorial indices are

$$\begin{aligned}\text{symmetric: } &(\gamma^{m_1})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_4})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_5})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_8})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_9})_{\alpha\beta}\,, \\ \text{antisymmetric: } &(\gamma^{m_1m_2})_\alpha{}^\beta\,, \quad (\gamma^{m_1m_2m_3})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_6})_\alpha{}^\beta\,, \quad (\gamma^{m_1\ldots m_7})_{\alpha\beta}\,, \quad (\gamma^{m_1\ldots m_{10}})_\alpha{}^\beta\,,\end{aligned} \tag{A.16}$$

where for example $(\gamma^{m_1m_2})_\alpha{}^\beta = -(\gamma^{m_1m_2})^\beta{}_\alpha$, and the symmetry properties of the matrices with all upper spinorial indices is the same as those with all lower indices, i.e. $(\gamma^{m_1})^{\alpha\beta} = (\gamma^{m_1})^{\beta\alpha}$ etc.

The Fierz decompositions of spinor bilinears reads

$$\begin{aligned}\psi^\alpha\chi^\beta &= \frac{1}{16}\gamma^{\alpha\beta}_{m_1}(\psi\gamma^{m_1}\chi) + \frac{1}{96}(\gamma_{m_1\ldots m_3})^{\alpha\beta}(\psi\gamma^{m_1\ldots m_3}\chi) + \frac{1}{3840}(\gamma_{m_1\ldots m_5})^{\alpha\beta}(\psi\gamma^{m_1\ldots m_5}\chi)\,, \\ \psi_\alpha\chi^\beta &= \frac{1}{16}\delta^\beta_\alpha(\psi\chi) + \frac{1}{32}(\gamma_{m_1m_2})_\alpha{}^\beta(\psi\gamma^{m_1m_2}\chi) + \frac{1}{384}(\gamma_{m_1\ldots m_4})_\alpha{}^\beta(\psi\gamma^{m_1\ldots m_4}\chi)\,.\end{aligned} \tag{A.17}$$

For anticommuting spinors θ^α and bosonic pure spinors λ^α, important special cases of (A.17) are[62]

$$\lambda^\alpha\lambda^\beta = \frac{1}{3840}(\lambda\gamma^{mnpqr}\lambda)\gamma^{\alpha\beta}_{mnpqr}\,, \qquad \theta^\alpha\theta^\beta = \frac{1}{96}(\theta\gamma^{mnp}\theta)\gamma^{\alpha\beta}_{mnp}\,. \tag{A.18}$$

[62] The first identity in (A.18) is sometimes incorrectly stated with a coefficient $\frac{1}{1920}$ rather than $\frac{1}{3840}$. To see why the latter is correct, one has to pay attention to the epsilon term in the trace (A.25) and the self-duality (A.21) of $\gamma^{mnpqr}_{\alpha\beta}$.

Appendix A.3. Duality properties

In ten-dimensional Minkowski space $\mathbb{R}^{1,9}$, using the convention

$$\epsilon_{01\dots9} = 1\,, \qquad \epsilon^{01\dots9} = -1\,, \tag{A.19}$$

with (in particular, $\epsilon^{m_1m_2\dots m_{10}}\epsilon_{m_1m_2\dots m_{10}} = -10!$)

$$\epsilon_{n_1\dots n_{10}}\epsilon^{m_1\dots m_{10}} = -10!\delta^{m_1\dots m_{10}}_{n_1\dots n_{10}}\,, \tag{A.20}$$

the antisymmetric gamma matrices (n-forms) are related by the duality properties

$$\begin{aligned}
(\gamma^{m_1\dots m_5})_{\alpha\beta} &= \frac{1}{5!}\epsilon^{m_1\dots m_5 n_1\dots n_5}(\gamma_{n_1\dots n_5})_{\alpha\beta}\,, & (\gamma^{m_1\dots m_5})^{\alpha\beta} &= -\frac{1}{5!}\epsilon^{m_1\dots m_5 n_1\dots n_5}(\gamma_{n_1\dots n_5})^{\alpha\beta}\,,\\
(\gamma^{m_1\dots m_6})_\alpha{}^\beta &= \frac{1}{4!}\epsilon^{m_1\dots m_6 n_1\dots n_4}(\gamma_{n_1\dots n_4})_\alpha{}^\beta\,, & (\gamma^{m_1\dots m_6})^\alpha{}_\beta &= -\frac{1}{4!}\epsilon^{m_1\dots m_6 n_1\dots n_4}(\gamma_{n_1\dots n_4})^\alpha{}_\beta\,,\\
(\gamma^{m_1\dots m_7})_{\alpha\beta} &= -\frac{1}{3!}\epsilon^{m_1\dots m_7 n_1\dots n_3}(\gamma_{n_1\dots n_3})_{\alpha\beta}\,, & (\gamma^{m_1\dots m_7})^{\alpha\beta} &= \frac{1}{3!}\epsilon^{m_1\dots m_7 n_1\dots n_3}(\gamma_{n_1\dots n_3})^{\alpha\beta}\,,\\
(\gamma^{m_1\dots m_8})_\alpha{}^\beta &= -\frac{1}{2!}\epsilon^{m_1\dots m_8 n_1 n_2}(\gamma_{n_1 n_2})_\alpha{}^\beta\,, & (\gamma^{m_1\dots m_8})^\alpha{}_\beta &= \frac{1}{2!}\epsilon^{m_1\dots m_8 n_1 n_2}(\gamma_{n_1 n_2})^\alpha{}_\beta\,,\\
(\gamma^{m_1\dots m_9})_{\alpha\beta} &= \epsilon^{m_1\dots m_9 n_1}(\gamma_{n_1})_{\alpha\beta}\,, & (\gamma^{m_1\dots m_9})^{\alpha\beta} &= -\epsilon^{m_1\dots m_9 n_1}(\gamma_{n_1})^{\alpha\beta}\,,\\
(\gamma^{m_1\dots m_{10}})_\alpha{}^\beta &= \epsilon^{m_1\dots m_{10}}\delta^\beta_\alpha\,, & (\gamma^{m_1\dots m_{10}})^\alpha{}_\beta &= -\epsilon^{m_1\dots m_{10}}\delta^\alpha_\beta\,.
\end{aligned} \tag{A.21}$$

A good exercise is to check them explicitly using the $SO(1,9)$ parameterization in (A.4). In ten-dimensional Euclidean space where $\epsilon^{12\dots10} = \epsilon_{12\dots10} = 1$, the equations above are still valid after redefining the Levi–Civita epsilon tensor $\epsilon \to i\epsilon$.

Appendix A.4. Traces of gamma matrices

In ten dimensions there are no invariant tensors with k antisymmetrized vector indices except when $k = 10$, so all the k-forms with even $2 \le k \le 8$ are traceless,

$$(\gamma^{m_1m_2\dots m_k})^\alpha{}_\alpha = 0\,, \quad k = 2,4,6,8\,. \tag{A.22}$$

When $k = 10$ compatibility with the duality conditions (A.21) implies

$$(\gamma^{m_1\dots m_{10}})_\alpha{}^\alpha = 16\epsilon^{m_1\dots m_{10}}\,, \quad (\gamma^{m_1\dots m_{10}})^\alpha{}_\alpha = -16\epsilon^{m_1\dots m_{10}}\,. \tag{A.23}$$

The trace relations for the $1,2,3,4$ and 5 forms are given by [445]

$$\begin{aligned}
\mathrm{Tr}(\gamma^{m_1}\gamma_{n_1}) &= (\gamma^{m_1})_{\alpha\beta}(\gamma_{n_1})^{\beta\alpha} = 16\delta^{m_1}_{n_1}\,, \qquad (A.24)\\
\mathrm{Tr}(\gamma^{m_1m_2}\gamma_{n_1n_2}) &= (\gamma^{m_1m_2})_\alpha{}^\beta(\gamma_{n_1n_2})_\beta{}^\alpha = -16\cdot 2!\delta^{m_1m_2}_{n_1n_2}\,,\\
\mathrm{Tr}(\gamma^{m_1\dots m_3}\gamma_{n_1\dots n_3}) &= (\gamma^{m_1\dots m_3})_{\alpha\beta}(\gamma_{n_1\dots n_3})^{\beta\alpha} = -16\cdot 3!\delta^{m_1\dots m_3}_{n_1\dots n_3}\,,\\
\mathrm{Tr}(\gamma^{m_1\dots m_4}\gamma_{n_1\dots n_4}) &= (\gamma^{m_1\dots m_4})_\alpha{}^\beta(\gamma_{n_1\dots n_4})_\beta{}^\alpha = 16\cdot 4!\delta^{m_1\dots m_4}_{n_1\dots n_4}\,,\\
\mathrm{Tr}(\gamma^{m_1\dots m_5}\gamma_{n_1\dots n_5}) &= (\gamma^{m_1\dots m_5})_{\alpha\beta}(\gamma_{n_1\dots n_5})^{\beta\alpha} = 16\cdot 5!\delta^{m_1\dots m_5}_{n_1\dots n_5} + 16\epsilon^{m_1\dots m_5}{}_{n_1\dots n_5}\,, \qquad (A.25)
\end{aligned}$$

where $\mathrm{Tr}(\gamma^M\gamma^N) = \gamma^M_{\alpha\beta}(\gamma^N)^{\beta\alpha}$ or $\mathrm{Tr}(\gamma^M\gamma^N) = (\gamma^M)_\alpha{}^\beta(\gamma^N)_\beta{}^\alpha$ depending on the lengths of the multi-indices M and N. We also note the vanishing of the traces with unequal lengths of M and N

$$\mathrm{Tr}(\gamma^{m_1\dots m_i}\gamma_{n_1\dots n_j}) = 0\,, \quad i \neq j\,. \tag{A.26}$$

These identities can be conveniently summarized using word notation as[63]

$$\mathrm{Tr}\big(\gamma^P\gamma_Q\big) = 16\delta^{p,q}\Big[p!\delta^P_{\tilde{Q}} + \delta^{p,5}\epsilon^P{}_Q\Big]\,, \quad |P| = p, \quad |Q| = q\,. \tag{A.27}$$

[63]The reversal $\tilde{Q}$ in (A.27) is explained by noting $[mn] = -[nm]$, $[mnp] = -[pnm]$. In general, $[P] = \begin{cases} [\tilde{P}] & : p = 0,1 \bmod 4 \\ -[\tilde{P}] & : p = 2,3 \bmod 4 \end{cases}$.

Appendix A.5. Products of gamma matrices

The antisymmetrized products of gamma matrices form a basis in the space of bispinor indices, as evidenced by the Fierz identities. In order to freely move between upstairs and downstairs indices with the Euclidean metric, we consider the Clifford algebra after a Wick rotation, $\{\Gamma^m, \Gamma^n\} = 2\delta^{mn}$. Since in the pure spinor formalism it is convenient to consider a Weyl representation leading to off-diagonal γ^m matrices as in (A.2), the Clifford algebra for the 32×32 matrices Γ^m reduces to

$$\{\gamma^m, \gamma^n\} = 2\delta^{mn} \tag{A.28}$$

in terms of 16×16 chiral gamma matrices. The explicit construction of such matrices can be found in Appendix B.1.

We now want to convert products of gamma matrices into sums over antisymmetrized gammas in the spinorial index basis. The starting point is the Clifford algebra (A.28) which implies

$$\gamma^m \gamma^n = \gamma^{mn} + \delta^{mn} \,. \tag{A.29}$$

This formula can be used iteratively when more indices are present, but the amount of generated terms grows quickly when doing so. General formulae and strategies to handle the combinatorics exist in the literature. For instance, a general formula for the product $\gamma_{m_1 \ldots m_i} \gamma^{n_1 \cdots n_j}$ has been written in [446] using a diagrammatic method, while in [447] an OPE-like algorithm was presented. A nice formula was given in [443] with the combinatorics conveniently organized as (note that the convention here has $1/k!$ in $[a_1 \ldots a_k]$)

$$\gamma_{a_1 a_2 \ldots a_p} \gamma^{b_1 b_2 \ldots b_q} = \sum_{k=0}^{\min(p,q)} k! \binom{p}{k} \binom{q}{k} \delta^{[b_1}_{[a_p} \delta^{b_2}_{a_{p-1}} \cdots \delta^{b_k}_{a_{p-k+1}} \gamma_{a_1 \ldots a_{p-k}]}{}^{b_{k+1} \ldots b_q]} \,, \tag{A.30}$$

where all the signs in the sum (prior to antisymmetrization) are uniformly positive due to the reverse ordering chosen for some indices on the right-hand side. This formula can be further decluttered using a notation based on words. If we adopt the convention where a lower case letter corresponding to the word denotes the length of the word, $|A| := a$, we can rewrite (A.30) more compactly as (here we have $1/k!$ in $[a_1 \ldots a_k]$)

$$\gamma_A \gamma^B = \sum_{\substack{XY=A \\ ZW=B \\ z=y}} y! \binom{a}{y} \binom{b}{y} \delta^{[Z}_{[\tilde{Y}} \gamma_{X]}{}^{W]} \,, \tag{A.31}$$

where $\tilde{Y}$ denotes the reversal of Y and we note the constraint $y = z$ on the lengths of Y and Z due to the generalized Kronecker delta. The combinatorial coefficients compensate the overall $1/(a!b!)$ due to the antisymmetrizations over the A and B indices and the normalization of the generalized Kronecker delta (A.9) such that in the expanded result all terms have a ± 1 coefficient, as follows from the iterated use of (A.29). For example,

$$\begin{aligned} \gamma_{a_1 a_2} \gamma^{b_1 b_2} &= \gamma_{a_1 a_2}{}^{b_1 b_2} + 4 \delta^{[b_1}_{[a_2} \gamma_{a_1]}{}^{b_2]} + 2 \delta^{b_1 b_2}_{a_2 a_1} \\ &= \gamma_{a_1 a_2}{}^{b_1 b_2} + \delta^{b_1}_{a_2} \gamma_{a_1}{}^{b_2} - \delta^{b_2}_{a_2} \gamma_{a_1}{}^{b_1} + \delta^{b_2}_{a_1} \gamma_{a_2}{}^{b_1} - \delta^{b_1}_{a_1} \gamma_{a_2}{}^{b_2} + \delta^{b_1}_{a_2} \delta^{b_2}_{a_1} - \delta^{b_2}_{a_2} \delta^{b_1}_{a_1} \,. \end{aligned} \tag{A.32}$$

Another example, which when fully expanded generates 136 terms in total, is given by

$$\gamma_{a_1 a_2 a_3 a_4 a_5} \gamma^{b_1 b_2 b_3} = \gamma_{a_1 a_2 a_3 a_4 a_5}{}^{b_1 b_2 b_3} + 15 \delta^{[b_1}_{[a_5} \gamma_{a_1 a_2 a_3 a_4]}{}^{b_2 b_3]} + 60 \delta^{[b_1 b_2}_{[a_5 a_4} \gamma_{a_1 a_2 a_3]}{}^{b_3]} + 60 \delta^{b_1 b_2 b_3}_{[a_5 a_4 a_3} \gamma_{a_1 a_2]} \,. \tag{A.33}$$

The different coefficients in front of each term correspond to the numbers of terms (with ± 1 coefficients) that are generated once the explicit antisymmetrization takes place (note $136 = 1 + 15 + 60 + 60$).

Related formulas for the commutator of gamma matrices can be found in [448].

Appendix A.6. Gamma matrix identities and pure spinors

A set of frequently used identities when manipulating pure spinor superspace expressions is listed below (repeated indices are contracted):

$$
\begin{aligned}
\gamma^m_{\alpha(\beta}\gamma^m_{\gamma\delta)} &= 0\,, \\
\gamma^{mnp}_{\alpha[\beta}\gamma^{mnp}_{\gamma\delta]} &= 0\,, \\
\gamma^{\alpha\beta}_{mnp}\gamma^{mnp}_{\gamma\delta} &= 48(\delta^\alpha_\gamma\delta^\beta_\delta - \delta^\beta_\gamma\delta^\alpha_\delta)\,, \\
\gamma^{mnp}_{\alpha\beta}\gamma^{mnp}_{\gamma\delta} &= 12(\gamma^m_{\alpha\delta}\gamma^m_{\beta\gamma} - \gamma^m_{\alpha\gamma}\gamma^m_{\beta\delta})\,, \\
\gamma^m_{\alpha\beta}\gamma^m_{\delta\sigma} &= -\frac{1}{2}\gamma^m_{\alpha\delta}\gamma^m_{\beta\sigma} - \frac{1}{24}\gamma^{mnp}_{\alpha\delta}\gamma^{mnp}_{\beta\sigma}\,, \\
\gamma^{mnp}_{\alpha\beta}\gamma^{mnp}_{\delta\sigma} &= -18\gamma^m_{\alpha\delta}\gamma^m_{\beta\sigma} + \frac{1}{2}\gamma^{mnp}_{\alpha\delta}\gamma^{mnp}_{\beta\sigma}\,, \\
(\gamma^{mn})_\alpha{}^\delta(\gamma_{mn})_\beta{}^\sigma &= -8\delta^\sigma_\alpha\delta^\delta_\beta - 2\delta^\delta_\alpha\delta^\sigma_\beta + 4\gamma^m_{\alpha\beta}\gamma^{\delta\sigma}_m\,, \\
(\gamma^{mnpq})_\alpha{}^\beta(\gamma_{mnpq})_\sigma{}^\delta &= 315\delta^\delta_\alpha\delta^\beta_\sigma + \frac{21}{2}(\gamma^{mn})_\alpha{}^\delta(\gamma_{mn})_\sigma{}^\beta + \frac{1}{8}(\gamma^{mnpq})_\alpha{}^\delta(\gamma_{mnpq})_\sigma{}^\beta\,, \\
(\gamma^{mnpq})_\alpha{}^\beta(\gamma_{mnpq})_\sigma{}^\delta &= -48\delta^\beta_\alpha\delta^\delta_\sigma + 288\delta^\delta_\alpha\delta^\beta_\sigma + 48\gamma^m_{\alpha\sigma}\gamma^{\beta\delta}_m\,.
\end{aligned}
\tag{A.34}
$$

They can be derived by using that the gamma matrices form a complete basis for the spinorial indices, see [449] for more examples.

Some of these identities are particularly useful when contracted with pure spinors. The first identity of (A.34), for instance, implies

$$(\lambda\gamma^m)_\alpha(\lambda\gamma_m)_\beta = 0\,. \tag{A.35}$$

To see this it suffices to contract $\gamma^m_{\alpha(\beta}\gamma^m_{\gamma\delta)} = 0$ with $\lambda^\alpha\lambda^\gamma$ and use the symmetry (A.16) of the one-form $\gamma^m_{\beta\gamma} = \gamma^m_{\gamma\beta}$ to obtain $(\lambda\gamma^m)_\beta(\lambda\gamma_m)_\delta = -(\lambda\gamma^m)_\delta(\lambda\gamma_m)_\beta - (\lambda\gamma^m\lambda)\gamma^m_{\delta\beta} = -(\lambda\gamma^m)_\delta(\lambda\gamma_m)_\beta = 0$ based on the pure spinor constraint (3.26). An important corollary of (A.35) is

$$(\lambda\gamma_m)_\alpha(\lambda\gamma^{mnpqr}\lambda) = 0\,, \tag{A.36}$$

which can be proven by decomposing the five-form using (A.31) as $(\lambda\gamma^{mnpqr}\lambda) = (\lambda\gamma^m\gamma^{npqr}\lambda) + (\lambda\gamma^{npq}\lambda)\delta^{mr} - (\lambda\gamma^{npr}\lambda)\delta^{mq} + (\lambda\gamma^{nqr}\lambda)\delta^{mp} - (\lambda\gamma^{pqr}\lambda)\delta^{mn}$ and observing that all the three-forms contracted with two pure spinors vanish by the antisymmetry (A.16). We can thus rewrite $(\lambda\gamma_m)_\alpha(\lambda\gamma^{mnpqr}\lambda) = (\lambda\gamma_m)_\alpha(\lambda\gamma^m\gamma^{npqr}\lambda)$ which vanishes by (A.35).

Appendix B. The $U(5)$ decomposition of $SO(10)$

In this appendix we will list some of the formulae relevant for the decomposition of various $SO(10)$ representations into their $U(5)$ components. Some useful references are [450] and the appendices of [451, 452, 453].

Appendix B.1. The Clifford algebra in $\mathbb{R}^{10}$

For convenience, we will consider the Wick-rotated version $SO(10)$ of the Lorentz group $SO(1,9)$. The ten-dimensional Clifford algebra in Euclidean signature

$$\{\Gamma^m, \Gamma^n\} = 2\delta^{mn}\,, \qquad m,n = 1,2,\ldots 10 \tag{B.1}$$

admits a recursive construction [454] starting from the 2×2 representation in terms of Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\,, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,, \tag{B.2}$$

satisfying $\{\sigma_i, \sigma_j\} = 2\delta_{ij}$ for $i, j = 1, 2, 3$. To assemble the explicit $2^5 \times 2^5$ gamma matrices Γ^m in ten dimensions we use the Kronecker product of Pauli sigma matrices as follows [455] ($\mathbb{1} := 1_{2\times 2}$; also see (A.4) for the analogous numerical representation of Γ^m for Minkowski spacetime $\mathbb{R}^{1,9}$):

$$\begin{aligned}
\Gamma^1 &= \sigma_2 \otimes \sigma_1 \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1}\,, & \Gamma^6 &= \sigma_2 \otimes \sigma_2 \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1}\,,\\
\Gamma^2 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_1 \otimes \mathbb{1} \otimes \mathbb{1}\,, & \Gamma^7 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_2 \otimes \mathbb{1} \otimes \mathbb{1}\,,\\
\Gamma^3 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_1 \otimes \mathbb{1}\,, & \Gamma^8 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_2 \otimes \mathbb{1}\,,\\
\Gamma^4 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_1\,, & \Gamma^9 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_2\,,\\
\Gamma^5 &= \sigma_2 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_3\,, & \Gamma^{10} &= -\sigma_1 \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1}\,.
\end{aligned} \tag{B.3}$$

The properties of the Kronecker product, $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ and $(A \otimes B)^T = A^T \otimes B^T$, imply that the Clifford algebra (B.1) is satisfied. Moreover, the symmetry properties of gamma matrices are

$$\Gamma_m^T = \begin{cases} -\Gamma_m\,, & m = 1, \ldots, 5 \\ +\Gamma_m\,, & m = 6, \ldots, 10\,, \end{cases} \tag{B.4}$$

while at the same time the gamma matrices are purely imaginary for $m = 1, \ldots, 5$ and real for $m = 6, \ldots, 10$, as they are constructed with an odd or even number of σ_2. This means that the representation (B.3) is hermitian

$$\Gamma_m^\dagger = \Gamma_m\,. \tag{B.5}$$

Charge conjugation and chirality matrices. Given the above symmetry property of Γ_m, the charge conjugation matrix C satisfying

$$C\Gamma_m = -\Gamma_m^T C \tag{B.6}$$

is obtained by the product of all antisymmetric Γ's [456]

$$C = \Gamma_1\Gamma_2\Gamma_3\Gamma_4\Gamma_5 = -\sigma_2 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_1 \otimes \sigma_2\,. \tag{B.7}$$

It is easy to see that C is antisymmetric, off-diagonal and satisfies $C^2 = 1_{32\times 32}$. In addition, the chirality matrix is given by

$$\Gamma_{11} = -i\Gamma_1 \ldots \Gamma_{10} = \sigma_3 \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} = \begin{pmatrix} 1_{16\times 16} & 0 \\ 0 & -1_{16\times 16} \end{pmatrix}, \tag{B.8}$$

and has the same numerical value as in the Lorentzian version (A.7).

Appendix B.2. Vectors and Lorentz generators

The generators M^{mn} of the group $SO(10)$ are antisymmetric $M^{mn} = -M^{nm}$ and satisfy the Lie-algebra relations

$$[M^{mn}, M^{pq}] = \delta^{mp}M^{nq} - \delta^{np}M^{mq} - \delta^{mq}M^{np} + \delta^{nq}M^{mp}\,. \tag{B.9}$$

The vector V^p and spinor Ψ representations of $SO(10)$ are defined by the following transformations[64]

$$[M^{mn}, V^p] = \delta^{mp}V^n - \delta^{np}V^m\,, \tag{B.10}$$

[64]Note the consistency between (B.9) and (B.10) as the transformation (B.9) can be viewed as the two-form version of (B.10) using $[M^{mn}, V^r \otimes V^s] = [M^{mn}, V^r] \otimes V^s + V^r \otimes [M^{mn}, V^s]$ and setting $M^{pq} = V^p \otimes V^q - V^q \otimes V^p$. The sign on the right-hand side of (B.11) may naively appear to conflict with the fact that the Lorentz algebra (B.9) is obeyed by Gamma matrices in the normalization of $M^{mn} \to -\frac{1}{2}\Gamma^{mn}$. However, repeated Lorentz transformations lead to Gamma matrices in the opposite multiplication order $[M^{pq}, [M^{mn}, \Psi]] = \frac{1}{4}(\Gamma^{mn}\Gamma^{pq})\Psi$ such that $[[M^{pq}, M^{mn}], \Psi] = \frac{1}{4}[\Gamma^{mn}, \Gamma^{pq}]\Psi = -\frac{1}{2}(\delta^{p[m}\Gamma^{n]q} - \delta^{q[m}\Gamma^{n]p})\Psi$, and the normalization on the right-hand side of (B.11) is determined by consistency with (B.9). In case of the pure spinor ghost $\Psi \to \lambda^\alpha$ and its Lorentz current $M^{mn} \to N^{mn}$, a more detailed version of this calculation can be found in (3.34), where (B.11) is implemented through the OPE (3.32).

$$[M^{mn}, \Psi] = \frac{1}{2}\Gamma^{mn}\Psi\,. \tag{B.11}$$

To decompose the vectorial representation $SO(10) \to U(5)$ we shall split the ten components of the $SO(10)$ vector V^m with $m = 1, \dots, 10$ into two vectors v^a, v_a labelled by an index $a = 1, \dots, 5$ as

$$v^a = \frac{1}{\sqrt{2}}\big(V^a + iV^{a+5}\big)\,, \qquad v_a = \frac{1}{\sqrt{2}}\big(V^a - iV^{a+5}\big)\,, \quad a = 1, \dots, 5\,. \tag{B.12}$$

Consequently, the components of tensors of $SO(10)$ are split following the tensor products of the vector decompositions (B.12) with the corresponding symmetry conditions. This implies, for example, that the generators M^{mn} split as follows:

$$\begin{aligned} m^{ab} &= \frac{1}{2}\left(M^{ab} + iM^{a(b+5)} + iM^{(a+5)b} - M^{(a+5)(b+5)}\right)\,, \\ m_{ab} &= \frac{1}{2}\left(M^{ab} - iM^{a(b+5)} - iM^{(a+5)b} - M^{(a+5)(b+5)}\right)\,, \\ m^a_b &= \frac{1}{2}\left(M^{ab} - iM^{a(b+5)} + iM^{(a+5)b} + M^{(a+5)(b+5)}\right)\,. \end{aligned} \tag{B.13}$$

Moreover, the trace of m^a_b is given by

$$m = \sum_{a=1}^{5} m^a_a = i\sum_{a=1}^{5} M^{(a+5)a}\,. \tag{B.14}$$

From the above it follows that

$$\begin{aligned} [m^{ab}, v^c] &= 0\,, & [m_{ab}, v_c] &= 0\,, \\ [m_{ab}, v^c] &= \delta^c_a v_b - \delta^c_b v_a\,, & [m^{ab}, v_c] &= \delta^a_c v^b - \delta^b_c v^a\,, \\ [m^a_b, v^c] &= -\delta^c_b v^a\,, & [m^a_b, v_c] &= \delta^a_c v_b\,, \\ [m, v^c] &= -v^c\,, & [m, v_c] &= v_c\,. \end{aligned} \tag{B.15}$$

To derive the above commutators, the decompositions (B.13) and (B.12) have been used to rewrite them in $SO(10)$ language, where (B.10) can be applied with its outcome cast back in $U(5)$ variables using (B.12).

Similarly, the same strategy shows that the $SO(10)$ Lie algebra (B.9) decomposes to $U(5)$ as

$$\begin{aligned} [m_{ab}, m_{cd}] &= 0\,, & [m^{ab}, m^{cd}] &= 0\,, \\ [m_{ab}, m^{cd}] &= -\delta^c_a m^d_b + \delta^d_a m^c_b + \delta^c_b m^d_a - \delta^d_b m^c_a\,, & [m_{ab}, m^c_d] &= -\delta^c_b m_{ad} + \delta^c_a m_{bd}\,, \\ [m^{ab}, m^c_d] &= -\delta^a_d m^{bc} + \delta^b_d m^{ac}\,, & [m^a_b, m^c_d] &= -\delta^c_b m^a_d + \delta^a_d m^c_b\,, \\ [m, m_{ab}] &= 2m_{ab}\,, & [m, m^{ab}] &= -2m^{ab}\,, \\ [m, m^a_b] &= 0\,, & [m, m] &= 0\,. \end{aligned} \tag{B.16}$$

This shows that m^a_b are the generators of $U(5)$ embedded in $SO(10)$. Moreover m^{ab} and m_{ab} transform as two-forms under $U(5)$, and v_a, v^a transform in the defining representations $\mathbf{5}$ and $\overline{\mathbf{5}}$ of $U(5)$. The trace m is the $U(1)$ generator in the decomposition $U(5) = SU(5) \otimes U(1)$. The $U(1)$ charge q_R of a representation R is defined by $[m, R] = q_R R$ and denoted by a subscript $\overline{\mathbf{N}}_{q_R}$ for an N-dimensional representation of $SU(5)$. We conclude that the vector V^m and the antisymmetric tensor M^{mn} transform as follows under the $SU(5) \otimes U(1)$ decomposition of $SO(10)$,

$$\begin{aligned} V^m &\to v^a \oplus v_a\,, & M^{mn} &\to m^{ab} \oplus m_{ab} \oplus m^a_b \oplus m\,, \\ \mathbf{10} &\to \mathbf{5}_{-1} \oplus \overline{\mathbf{5}}_1\,, & \mathbf{45} &\to \mathbf{10}_{-2} \oplus \overline{\mathbf{10}}_2 \oplus \mathbf{24}_0 \oplus \mathbf{1}_0\,. \end{aligned} \tag{B.17}$$

Decomposition of the Lorentz currents OPE. In the pure spinor formalism the $SO(10)$ to $U(5)$ decomposition must be applied to the OPE between the Lorentz generators for the pure spinor variables,

$$N^{mn}(z)N^{pq}(w) \sim \frac{\delta^{mp}N^{nq}(w) - \delta^{np}N^{mq}(w) - \delta^{mq}N^{np}(w) + \delta^{nq}N^{mp}(w)}{z-w} - 3\frac{\left(\delta^{mq}\delta^{np} - \delta^{mp}\delta^{nq}\right)}{(z-w)^2}\,. \tag{B.18}$$

Starting from the generators in (B.13) and (B.14), we perform the redefinitions

$$n = \frac{m}{\sqrt{5}}\,, \qquad n^a_b = m^a_b - \frac{1}{5}\delta^a_b m\,, \tag{B.19}$$

which identify n^a_b as the traceless generator of $SU(5)$. Using the same strategy as above leads to the following OPEs among the $SU(5)\otimes U(1)$ decompositions

$$\begin{aligned}
n_{ab}(z)n_{cd}(w) &\sim \text{regular}\,, & n^{ab}(z)n^{cd}(w) &\sim \text{regular}\,, \\
n_{ab}(z)n^{cd}(w) &\sim \frac{-\delta^c_{[a}n^d_{b]} + \delta^d_{[a}n^c_{b]} - \frac{2}{\sqrt{5}}\delta^c_{[a}\delta^d_{b]}n}{z-w} - 3\frac{\delta^c_b\delta^d_a - \delta^c_a\delta^d_b}{(z-w)^2}\,, & n(z)n^a_b(w) &\sim \text{regular}\,, \\
n^a_b(z)n^c_d(w) &\sim \frac{-\delta^c_b n^a_d + \delta^a_d n^c_b}{z-w} - 3\frac{\delta^a_d\delta^c_b - \frac{1}{5}\delta^a_b\delta^c_d}{(z-w)^2}\,, & n(z)n_{ab}(w) &\sim \frac{2}{\sqrt{5}}\frac{n_{ab}}{z-w}\,, \\
n^{ab}(z)n^c_d(w) &\sim \frac{-\delta^a_d n^{bc} + \delta^b_d n^{ac} - \frac{2}{5}\delta^c_d n^{ab}}{z-w}\,, & n(z)n^{ab}(w) &\sim -\frac{2}{\sqrt{5}}\frac{n^{ab}}{z-w}\,, \\
n_{ab}(z)n^c_d(w) &\sim \frac{-\delta^c_b n_{ad} + \delta^c_a n_{bd} + \frac{2}{5}\delta^c_d n_{ab}}{z-w}\,, & n(z)n(w) &\sim -\frac{3}{(z-w)^2}\,.
\end{aligned} \tag{B.20}$$

Redefining the $U(1)$ charge to $[n,R] = \frac{q_R}{\sqrt{5}}R$ in view of (B.19) we see that $(n, n^a_b, n^{ab}, n_{ab})$ transform as the $(\mathbf{1}_0, \mathbf{24}_0, \mathbf{10}_{-2}, \overline{\mathbf{10}}_2)$ representations of $SU(5)\otimes U(1)$.

Appendix B.3. Spinors

To obtain the decomposition of the spinorial representation of $SO(10)$ under $SU(5)\otimes U(1)$ it will be convenient to consider the linear combinations [450]

$$b^a = \frac{1}{2}\left(\Gamma^a + i\Gamma^{a+5}\right)\,, \qquad b_a = \frac{1}{2}\left(\Gamma^a - i\Gamma^{a+5}\right)\,, \tag{B.21}$$

where $a = 1, 2, \ldots, 5$. The Clifford algebra (B.1) implies the fermionic oscillator algebra

$$\{b_a, b^b\} = \delta^b_a\,, \qquad \{b_a, b_b\} = \{b^a, b^b\} = 0\,. \tag{B.22}$$

This means that the matrices b_a and b^b can be interpreted as annihilation and creation operators. To exploit this interpretation we define a vacuum $|0\rangle$ annihilated by all the b_a operators, $b_a|0\rangle := 0$ (also $\langle 0|\, b^a := 0$) and normalized as $\langle 0|0\rangle = 1$. States are created by acting with the creation operators b^a on the vacuum, for a maximum of 32 states. We will also define $\langle\psi| = |\psi\rangle^T$. These operators also satisfy

$$\begin{aligned}
b_a^\dagger &= b^a\,, & (b^a)^\dagger &= b_a\,, \\
b_a^T &= -b^a\,, & (b^a)^T &= -b_a\,,
\end{aligned} \tag{B.23}$$

for $a = 1, \ldots, 5$, as can be verified from (B.4) and (B.5). In this language, the charge conjugation and chirality matrices in (B.7) and (B.8) become

$$C = \prod_{j=1}^{5}(b_j + b^j)\,, \qquad \Gamma_{11} = \prod_{j=1}^{5}(b^j b_j - b_j b^j)\,. \tag{B.24}$$

To connect this description with the $U(5)$ decomposition of $SO(10)$ above, we write the generators M^{mn} for the spinor representation as

$$M^{mn} \to -\frac{1}{2}\Gamma^{mn} = -\frac{1}{4}[\Gamma^m, \Gamma^n]\,, \tag{B.25}$$

which satisfy the commutation relations (B.9). Therefore, from the expressions (B.13) and (B.14), the $U(5)$ Lorentz generators become

$$\begin{aligned} m_b^a &= -\frac{1}{2}\big(b^a b_b - b_b b^a\big)\,, & m &= -\frac{1}{2}\big(b^a b_a - b_a b^a\big) = -b^a b_a + \frac{5}{2}\,, \\ m^{ab} &= -b^a b^b, & m_{ab} &= -b_a b_b\,. \end{aligned} \tag{B.26}$$

These expressions can be verified by plugging the spinorial representation (B.25) into the decompositions (B.13) and using the inverse of (B.21). In this language, it is straightforward to verify the decompositions (B.16) using $[b_b b^a, b_d b^c] = [b_b b^a, b_d]b^c + b_d[b_b b^a, b^c]$ and $[b_d, b_b b^a] = \{b_d, b_b\}b^a - b_b\{b_d, b^a\}$. Furthermore,

$$\begin{aligned} [m_b^a, b^c] &= -\delta_b^c b^a\,, & [m_b^a, b_c] &= \delta_c^a b_b\,, \\ [m^{ab}, b^c] &= 0\,, & [m^{ab}, b_c] &= \delta_c^a b^b - \delta_c^b b^a\,, \\ [m_{ab}, b^c] &= \delta_a^c b_b - \delta_b^c b_a\,, & [m_{ab}, b_c] &= 0\,, \\ [m, b^c] &= -b^c\,, & [m, b_c] &= b_c\,. \end{aligned} \tag{B.27}$$

These relations identify m_b^a as the generators of $U(5)$ and m as the generator of $U(1)$. Moreover (B.27) implies that b^c and b_c transform in the $\mathbf{5}_{-1}$ and $\overline{\mathbf{5}}_1$ representations of $SU(5) \otimes U(1)$, respectively.

Pure spinors. Recall that a ten-dimensional pure spinor was defined by Cartan as a bosonic Weyl spinor Λ satisfying the equation [104]

$$\Lambda^T C \Gamma^m \Lambda = 0\,, \tag{B.28}$$

where C is the 32×32 charge conjugation matrix (B.7) and $\Gamma_{11}\Lambda = -\Lambda$. From (B.8) we obtain $\Lambda^T = (\lambda\ \ 0)$ for a 16 dimensional bosonic spinor λ^α, and this implies the familiar equation $\lambda^\alpha \gamma^m_{\alpha\beta}\lambda^\beta = 0$. We will now describe the pure spinor constraint using the creation and annihilation operators (B.21). To do this, first we will need to characterize a Weyl spinor in this language.

Lemma 14. *The $16 \oplus 16'$ states of ten-dimensional Weyl $|\lambda\rangle$ and anti-Weyl $|\Omega\rangle$ spinors satisfying $\Gamma_{11}|\lambda\rangle = -|\lambda\rangle$ and $\Gamma_{11}|\Omega\rangle = |\Omega\rangle$ are created by*

$$\begin{aligned} |\lambda\rangle &= \lambda_+ |0\rangle + \frac{1}{2}\lambda_{ab} b^b b^a |0\rangle + \frac{1}{4!}\lambda^a \epsilon_{abcde} b^e b^d b^c b^b |0\rangle\,, \\ |\Omega\rangle &= \frac{1}{5!}\omega_+ \varepsilon_{abcde} b^a b^b b^c b^d b^e |0\rangle + \frac{1}{2!3!}\omega^{ab}\varepsilon_{abcde} b^c b^d b^e |0\rangle + \omega_a b^a |0\rangle\,. \end{aligned} \tag{B.29}$$

These expressions correspond to the following representation decompositions of $SO(10) \to U(5)$:

$$\begin{aligned} \lambda^\alpha &\to (\lambda^+, \lambda_{ab}, \lambda^a)\,, & \omega_\alpha &\to (\omega_+, \omega^{ab}, \omega_a)\,, \\ \mathbf{16} &\to (\mathbf{1}, \overline{\mathbf{10}}, \mathbf{5})\,, & \mathbf{16}' &\to (\mathbf{1}, \mathbf{10}, \overline{\mathbf{5}})\,. \end{aligned} \tag{B.30}$$

Proof. The chirality matrix in terms of the creation and annihilation algebra is given by (B.8) such that $\Gamma_{11}|0\rangle = -|0\rangle$ and $\{\Gamma_{11}, b^a\} = 0$. This means that states with an even (odd) number of creation operators acting on the vacuum have eigenvalue -1 ($+1$) under Γ_{11}. This explains the expressions (B.29). The number of independent components of each $U(5)$ representation in (B.30) follows easily from the fermionic nature of the creation operators b^a as $\#(b^{a_1}\ldots b^{a_k}) = \binom{5}{k}$. □

Note that the $U(5)$ components of the Weyl and anti-Weyl spinors can be extracted as

$$\lambda^+ = \langle 0|\lambda\rangle\,, \qquad \lambda_{ab} = \langle 0|\, b_a b_b\, |\lambda\rangle\,, \qquad \lambda^a = \frac{1}{4!}\epsilon^{abcde}\langle 0|\, b_b b_c b_d b_e\, |\lambda\rangle\,,$$

$$\omega_+ = \frac{1}{5!}\epsilon^{abcde} \langle 0| b_a b_b b_c b_d b_e |\omega\rangle \ , \qquad \omega^{ab} = -\frac{1}{3!}\epsilon^{abcde} \langle 0| b_c b_d b_e |\omega\rangle \ , \qquad \omega_a = \langle 0| b_a |\omega\rangle \ . \tag{B.31}$$

In order to obtain the number of degrees of freedom of a ten-dimensional pure spinor we will need the following results

$$\langle 0| Cb^a b^b b^c b^d b^e |0\rangle = \epsilon^{abcde} \, , \qquad Cb_a = b^a C \, , \quad Cb^a = b_a C \, , \tag{B.32}$$

which can be obtained from (B.8) and (B.22) together with the normalization $\langle 0|0\rangle = 1$. Using the Weyl spinor decomposition (B.29) one can also show

$$\begin{aligned} \langle 0| Cb^a |\lambda\rangle &= \lambda^a \, , \\ \langle 0| Cb^a b^b b^c |\lambda\rangle &= -\frac{1}{2}\epsilon^{abcde}\lambda_{de} \, , \\ \langle 0| Cb^a b^b b^c b^d b^e |\lambda\rangle &= \epsilon^{abcde}\lambda^+ \, . \end{aligned} \tag{B.33}$$

We are now ready to show

Proposition 22. *A ten-dimensional pure spinor has eleven complex degrees of freedom.*

Proof. Under the decomposition of $SO(10) \to U(5)$, the constraint (B.28) generates two sets of independent equations (with $i = 1, 2, 3, 4, 5$ each)[65]:

$$\langle\lambda| Cb^i |\lambda\rangle = 0 \, , \tag{B.34}$$

$$\langle\lambda| Cb_i |\lambda\rangle = 0 \, . \tag{B.35}$$

The transpose relation $(b^i)^T = -b_i$ in (B.22) implies $\langle\lambda| = \langle 0| \lambda_+ + \frac{1}{2} \langle 0| b_a b_b \lambda_{ab} + \frac{1}{24} \langle 0| b_b b_c b_d b_e \lambda^a \epsilon_{abcde}$. So the equation (B.34) becomes

$$\begin{aligned} 0 = \langle\lambda| Cb^p |\lambda\rangle &= \lambda^+ \langle 0| Cb^p |\lambda\rangle + \frac{1}{2}\lambda_{ij} \langle 0| b_i b_j Cb^p |\lambda\rangle + \frac{1}{24}\lambda^i \epsilon_{ijklm} \langle 0| b_j b_k b_l b_m Cb^p |\lambda\rangle \\ &= 2\lambda^+\lambda^a - \frac{1}{4}\epsilon^{abcde}\lambda_{bc}\lambda_{de} \, , \end{aligned} \tag{B.36}$$

where we used $b_a C = Cb^a$ from (B.32) and (B.33). Hence, (B.36) implies that we can write the five components λ^a in terms of the others

$$\lambda^a = \frac{1}{8\lambda^+}\epsilon^{abcde}\lambda_{bc}\lambda_{de} \, , \tag{B.37}$$

which solves the first equation (B.34). Moreover, the second equation (B.35) yields

$$\langle 0| Cb_b |0\rangle = 2\lambda^a\lambda_{ab} \, , \tag{B.38}$$

which is automatically satisfied when inserting the solution (B.37) due to an over-antisymmetrization of five-dimensional indices. Therefore the pure spinor constraint in ten dimensions removes only the five components (B.37) from the 16-component Weyl spinor, leaving a total of eleven degrees of freedom. ☐

Spinorial transformations in $U(5)$ language. In the fundamentals of the pure spinor formalism it is necessary to know how the $U(5)$ components of the pure spinor transform under the $SO(10)$ rotations. To do this we note the interpretation [450] $O\,|v\rangle = |Ov\rangle$ for an arbitrary operator O that allows one to read off how the different tensor components transform under O. Straightforward calculations using the operators (B.26) imply that the right-hand side of the spinorial transformation (B.11), given by the action of $\frac{1}{2}\Gamma^{mn} = -M^{mn}$, decomposes as follows

$$m_{ab} |\lambda\rangle = \lambda_{ab} |0\rangle + \frac{1}{2}\epsilon_{abcde}\lambda^e b^d b^c |0\rangle \ ,$$

[65] The constraint (B.28) for both $\Gamma^1 = b^1 + b_1$ and $\Gamma^6 = -i(b^1 - b_1)$ implies $\langle\lambda| Cb_1 |\lambda\rangle = 0$ and $\langle\lambda| Cb^1 |\lambda\rangle = 0$ from suitable linear combinations.

$$m^{ab}\,|\lambda\rangle = \lambda^+ b^a b^b\,|0\rangle + \frac{1}{2}\lambda_{cd} b^a b^b b^d b^c\,|0\rangle\,, \tag{B.39}$$
$$m^a_b\,|\lambda\rangle = -\frac{1}{2}\delta^a_b\,|\lambda\rangle + \lambda_{cb} b^a b^c\,|0\rangle + \frac{1}{3!}\lambda^c\epsilon_{cdefb} b^a b^f b^e b^d\,|0\rangle\,,$$
$$m\,|\lambda\rangle = -\frac{5}{2}\lambda_+\,|0\rangle - \frac{1}{4}\lambda_{ab} b^b b^a\,|0\rangle + \frac{1}{16}\lambda^a\epsilon_{abcde} b^e b^d b^c b^b\,|0\rangle\,,$$

where we used that $\frac{1}{4!}\epsilon_{bnpqr}\lambda^a b^r b^q b^p b^n\,|0\rangle = \frac{1}{3!}\lambda^c\epsilon_{cdefb} b^a b^f b^e b^d\,|0\rangle$ due to the restricted range 1 to 5 of the indices. Note the factor of $|\lambda\rangle$ instead of $|0\rangle$ in the first term of (B.39).

Using the projections to the $U(5)$ components (B.31) these transformations imply

$$\begin{aligned}
&m_{ab}\lambda^+ = \lambda_{ab}\,, && m_{ab}\lambda_{cd} = \epsilon_{abcde}\lambda^e\,, && m_{ab}\lambda^c = 0\,,\\
&m^{ab}\lambda^+ = 0\,, && m^{ab}\lambda_{cd} = (\delta^a_d\delta^b_c - \delta^a_c\delta^b_d)\lambda^+\,, && m^{ab}\lambda^c = -\frac{1}{2}\epsilon^{abcde}\lambda_{de}\,,\\
&m^a_b\lambda^+ = -\frac{1}{2}\delta^a_b\lambda^+\,, && m^a_b\lambda_{cd} = \delta^a_d\lambda_{cb} - \delta^a_c\lambda_{db} - \frac{1}{2}\delta^a_b\lambda_{cd}\,, && m^a_b\lambda^c = -\delta^c_b\lambda^a + \frac{1}{2}\delta^a_b\lambda^c\,,\\
&m\lambda^+ = -\frac{5}{2}\lambda^+\,, && m\lambda_{cd} = -\frac{1}{2}\lambda_{cd}\,, && m\lambda^c = +\frac{3}{2}\lambda^c\,.
\end{aligned} \tag{B.40}$$

For example, $|m\lambda\rangle_{ab} := \langle 0|\, b_a b_b m\,|\lambda\rangle = -\frac{1}{4}\lambda_{fg}\,\langle 0|\, b_a b_b b^g b^f\,|0\rangle = -\frac{1}{2}\lambda_{ab}$, where $|m\lambda\rangle_{ab}$ denotes the projection of $|m\lambda\rangle$ into its **10** component of $SU(5)$.

After identifying the $SU(5)\otimes U(1)$ Lorentz currents with a traceless n^a_b as

$$(n, n^a_b, n^{ab}, n_{ab}) = \left(\frac{m}{\sqrt{5}}, m^a_b - \frac{1}{5}\delta^a_b m, m^{ab}, m_{ab}\right), \tag{B.41}$$

we arrive at the following $SO(10)\to SU(5)\otimes U(1)$ decompositions

$$\begin{aligned}
&n_{ab}\lambda^+ = \lambda_{ab}\,, && n_{ab}\lambda_{cd} = \epsilon_{abcde}\lambda^e\,, && n_{ab}\lambda^c = 0\,,\\
&n^{ab}\lambda^+ = 0\,, && n^{ab}\lambda_{cd} = -\delta^{[a}_c\delta^{b]}_d\lambda^+\,, && n^{ab}\lambda^c = -\frac{1}{2}\epsilon^{abcde}\lambda_{de}\,,\\
&n^a_b\lambda^+ = 0\,, && n^a_b\lambda_{cd} = \delta^a_d\lambda_{cb} - \delta^a_c\lambda_{db} - \frac{2}{5}\delta^a_b\lambda_{cd}\,, && n^a_b\lambda^c = -\delta^c_b\lambda^a + \frac{1}{5}\delta^a_b\lambda^c\,,\\
&n\lambda^+ = -\frac{\sqrt{5}}{2}\lambda^+\,, && n\lambda_{cd} = -\frac{1}{2\sqrt{5}}\lambda_{cd}\,, && n\lambda^c = \frac{3}{2\sqrt{5}}\lambda^c\,.
\end{aligned} \tag{B.42}$$

These are the coefficients of the single pole in the OPE $N^{mn}\lambda^\alpha$ in (3.32).

To find the $U(1)$ charges of the anti-Weyl spinor we compute

$$m\,|\Omega\rangle = \frac{1}{48}\omega_+\varepsilon_{abcde} b^a b^b b^c b^d b^e\,|0\rangle + \frac{1}{24}\omega^{ab}\varepsilon_{abcde} b^c b^d b^e\,|0\rangle - \frac{3}{2}\omega_a b^a\,|0\rangle \tag{B.43}$$

which implies, after projecting to the components via (B.31),

$$m\omega_+ = \frac{5}{2}\omega_+\,, \qquad m\omega^{ab} = \frac{1}{2}\omega^{ab}\,, \qquad m\omega_a = -\frac{3}{2}\omega_a\,. \tag{B.44}$$

Appendix C. Combinatorics on words

In this appendix we list some of the most common maps on words used throughout this review. With the exception of the letterification defined in [85], these definitions are standard and can be found in the books [457, 143].

The left-to-right bracketing map $\ell(A)$, also called the Dynkin bracket, is defined recursively by

$$\ell(123..n) := \ell(123...n-1)n - n\ell(123...n-1)\,, \quad \ell(i) = i\,, \quad \ell(\emptyset) = 0\,. \tag{C.1}$$

For example,

$$\begin{aligned}\ell(12) &= 12 - 21\,,\\ \ell(123) &= 123 - 213 - 312 + 321\,.\end{aligned}\tag{C.2}$$

In addition, the map $\rho(A)$ is defined by

$$\rho(123\ldots n) := \rho(123\ldots n{-}1)n - \rho(23\ldots n)1\,.\tag{C.3}$$

For example,

$$\begin{aligned}\rho(12) &= 12 - 21\,,\\ \rho(123) &= 123 - 213 - 231 + 321\,.\end{aligned}\tag{C.4}$$

The *shuffle product* $\sqcup\!\sqcup$ is defined recursively by

$$\emptyset \sqcup\!\sqcup P = P \sqcup\!\sqcup \emptyset := P\,, \qquad iP \sqcup\!\sqcup jQ := i(P \sqcup\!\sqcup jQ) + j(Q \sqcup\!\sqcup iP)\,,\tag{C.5}$$

where i and j are letters, P and Q are words while $\emptyset$ represents the empty word. For example,

$$\begin{gathered}1 \sqcup\!\sqcup 2 = 12 + 21\,, \quad 1 \sqcup\!\sqcup 23 = 123 + 213 + 231\,,\\ 12 \sqcup\!\sqcup 34 = 1234 + 1324 + 1342 + 3124 + 3142 + 3412\,.\end{gathered}\tag{C.6}$$

The *deshuffle* $\delta(P) = X \otimes Y$ of P (sometimes denoted as $P = X \sqcup\!\sqcup Y$) is the sum of all pairs of words X, Y such that P is in the shuffle of X and Y. An efficient algorithm to obtain the words X and Y in $\delta(P) = X \otimes Y$ is given by [143]

$$\delta(a_1a_2\ldots a_n) := \delta(a_1)\delta(a_2)\ldots\delta(a_n)\,, \quad \delta(a_i) := \emptyset \otimes a_i + a_i \otimes \emptyset\,, \quad \delta(\emptyset) := \emptyset \otimes \emptyset\,,\tag{C.7}$$

where the product is defined as

$$(A \otimes B)(R \otimes S) := (AR \otimes BS)\,.\tag{C.8}$$

For example,

$$\begin{aligned}\delta(1) &= \emptyset \otimes 1 + 1 \otimes \emptyset\,,\\ \delta(12) &= \delta(1)\delta(2) = (\emptyset \otimes 1 + 1 \otimes \emptyset)(\emptyset \otimes 2 + 2 \otimes \emptyset) = \emptyset \otimes 12 + 1 \otimes 2 + 2 \otimes 1 + 12 \otimes \emptyset\,,\\ \delta(123) &= \delta(12)\delta(3) = (\emptyset \otimes 12 + 1 \otimes 2 + 2 \otimes 1 + 12 \otimes \emptyset)(\emptyset \otimes 3 + 3 \otimes \emptyset)\\ &= \emptyset \otimes 123 + 1 \otimes 23 + 2 \otimes 13 + 12 \otimes 3 + 3 \otimes 12 + 13 \otimes 2 + 23 \otimes 1 + 123 \otimes \emptyset\,.\end{aligned}\tag{C.9}$$

An alternative characterization of the deshuffle map is

$$\delta(P) = \sum_{X,Y} \langle P, X \sqcup\!\sqcup Y\rangle\, X \otimes Y\,,\tag{C.10}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product on words[66]

$$\langle A, B\rangle := \delta_{A,B}\,, \qquad \delta_{A,B} = \begin{cases} 1\,, & \text{if } A = B\\ 0\,, & \text{otherwise}\,.\end{cases}\tag{C.11}$$

In addition, the *letterfication* maps a *word* Q to a *letter* $\dot{q}$,

$$Q \to \dot{q}\,.\tag{C.12}$$

[66]Not to be confused with the pure spinor zero-mode measure.

The purpose of this map is to avoid deconcatenation of $\dot{q}$ since a letter can not be deconcatenated. For example, suppose that the word $Q = 12$ has been letterified to $\dot{q} = 12$ and that $P = 3$. Then deconcatenating QP is different from deconcatenating $\dot{q}P$. For example, one gets only one term $\sum_{XY=\dot{q}P} S_X T_Y = S_{\dot{q}} T_3 = S_{12}T_3$ instead of the usual two ($\sum_{XY=\dot{Q}P} S_X T_Y = S_1 T_{23} + S_{12}T_3$) if Q is not letterified.

In the proof of (4.132) we used a result that was already implicit in the proof of the shuffle symmetry of Berends–Giele currents in the appendix A of [84]:

Lemma 15. *For A and B non-empty words*

$$\Delta(A \sqcup\!\sqcup B) = \emptyset \otimes (A \sqcup\!\sqcup B) + (A \sqcup\!\sqcup B) \otimes \emptyset + A \otimes B + B \otimes A + \sum_{PQ=A}{}' \sum_{RS=B}{}' (P \sqcup\!\sqcup R) \otimes (Q \sqcup\!\sqcup S)\,, \tag{C.13}$$

where $\sum'$ denotes deconcatenation over non-empty words.

Proof. The deconcatenation coproduct $\Delta(P) = \sum_{XY=P} X \otimes Y$ is a homomorphism with respect to the shuffle product, $\Delta(A \sqcup\!\sqcup B) = \Delta(A) \sqcup\!\sqcup \Delta(B)$ (see Proposition 1.9 of [143]). Noting that $(P \otimes Q) \sqcup\!\sqcup (R \otimes S) = (P \sqcup\!\sqcup R) \otimes (Q \sqcup\!\sqcup S)$ we get for $A, B \neq \emptyset$

$$\begin{aligned}\Delta(A \sqcup\!\sqcup B) = \Delta(A) \sqcup\!\sqcup \Delta(B) &= \sum_{PQ=A} \sum_{RS=B} (P \otimes Q) \sqcup\!\sqcup (R \otimes S) \\ &= \emptyset \otimes (A \sqcup\!\sqcup B) + (A \sqcup\!\sqcup B) \otimes \emptyset + A \otimes B + B \otimes A + \sum_{PQ=A}{}' \sum_{RS=B}{}' (P \sqcup\!\sqcup R) \otimes (Q \sqcup\!\sqcup S)\,,\end{aligned} \tag{C.14}$$

where the first four terms in the second line come from separating off the empty words in the sums such that the deconcatenation words in $\sum'$ are not empty. $\square$

Appendix C.1. The dual Lie polynomials

The dual Lie polynomials in $\mathcal{L}^*$ are characterized by the dual basis iQ satisfying

$$\langle iQ, \ell(iP) \rangle = \delta_{Q,P}\,, \tag{C.15}$$

where $\ell(iP)$ is the Lyndon basis of Lie polynomials when i is the minimum letter of iQ. Given a dual Lie polynomial P^* and a Lie polynomial P, their expansions in their respective bases are given by [458]

$$P^* = \sum_Q \langle P^*, \ell(iQ) \rangle iQ\,, \tag{C.16}$$

$$P = \sum_Q \langle P, iQ \rangle \ell(iQ)\,. \tag{C.17}$$

Using Ree's theorem (4.134) it is easy to see that dual Lie polynomials are unchanged by proper shuffles and therefore define equivalence classes $P^* + \sum R \sqcup\!\sqcup S \sim P^*$. For related work, see [459] and also [180].

Appendix D. Dynkin labels of $SO(10)$

In this appendix we will very briefly summarize the representation theory of $SO(10)$ in the language of Dynkin labels that was used in the main text. The practical calculations were done using [146]. For the mathematical background, see [460, 461].

An irreducible representation of the Lie algebra of $SO(10)$ is labelled by five indices $(a_1a_2a_3a_4a_5)$ characterizing its highest-weight vector. For instance a scalar of $SO(10)$ is represented by (00000), while a vector is the (10000); see the table below for more examples. The dimension of the representation labelled by $(a_1 \ldots a_5)$ is given by [462]

$$87091200 \dim(a_1a_2a_3a_4a_5) = (1+a_1)(1+a_2)(1+a_3)(1+a_4)(1+a_5)(2+a_1+a_2)(2+a_2+a_3)$$

Dynkin label	$SO(10)$ content
(00000)	0-form A
(10000)	1-form A^m
(01000)	2-form $A^{[mn]}$
(00100)	3-form $A^{[mnp]}$
(00011)	4-form $A^{[mnpq]}$
$(00020) \oplus (00002)$	5-form $A^{[mnpqr]}$
(00010)	anti-Weyl spinor ψ_α
(00001)	Weyl spinor ψ^α
$(0000n)$	$\lambda^{\alpha_1}\lambda^{a_2}\dots\lambda^{\alpha_n}$ pure spinors

$$
\begin{aligned}
&\times (2+a_3+a_4)(2+a_3+a_5)(3+a_1+a_2+a_3)(3+a_2+a_3+a_4)(3+a_2+a_3+a_5)\\
&\times (3+a_3+a_4+a_5)(4+a_1+a_2+a_3+a_4)(4+a_1+a_2+a_3+a_5)\\
&\times (4+a_2+a_3+a_4+a_5)(5+a_1+a_2+a_3+a_4+a_5)(5+a_2+2a_3+a_4+a_5)\\
&\times (6+a_1+a_2+2a_3+a_4+a_5)(7+a_1+2a_2+2a_3+a_4+a_5)\,.
\end{aligned} \tag{D.1}
$$

For example, dim(00000) = 1, dim(10000) = 10, and dim(00001) = 16. From the formula above it is easy to be convinced that these calculations are better handled by computers, see [146, 462].

Many calculations in this review require to know the decomposition of product representations. A common example is the familiar fact that two vectors decompose into a symmetric and traceless, antisymmetric and trace parts; $V^mW^n = \frac{1}{2}(V^mW^n + V^nW^m - \frac{\delta^{mn}}{5}V\cdot W) + \frac{1}{2}(V^mW^n - V^mW^n) + \frac{1}{10}\delta^{mn}V\cdot W$. In terms of the Dynkin labels, this is represented by

$$
(10000)\otimes(10000) = (20000)\oplus(01000)\oplus(00000)\,, \tag{D.2}
$$

where (20000) is the symmetric traceless and (01000) is the antisymmetric part. The dimensions match as $10\times 10 = 54+45+1$.

Of special importance for us is the pure spinor representation. A single pure spinor λ^α is a Weyl spinor (00001), but a product of n pure spinors $\lambda^{\alpha_1}\lambda^{\alpha_2}\dots\lambda^{\alpha_n}$ is $(0000n)$. The dimensions of the pure spinor representation $\lambda^n = (0000n)$ are $16, 126, 672, 2772, 9504, 28314, \dots$ for $n = 1, 2, 3, 4, 5, 6, \dots$.

Appendix E. Pure spinor superspace correlators

The result (3.94) of Lemma 2 guarantees that any pure spinor superspace expression with three pure spinors and five thetas can be reduced to the unique scalar proportional to $(\lambda^3\theta^5)$ with coefficients carrying the tensorial structure whose normalization is uniquely fixed by the condition (3.79). Therefore one can assemble a catalog of correlators using symmetry arguments alone. For instance, (in contrast to the rest of this work, the antisymmetrization brackets $[m_1m_2\dots m_k]$ enclosing k indices here include $1/k!$, e.g. $V^{[m}W^{n]} = \frac{1}{2}(V^mW^n - V^nW^m)$)

$$
\begin{aligned}
\langle(\lambda\gamma^m\theta)(\lambda\gamma^s\theta)(\lambda\gamma^u\theta)(\theta\gamma_{fgh}\theta)\rangle &= 24\delta^{msu}_{fgh}\,, \qquad\qquad \text{(E.1)}\\
\langle(\lambda\gamma_m\theta)(\lambda\gamma_s\theta)(\lambda\gamma^{ptu}\theta)(\theta\gamma_{fgh}\theta)\rangle &= \frac{288}{7}\delta^{[p}_{[m}\eta_{s][f}\delta^t_g\delta^{u]}_{h]}\,,\\
\langle(\lambda\gamma_m\theta)(\lambda\gamma^{nrs}\theta)(\lambda\gamma^{ptu}\theta)(\theta\gamma_{fgh}\theta)\rangle &= \frac{12}{35}\epsilon_{fghm}{}^{nprstu} + \frac{144}{7}\Big[\delta^{[n}_m\delta^r_{[f}\eta^{s][p}\delta^t_g\delta^{u]}_{h]} - \delta^{[p}_m\delta^t_{[f}\eta^{u][n}\delta^r_g\delta^{s]}_{h]}\Big]\\
&\quad - \frac{72}{7}\Big[\eta_{m[f}\eta^{v[p}\delta^t_g\eta^{u][n}\delta^r_{h]}\delta^{s]}_v - \eta_{m[f}\eta^{v[n}\delta^r_g\eta^{s][p}\delta^t_{h]}\delta^{u]}_v\Big]\,,\\
\langle(\lambda\gamma^{mnpqr}\theta)(\lambda\gamma_d\theta)(\lambda\gamma_e\theta)(\theta\gamma_{fgh}\theta)\rangle &= -\frac{480}{7}\big(\delta^{mnpqr}_{defgh} - \frac{1}{120}\epsilon^{mnpqr}{}_{defgh}\big)\,,\\
\langle(\lambda\gamma^{mnpqr}\theta)(\lambda\gamma_{stu}\theta)(\lambda\gamma^v\theta)(\theta\gamma_{fgh}\theta)\rangle &= \frac{576}{7}\eta^{v[m}\delta^n_{[s}\delta^p_t\eta_{u][f}\delta^q_g\delta^{r]}_{h]} - \frac{1152}{7}\delta^{[m}_{[s}\delta^n_t\delta^p_{u]}\delta^q_{[f}\delta^{r]}_g\delta^v_{h]}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{120}\epsilon^{mnpqr}{}_{abcde}\left(\frac{576}{7}\eta^{v[a}\delta^{b}_{[s}\delta^{c}_{t}\eta_{u][f}\delta^{d}_{g}\delta^{e]}_{h]}-\frac{1152}{7}\delta^{[a}_{[s}\delta^{b}_{t}\delta^{c}_{u]}\delta^{d}_{[f}\delta^{e]}_{g}\delta^{v}_{h]}\right)\,,\\
\langle(\lambda\gamma^{mnp}\theta)(\lambda\gamma_{qrs}\theta)(\lambda\gamma_{tuv}\theta)(\theta\gamma^{ijk}\theta)\rangle &= -\frac{1728}{35}\Big[\delta^{[i}_{a}\delta^{j}_{[t}\delta^{k]}_{u}\delta^{[m}_{v]}\delta^{n}_{[q}\delta^{p]}_{r}\delta^{a}_{s]}-\delta^{[i}_{a}\delta^{j}_{[q}\delta^{k]}_{r}\delta^{[m}_{s]}\delta^{n}_{[t}\delta^{p]}_{u}\delta^{a}_{v]}+\delta^{[i}_{[q}\delta^{j}_{r}\eta^{k][m}\eta_{s][t}\delta^{n}_{u}\delta^{p]}_{v]}\\
&+\delta^{a}_{[t}\eta^{b[i}\delta^{j}_{u}\eta^{k][m}\eta_{v][q}\delta^{n}_{r}\eta_{s]a}\delta^{p]}_{b}-\delta^{a}_{[q}\eta^{b[i}\delta^{j}_{r}\eta^{k][m}\eta_{s][t}\delta^{n}_{u}\eta_{v]a}\delta^{p]}_{b}-\delta^{[i}_{[t}\delta^{j}_{u}\eta^{k][m}\eta_{v][q}\delta^{n}_{r}\delta^{p]}_{s]}\Big]\,,
\end{aligned}
$$

where δ^{msu}_{fgh} is the antisymmetrized combination of Kronecker deltas beginning with $\frac{1}{3!}\delta^m_f\delta^s_g\delta^u_h$, see (A.9). To justify the first line of (E.1), note that its right-hand side is the only tensor antisymmetric both in $[msu]$ and $[fgh]$ and which is normalized to yield 2880 upon contraction (because $\delta^{msu}_{msu}=120$, see (A.11)), therefore respecting the normalization (3.80). The other identities are justified by similar means.

Basis of zero-mode correlators. Using the Fierz identities (A.18) appropriately, all pure spinor superspace expressions can be written as a linear combination of the following three correlators [378, 463]

$$
\begin{aligned}
\langle(\lambda\gamma^{mnpqr}\lambda)(\lambda\gamma^{s}\theta)(\theta\gamma_{fgh}\theta)(\theta\gamma_{jkl}\theta)\rangle &= \frac{1152}{7}\delta^m_f\delta^n_g\delta^h_j\delta^p_k\delta^q_l\delta^r_s-\frac{2304}{7}\delta^m_f\delta^n_g\delta^p_h\delta^q_j\delta^r_k\delta^s_l\\
&+\frac{1}{120}\epsilon^{mnpqr}{}_{abcde}\Big[\frac{1152}{7}\delta^a_f\delta^b_g\delta^h_j\delta^c_k\delta^d_l\delta^e_s-\frac{2304}{7}\delta^a_f\delta^b_g\delta^c_h\delta^d_j\delta^e_k\delta^s_l\Big]\\
&+[mnpqr][fgh][jkl]+(fgh\leftrightarrow jkl)\,,\\
\langle(\lambda\gamma^{mnpqr}\lambda)(\lambda\gamma^{stu}\theta)(\theta\gamma_{fgh}\theta)(\theta\gamma_{jkl}\theta)\rangle &= \frac{6917}{7}\Big[\delta^v_s\delta^t_f\delta^m_u\delta^n_g\delta^h_j\delta^p_k\delta^q_l\delta^r_v-\delta^s_f\delta^t_g\delta^m_u\delta^n_h\delta^p_j\delta^q_k\delta^r_l\Big]\\
&+\frac{3456}{1125}\epsilon^{mnpqr}{}_{abcde}\Big[\delta^v_s\delta^t_f\delta^a_u\delta^b_g\delta^h_j\delta^c_k\delta^d_l\delta^e_v-\delta^s_f\delta^t_g\delta^a_u\delta^b_h\delta^c_j\delta^d_k\delta^e_l\Big]\\
&+[mnpqr][stu][fgh][jkl]+(fgh\leftrightarrow jkl)\,,\\
\langle(\lambda\gamma^{mnpqr}\lambda)(\lambda\gamma^{stuvx}\theta)(\theta\gamma^{fgh}\theta)(\theta\gamma^{jkl}\theta)\rangle &= 2880\Big[\frac{8}{7}\delta^m_s\delta^n_t\delta^p_f\delta^q_g\delta^r_h\delta^u_j\delta^v_k\delta^x_l-\frac{8}{7}\delta^m_s\delta^n_t\delta^p_u\delta^q_f\delta^r_g\delta^v_j\delta^x_k\delta^h_l\\
&+\frac{16}{7}\delta^m_s\delta^n_t\delta^p_u\delta^q_f\delta^r_j\delta^v_g\delta^x_k\delta^h_l-\frac{24}{7}\delta^m_s\delta^n_t\delta^p_f\delta^q_g\delta^r_j\delta^u_h\delta^v_k\delta^x_l\Big]\\
&+24\epsilon^{mnpqr}{}_{abcde}\Big[\frac{8}{7}\delta^a_s\delta^b_t\delta^c_f\delta^d_g\delta^e_h\delta^u_j\delta^v_k\delta^x_l-\frac{8}{7}\delta^a_s\delta^b_t\delta^c_u\delta^d_f\delta^e_g\delta^v_j\delta^x_k\delta^h_l\\
&+\frac{16}{7}\delta^a_s\delta^b_t\delta^c_u\delta^d_f\delta^e_j\delta^v_g\delta^x_k\delta^h_l-\frac{24}{7}\delta^a_s\delta^b_t\delta^c_f\delta^d_g\delta^e_j\delta^u_h\delta^v_k\delta^x_l\Big]\\
&+[mnpqr][stuvx][fgh][jkl]+(fgh\leftrightarrow jkl)\,,
\end{aligned}
\tag{E.2}
$$

where the notation $+[i_1\ldots i_k]\ldots$ instructs to antisymmetrize the indices $i_1,\ldots,i_k$ including the normalization $1/k!$. In the correlations above we obtained the epsilon terms by considering the duality (A.21) of the five-form gamma matrix, explaining their relative factors of the form (parity even) $+\frac{1}{120}\epsilon_{10}$(parity even).

Suppose one wants to compute the following pure spinor correlator

$$
\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\gamma^{rs}\theta)(\lambda\gamma^p\gamma^{tu}\theta)(\theta\gamma_{fgh}\theta)\rangle\,.
\tag{E.3}
$$

Using the gamma-matrix identity $\gamma^m\gamma^{np}=\gamma^{mnp}+\eta^{mn}\gamma^p-\eta^{mp}\gamma^n$, we obtain a linear combinations of correlators present in the catalog above:

$$
\begin{aligned}
\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\gamma^{rs}\theta)(\lambda\gamma^p\gamma^{tu}\theta)(\theta\gamma_{fgh}\theta)\rangle &= \langle(\lambda\gamma^m\theta)(\lambda\gamma^{nrs}\theta)(\lambda\gamma^{ptu}\theta)(\theta\gamma_{fgh}\theta)\rangle\\
&+2\langle(\lambda\gamma^m\theta)\eta^{n[r}(\lambda\gamma^{s]}\theta)(\lambda\gamma^{ptu}\theta)(\theta\gamma_{fgh}\theta)\rangle\\
&+2\langle(\lambda\gamma^m\theta)(\lambda\gamma^{nrs}\theta)\eta^{p[t}(\lambda\gamma^{u]}\theta)(\theta\gamma_{fgh}\theta)\rangle\\
&+4\langle(\lambda\gamma^m\theta)\eta^{n[r}(\lambda\gamma^{s]}\theta)\eta^{p[t}(\lambda\gamma^{u]}\theta)(\theta\gamma_{fgh}\theta)\rangle\,.
\end{aligned}
\tag{E.4}
$$

Proceeding in this way we can quickly calculate any zero-mode correlator, and a `FORM` implementation can be found in [153].

Practicalities of pure spinor superspace component expansions. By virtue of Fierz identities, all possible pure spinor superspace expressions can be written in the basis (E.2) of three fundamental zero-mode correlators: $(\lambda\gamma^{[5]}\lambda)(\lambda\gamma^{[n]}\theta)(\theta\gamma^{[3]}\theta)(\theta\gamma^{[3]}\theta)$ with $n=1,3$ or 5, where the notation $\gamma^{[n]}$ for an integer n means an antisymmetric gamma matrix with n vectorial indices. The explicit form of this basis can be found in (E.2).

However, it is often more efficient to assemble beforehand a catalog of common correlators and use them out of storage rather than performing the Fierz manipulations to go to the above basis. This avoids wasteful manipulations that dramatically simplify in the end, such as computing the simple correlator (3.95) via

$$\langle(\lambda\gamma^m\theta)(\lambda\gamma^n\theta)(\lambda\gamma^p\theta)(\theta\gamma_{abc}\theta)\rangle = \frac{1}{96}\langle(\lambda\gamma^{mrstn}\lambda)(\lambda\gamma^p\theta)(\theta\gamma_{abc}\theta)(\theta\gamma_{rst}\theta)\rangle\,, \tag{E.5}$$

which, as can be seen in the expression (E.2), leads to many intermediate terms.

Another approach is to evaluate the correlators of three lambdas and five thetas by brute-force in terms of the tensor [464, 463]

$$\langle\lambda^{\alpha_1}\lambda^{\alpha_2}\lambda^{\alpha_3}\theta^{\delta_1}\theta^{\delta_2}\theta^{\delta_3}\theta^{\delta_4}\theta^{\delta_5}\rangle := T^{\alpha_1\alpha_2\alpha_3;\delta_1\delta_2\delta_3\delta_4\delta_5} = \frac{1}{1792}\gamma_m^{\alpha_1\delta_1}\gamma_n^{\alpha_2\delta_2}\gamma_p^{\alpha_3\delta_3}\gamma_{mnp}^{\delta_4\delta_5} + [\delta_1\delta_2\delta_3\delta_4\delta_5] \tag{E.6}$$

where $+[\delta_1\ldots\delta_5]$ instructs to antisymmetrize over the indices $\delta_1,\ldots,\delta_5$ including the normalization factor $1/5!$[67]. Note that the right-hand side of (E.6) is found to be symmetric in $(\alpha_1\alpha_2\alpha_3)$ after taking the antisymmetrizations over $\delta_1,\ldots,\delta_5$ and m,n,p into account. In addition, it is straightforward to see that $T^{\alpha\beta\gamma;\delta_1\delta_2\delta_3\delta_4\delta_5}\gamma^m_{\alpha\delta_1}\gamma^n_{\beta\delta_2}\gamma^p_{\gamma\delta_3}\gamma^{mnp}_{\delta_4\delta_5} = 2880$ recovers the normalization (3.80).

Evaluating pure spinor superspace expressions using this method follows from

$$\langle\lambda^\alpha\lambda^\beta\lambda^\gamma\theta^{\delta_1}\theta^{\delta_2}\theta^{\delta_3}\theta^{\delta_4}\theta^{\delta_5}f_{\alpha\beta\gamma\delta_1\delta_2\delta_3\delta_4\delta_5}(e,\chi,k)\rangle = T^{\alpha\beta\gamma;\delta_1\delta_2\delta_3\delta_4\delta_5}f_{\alpha\beta\gamma\delta_1\delta_2\delta_3\delta_4\delta_5}(e,\chi,k)\,, \tag{E.7}$$

but this usually leads to the calculation of many gamma-matrix traces, often with many free indices. While there are closed formulae for these traces in the Appendix A.1, doing these calculations on demand tends to become a time-consuming task. Therefore, to avoid any spurious inefficiencies, the catalog approach is used in the program [153]. As we have seen, pure spinor superspace expressions with many external particles can be evaluated very efficiently using multiparticle superfields in the Harnad–Shnider gauge.

Appendix F. θ-expansion of SYM superfields

A convenient gauge choice to expand the superfields of ten-dimensional SYM in theta is the Harnad–Shnider gauge [88],

$$\theta^\alpha\mathbb{A}^{\rm HS}_\alpha = 0\,. \tag{F.1}$$

At the linearized level, the gauge $\theta^\alpha A^i_\alpha = 0$ has been used in [90, 89] to obtain the θ-expansions (2.17) of the single-particle superfields K_i to arbitrary order. Since the recursive definition of multiparticle Berends–Giele currents $\mathcal{A}^P_\alpha$ in (4.95) quickly generates many terms, it would be expensive to follow the recursion up to single-particle level and then expand the multiparticle superfields using the Harnad–Shnider gauge fixing (F.1). Luckily it was shown in [84] that one can exploit the gauge-transformation properties of multiparticle superfields to arrive at Berends–Giele currents satisfying the Harnad–Shnider gauge (F.1).

It is easy to see that Berends–Giele currents in Lorenz gauge do not satisfy the condition (F.1), i.e. $\theta^\alpha\mathbb{A}^{\rm L}_\alpha \neq 0$. The idea is to find a non-linear gauge transformation $\mathbb{L}$

$$\mathbb{A}^{\rm HS}_\alpha = \mathbb{A}^{\rm L}_\alpha - [D_\alpha,\mathbb{L}] + [\mathbb{A}^{\rm L}_\alpha,\mathbb{L}] \tag{F.2}$$

such that $\theta^\alpha\mathbb{A}^{\rm HS}_\alpha = 0$. Assuming that the superfields have been brought to this gauge, the derivation of their θ-expansions proceeds in a similar way as in their single-particle counterpart.

[67]For definiteness, the definition (E.6) has 60 terms starting with $\frac{1}{107520}\gamma_m^{\alpha_1\delta_1}\gamma_n^{\alpha_2\delta_2}\gamma_p^{\alpha_3\delta_3}\gamma_{mnp}^{\delta_4\delta_5} - \cdots$.

We start by contracting the non-linear equations of motion (2.11) with θ^α while assuming the Harnad–Shnider gauge $\theta^\alpha \mathbb{A}_\alpha = 0$. The result [88]

$$\begin{aligned} \big(\mathcal{D}+1\big)\mathbb{A}_\beta &= (\theta\gamma^m)_\beta \mathbb{A}_m\,, \qquad & \mathcal{D}\mathbb{A}_m &= (\theta\gamma_m \mathbb{W})\,, \\ \mathcal{D}\mathbb{W}^\beta &= \frac{1}{4}(\theta\gamma^{mn})^\beta \mathbb{F}_{mn}\,, \qquad & \mathcal{D}\mathbb{F}^{mn} &= -(\mathbb{W}^{[m}\gamma^{n]}\theta) \end{aligned} \tag{F.3}$$

is most conveniently expressed in terms of the Euler operator

$$\mathcal{D} := \theta^\alpha D_\alpha = \theta^\alpha \frac{\partial}{\partial\theta^\alpha} \tag{F.4}$$

that weights the k^{th} order in θ by a factor of k. One can therefore use (F.3) to reconstruct the entire θ-expansion of all SYM superfields from their zeroth orders $\mathbb{K}(\theta=0)$,

$$\begin{aligned} [\mathbb{A}_\alpha]_k &= \frac{1}{k+1}(\theta\gamma^m)_\alpha [\mathbb{A}_m]_{k-1}\,, \qquad & [\mathbb{A}_m]_k &= \frac{1}{k}(\theta\gamma_m [\mathbb{W}]_{k-1})\,, \\ [\mathbb{W}^\alpha]_k &= \frac{1}{4k}(\theta\gamma^{mn})^\alpha [\mathbb{F}_{mn}]_{k-1}\,, \qquad & [\mathbb{F}^{mn}]_k &= -\frac{1}{k}([\mathbb{W}^{[m}]_{k-1}\gamma^{n]}\theta)\,, \end{aligned} \tag{F.5}$$

where the notation $[\ldots]_k$ instructs to only keep terms of order $(\theta)^k$ of the enclosed superfields. The analogous relations for the superfields at higher mass dimensions in (2.20) are

$$\begin{aligned} [\mathbb{W}^\alpha_m]_k &= \frac{1}{k}\bigg\{\frac{1}{4}(\theta\gamma^{pq})^\alpha [\mathbb{F}_{m|pq}]_{k-1} - (\theta\gamma_m)_\beta \sum_{l=0}^{k-1}\{[\mathbb{W}^\beta]_l, [\mathbb{W}^\alpha]_{k-l-1}\}\bigg\}\,, \\ [\mathbb{F}^{m|pq}]_k &= -\frac{1}{k}\bigg\{([\mathbb{W}^{m[p}]_{k-1}\gamma^{q]}\theta) + (\theta\gamma^m)_\alpha \sum_{l=0}^{k-1}[[\mathbb{W}^\alpha]_l, [\mathbb{F}^{pq}]_{k-l-1}]\bigg\}\,, \\ [\mathbb{W}^\alpha_{mn}]_k &= \frac{1}{k}\bigg\{\frac{1}{4}(\theta\gamma^{pq})^\alpha [\mathbb{F}_{mn|pq}]_{k-1} + (\theta\gamma_m)_\beta \sum_{l=0}^{k-1}\{[\mathbb{W}^\beta]_l, [\mathbb{W}^\alpha_n]_{k-l-1}\} \\ &\quad + (\theta\gamma_n)_\beta \sum_{l=0}^{k-1}\Big(\{[\mathbb{W}^\beta_m]_l, [\mathbb{W}^\alpha]_{k-l-1}\} + \{[\mathbb{W}^\beta]_l, [\mathbb{W}^\alpha_m]_{k-l-1}\}\Big)\bigg\}\,. \end{aligned} \tag{F.6}$$

Using the notation $\mathcal{K}_P(X,\theta) := \mathcal{K}_P(\theta)e^{k_P\cdot X}$, the recursions (F.5) and (F.6) were shown in [84] to yield the following multiparticle θ-expansions,

$$\begin{aligned} \mathcal{A}^P_\alpha(\theta) &= \frac{1}{2}(\theta\gamma_m)_\alpha \mathfrak{e}^m_P + \frac{1}{3}(\theta\gamma_m)_\alpha(\theta\gamma^m\mathcal{X}_P) - \frac{1}{32}(\theta\gamma_m)^\alpha(\theta\gamma^{mnp}\theta)\mathfrak{f}^P_{np} \\ &\quad + \frac{1}{60}(\theta\gamma_m)_\alpha(\theta\gamma^{mnp}\theta)(\mathcal{X}^P_n\gamma_p\theta) + \frac{1}{1152}(\theta\gamma_m)_\alpha(\theta\gamma^{mnp}\theta)(\theta\gamma^{pqr}\theta)\mathfrak{f}^{n|qr}_P \\ &\quad + \sum_{XY=P}[\mathcal{A}^{X,Y}_\alpha]_5 + \ldots\,, \\ \mathcal{A}^m_P(\theta) &= \mathfrak{e}^m_P + (\theta\gamma^m\mathcal{X}_P) - \frac{1}{8}(\theta\gamma^{mpq}\theta)\mathfrak{f}^{pq}_P + \frac{1}{12}(\theta\gamma^{mnp}\theta)(\mathcal{X}^n_P\gamma^p\theta) \\ &\quad + \frac{1}{192}(\theta\gamma^m{}_{nr}\theta)(\theta\gamma^r{}_{pq}\theta)\mathfrak{f}^{n|pq}_P - \frac{1}{480}(\theta\gamma^m{}_{nr}\theta)(\theta\gamma^r{}_{pq}\theta)(\mathcal{X}^{np}_P\gamma^q\theta) \\ &\quad + \sum_{XY=P}\Big([\mathcal{A}^m_{X,Y}]_4 + [\mathcal{A}^m_{X,Y}]_5\Big) + \ldots\,, \\ \mathcal{W}^\alpha_P(\theta) &= \mathcal{X}^\alpha_P + \frac{1}{4}(\theta\gamma^{mn})^\alpha\mathfrak{f}^P_{mn} - \frac{1}{4}(\theta\gamma_{mn})^\alpha(\mathcal{X}^m_P\gamma^n\theta) - \frac{1}{48}(\theta\gamma_m{}^q)^\alpha(\theta\gamma_{qnp}\theta)\mathfrak{f}^{m|np}_P \\ &\quad + \frac{1}{96}(\theta\gamma_m{}^q)^\alpha(\theta\gamma_{qnp}\theta)(\mathcal{X}^{mn}_P\gamma^p\theta) - \frac{1}{1920}(\theta\gamma_m{}^r)^\alpha(\theta\gamma_{nr}{}^s\theta)(\theta\gamma_{spq}\theta)\mathfrak{f}^{mn|pq}_P \end{aligned} \tag{F.7}$$

$$+\sum_{XY=P}\Big([\mathcal{W}^{\alpha}_{X,Y}]_3+[\mathcal{W}^{\alpha}_{X,Y}]_4+[\mathcal{W}^{\alpha}_{X,Y}]_5\Big)+\ldots\,,$$

$$\begin{aligned}\mathcal{F}^{mn}_P(\theta)&=\mathfrak{f}^{mn}_P-(\mathcal{X}^{[m}_P\gamma^{n]}\theta)+\frac{1}{8}(\theta\gamma_{pq}{}^{[m}\theta)\mathfrak{f}^{n]|pq}_P-\frac{1}{12}(\theta\gamma_{pq}{}^{[m}\theta)(\mathcal{X}^{n]p}_P\gamma^q\theta)\\&-\frac{1}{192}(\theta\gamma_{ps}{}^{[m}\theta)\mathfrak{f}^{n]p|qr}_P(\theta\gamma^s{}_{qr}\theta)+\frac{1}{480}(\theta\gamma^{[m}{}_{ps}\theta)(\mathcal{X}^{n]pq}_P\gamma^r\theta)(\theta\gamma^s{}_{qr}\theta)\\&+\sum_{XY=P}\Big([\mathcal{F}^{mn}_{X,Y}]_2+[\mathcal{F}^{mn}_{X,Y}]_3+[\mathcal{F}^{mn}_{X,Y}]_4+[\mathcal{F}^{mn}_{X,Y}]_5\Big)+\sum_{XYZ=P}[\mathcal{F}^{mn}_{X,Y,Z}]_5+\ldots\,,\end{aligned}$$

with terms of order $\theta^{\geq 6}$ in the ellipsis. The non-linearities of the form $\sum_{XY=P}[\mathcal{K}_{X,Y}]_l$ can be traced back to the quadratic expressions in (F.6), e.g.

$$\begin{aligned}[][\mathcal{A}^{X,Y}_\alpha]_5&=\frac{1}{144}(\theta\gamma_m)_\alpha(\theta\gamma^{mnp}\theta)(\mathcal{X}^X\gamma_n\theta)(\mathcal{X}^Y\gamma_p\theta)\,,\\ [\mathcal{A}^m_{X,Y}]_4&=\frac{1}{24}(\theta\gamma^m{}_{np}\theta)(\mathcal{X}^X\gamma^n\theta)(\mathcal{X}^Y\gamma^p\theta)\,,\\ [\mathcal{W}^\alpha_{X,Y}]_3&=-\frac{1}{6}(\theta\gamma_{mn})^\alpha(\mathcal{X}_X\gamma^m\theta)(\mathcal{X}_Y\gamma^n\theta)\,,\\ [\mathcal{F}^{mn}_{X,Y}]_2&=-(\mathcal{X}_X\gamma^{[m}\theta)(\mathcal{X}_Y\gamma^{n]}\theta)\,,\end{aligned}\tag{F.8}$$

and further instances as to make the complete orders $\theta^{\leq 5}$ available are spelled out in the appendix of [84]. It is easy to see that these non-linear terms vanish in the single-particle case, and one recovers the linearized expansions (2.17) of [89, 90].

Analogous θ-expansions for the superfields of higher mass dimensions start with

$$\begin{aligned}\mathcal{W}^{m\alpha}_P(X,\theta)&=e^{k_P\cdot X}\Big(\mathcal{X}^{m\alpha}_P+\frac{1}{4}(\theta\gamma_{np})^\alpha\mathfrak{f}^{m|np}_P+\sum_{XY=P}\big[(\mathcal{X}_X\gamma^m\theta)\mathcal{X}^\alpha_Y-(X\leftrightarrow Y)\big]+\ldots\Big)\,,\\ \mathcal{F}^{m|pq}_P(X,\theta)&=e^{k_P\cdot X}\Big(\mathfrak{f}^{m|pq}_P-(\mathcal{X}^{m[p}_P\gamma^{q]}\theta)+\sum_{XY=P}\big[(\mathcal{X}_X\gamma^m\theta)\mathfrak{f}^{pq}_Y-(X\leftrightarrow Y)\big]+\ldots\Big)\,,\end{aligned}\tag{F.9}$$

where the lowest two orders $\sim\theta^2,\theta^3$ in the ellipsis along with generalizations to higher mass dimensions are spelled out in the appendix of [84].

Appendix G. Redefinitions from the Lorenz gauge to the BCJ gauge

As shown in [85], multiparticle superfields in the BCJ gauge can be generated by starting from the multiparticle superfields in the Lorenz gauge defined recursively in (4.33). In contrast to the hybrid gauge discussed in section 4.1.6, the redefinitions are more involved and require the following iterated redefinition,

$$K_{[P,Q]}=L_1(\hat{K}_{[P,Q]})\,,\tag{G.1}$$

where the operator L_j is defined as the local version of the perturbiner (4.114)

$$L_j(\hat{K}_{[P,Q]}):=\hat{K}_{[P,Q]}-\frac{1}{j}\big(\hat{H}\otimes L_{(j+1)}(\hat{K})\big)_{C([P,Q])}-\frac{1}{j}\begin{cases}D_\alpha\hat{H}_{[P,Q]} & :\ K=A_\alpha\,,\\ k^m_{PQ}\hat{H}_{[P,Q]} & :\ K=A^m\,,\\ 0 & :\ K=W^\alpha\,,\end{cases}\tag{G.2}$$

where we used the notation (4.31) and $\hat{H}$ is defined below in (G.6). The action of $L_j(\hat{K}_{[P,Q]})$ gives rise to $L_{(j+1)}(\hat{K}_{[A,B]})$ on the right-hand side with $|A|+|B|<|P|+|Q|$. Therefore this is a iteration over the index j, and it eventually stops as each step involves splitting the nested brackets in $[P,Q]$. The iteration built into the redefinition (G.1) yields the infinite series of non-linear terms [177] present in the finite gauge transformation of the corresponding perturbiner series.

The examples (4.62) of redefinitions from the hybrid to BCJ gauge have the following Lorenz to BCJ counterparts:

$$
\begin{aligned}
A^m_{[1,2]} &= \hat{A}^m_{[1,2]}\,, \qquad\qquad \text{(G.3)}\\
A^m_{[[1,2],3]} &= \hat{A}^m_{[[1,2],3]} - k^m_{123}\hat{H}_{[[1,2],3]}\,,\\
A^m_{[[1,2],[3,4]]} &= \hat{A}^m_{[[1,2],[3,4]]} - (k_1\cdot k_2)\Big(\hat{H}_{[1,[3,4]]}\hat{A}^m_2 - \hat{H}_{[2,[3,4]]}\hat{A}^m_1\Big)\\
&\quad + (k_3\cdot k_4)\Big(\hat{H}_{[[1,2],4]}\hat{A}^m_3 - \hat{H}_{[[1,2],3]}\hat{A}^m_4\Big) - k^m_{1234}\hat{H}_{[[1,2],[3,4]]}\,,\\
A^m_{[[[1,2],3],4]} &= \hat{A}^m_{[[[1,2],3],4]} - (k_1\cdot k_2)\Big(\hat{H}_{[[1,3],4]}\hat{A}^m_2 - \hat{H}_{[[2,3],4]}\hat{A}^m_1\Big)\\
&\quad - (k_{12}\cdot k_3)\Big(\hat{H}_{[[1,2],4]}\hat{A}^m_3\Big) - (k_{123}\cdot k_4)\Big(\hat{H}_{[[1,2],3]}\hat{A}^m_4\Big) - k^m_{1234}\hat{H}_{[[[1,2],3],4]}\,,\\
A^m_{[[[[1,2],3],4],5]} &= \hat{A}^m_{[[[[1,2],3],4],5]} - (k_1\cdot k_2)\Big(\hat{H}_{[[1,3],4]}\hat{A}^m_{[2,5]} + \hat{H}_{[[1,3],5]}\hat{A}^m_{[2,4]} + \hat{H}_{[[1,4],5]}\hat{A}^m_{[2,3]}\\
&\quad + \hat{H}_{[[[1,3],4],5]}\hat{A}^m_2 - (1\leftrightarrow 2)\Big)\\
&\quad - (k_{12}\cdot k_3)\Big(\hat{H}_{[[1,2],4]}\hat{A}^m_{[3,5]} + \hat{H}_{[[1,2],5]}\hat{A}^m_{[3,4]} + \hat{H}_{[[[1,2],4],5]}\hat{A}^m_3 - ([1,2]\leftrightarrow 3)\Big)\\
&\quad - (k_{123}\cdot k_4)\Big(\hat{H}_{[[1,2],3]}\hat{A}^m_{[4,5]} + \hat{H}_{[[[1,2],3],5]}\hat{A}^m_4\Big)\\
&\quad - (k_{1234}\cdot k_5)\Big(\hat{H}_{[[[1,2],3],4]}\hat{A}^m_5\Big) - \hat{H}_{[[[[1,2],3],4],5]}k^m_{12345}\,,\\
A^m_{[[[1,2],3],[4,5]]} &= \hat{A}^m_{[[[1,2],3],[4,5]]} - (k_1\cdot k_2)\Big(\hat{H}_{[1,[4,5]]}\hat{A}^m_{[2,3]} + \hat{H}_{[[1,3],[4,5]]}\hat{A}^m_2 - (1\leftrightarrow 2)\Big)\\
&\quad - (k_{12}\cdot k_3)\Big(\hat{H}_{[[1,2],[4,5]]}\hat{A}^m_3 - \hat{H}_{[3,[4,5]]}\hat{A}^m_{[1,2]}\Big)\\
&\quad - (k_{123}\cdot k_{45})\Big(\hat{H}_{[[1,2],3]}\hat{A}^m_{[4,5]}\Big)\\
&\quad + (k_4\cdot k_5)\Big(\hat{H}_{[[[1,2],3],5]}\hat{A}^m_4 - \hat{H}_{[[[1,2],3],4]}\hat{A}^m_5\Big) - k^m_{12345}\hat{H}_{[[[1,2],3],[4,5]]}\,. \qquad\qquad \text{(G.4)}
\end{aligned}
$$

For an example of the redefinition (G.1) for more than one iteration of L_j, it is enough to consider the superfield $\hat{A}^m_{[[12,34],56]}$. A long and tedious calculation yields [85]

$$
\begin{aligned}
A^m_{[[12,34],56]} &= \hat{A}^m_{[[12,34],56]} - k^m_{123456}\hat{H}_{[[12,34],56]} \qquad\qquad \text{(G.5)}\\
&\quad - (k_1\cdot k_2)\Big(\hat{A}^m_2\hat{H}_{[[1,34],56]} + \hat{A}^m_{[2,34]}\hat{H}_{[1,56]} + \hat{A}^m_{[2,56]}\hat{H}_{[1,34]}\\
&\qquad\qquad - \frac{1}{2}k^m_{234}\hat{H}_{[2,34]}\hat{H}_{[1,56]} - \frac{1}{2}k^m_{256}\hat{H}_{[2,56]}\hat{H}_{[1,34]} - (1\leftrightarrow 2)\Big)\\
&\quad - (k_{12}\cdot k_{34})\Big(\hat{A}^m_{34}\hat{H}_{[12,56]} - (12\leftrightarrow 34)\Big)\\
&\quad - (k_{1234}\cdot k_{56})\hat{A}^m_{56}\hat{H}_{[12,34]}\\
&\quad - (k_3\cdot k_4)\Big(\hat{A}^m_4\hat{H}_{[123,56]} + \hat{A}^m_{[12,4]}\hat{H}_{[3,56]} + \hat{A}^m_{[4,56]}\hat{H}_{[12,3]}\\
&\qquad\qquad - \frac{1}{2}k^m_{124}\hat{H}_{[12,4]}\hat{H}_{[3,56]} - \frac{1}{2}k^m_{456}\hat{H}_{[4,56]}\hat{H}_{[12,3]} - (3\leftrightarrow 4)\Big)\\
&\quad - (k_5\cdot k_6)\Big(\hat{A}^m_6\hat{H}_{[[12,34],5]} - (5\leftrightarrow 6)\Big)\,.
\end{aligned}
$$

The factors of 1/2 correspond to the appearance of quadratic terms of the redefining superfields $H_{[A,B]}$ in the finite gauge transformation given by the infinite series (4.113).

In the above redefinitions $\hat{H}_{[P,Q]}$ is given by

$$
\hat{H}_{[A,B]} = \hat{H}'_{[A,B]} - \frac{1}{2}(\hat{H}\otimes\hat{H})_{\tilde{C}([A,B])}\,, \qquad\qquad \text{(G.6)}
$$

$$\begin{aligned}
\hat{H}'_{[A,B]} &= H_{[A,B]} - \frac{1}{2}\Big[\big(\hat{H}'_A k_A^m - (\hat{H}\otimes\hat{H}^m)_{\tilde{C}'(A)}\big)A_m^B - (A\leftrightarrow B)\Big]\,,\\
\hat{H}'_i &= \hat{H}'_{[i,j]} = 0\,,
\end{aligned}$$

where the $H_{[A,B]}$ are defined as they were in (4.63) to (4.65), and $\hat{H}_A^m := k_A^m \hat{H}_A$. Furthermore, the maps $\tilde{C}$ and $\tilde{C}'$ in the subscripts of (G.6) are variants of the contact-term map C reviewed in section 4.1.1 and introduced in [85],

$$\tilde{C}(i) = 0\,, \qquad \tilde{C}([A,B]) = [C(A),B]_r + [A,C(B)]_r\,, \tag{G.7}$$

see (4.21) for the definition of the C map on the right-hand side, and we use the notation

$$[P\otimes Q, B]_r = [P,B]\otimes Q\,. \tag{G.8}$$

The definitions in (G.6) furthermore involve the map

$$\tilde{C}'([A,B]) = \tilde{C}([A,B]) - \frac{1}{2}(k_A\cdot k_B)(A\otimes B - B\otimes A)\,. \tag{G.9}$$

In this way, iterative use of (G.6) will reduce any $\hat{H}_{[A,B]}$ to combinations of A_C^m, Mandelstam invariants and the superfields $H_{[A,B]}$ defined in (4.63) to (4.65), for instance

$$\begin{aligned}
\hat{H}_{[[[[1,2],3],[4,5]],6]} = H_{[[[[1,2],3],[4,5]],6]}& \\
&- \frac{1}{2}H_{[[[1,2],3],[4,5]]}(k_{12345}\cdot A_6) + \frac{1}{4}H_{[[1,2],3]}(k_{123}\cdot A_{45})(k_{12345}\cdot A_6)\\
&- \frac{1}{2}(k_1\cdot k_2)\big(H_{[[1,3],6]}H_{[2,[4,5]]} - H_{[[2,3],6]}H_{[1,[4,5]]}\big)\\
&- \frac{1}{2}(k_{12}\cdot k_3)\big(H_{[[1,2],6]}H_{[3,[4,5]]}\big) - \frac{1}{2}(k_{123}\cdot k_{45})\big(H_{[[4,5],6]}H_{[[1,2],3]}\big)\,.
\end{aligned} \tag{G.10}$$

Appendix H. The contact-term map is nilpotent

To show that the contact-term map in (4.21) is nilpotent[68] we will first determine the action of the adjoint C^* on $X_1\wedge X_2\wedge X_3$, where $X_i\in\mathcal{L}^*$ are dual Lie polynomials (see Appendix C.1). From the definition (4.155) we know that the adjoint of the contact-term map is the S bracket. For convenience we can use

$$\langle X_1\wedge X_2, C(\Gamma_1)\rangle = \langle C^\star(X_1\wedge X_2),\Gamma_1\rangle \tag{H.1}$$

for a Lie monomial $\Gamma_1\in\mathcal{L}$ and dual words $X_1, X_2\in\mathcal{L}^*$, and $C^\star(X_1\wedge X_2) = 2\{X_1,X_2\}$ is the S bracket. Recall that in (4.26) the contact-term map C is extended to act on the antisymmetric product of Lie polynomials $\mathcal{L}\wedge\mathcal{L}$ as a graded derivation of grading $+1$ acting on Lie polynomials Γ_i of grading $+1$,

$$C(\Gamma_1\wedge\Gamma_2) = C(\Gamma_1)\wedge\Gamma_2 - \Gamma_1\wedge C(\Gamma_2)\,. \tag{H.2}$$

To compute $C^\star(X_1\wedge X_2\wedge X_3)$ we use the definition

$$\langle C^\star(X_1\wedge X_2\wedge X_3),\Gamma_1\wedge\Gamma_2\rangle = \langle X_1\wedge X_2\wedge X_3, C(\Gamma_1\wedge\Gamma_2)\rangle \tag{H.3}$$

which exploits the fact that $C(\Gamma_1\wedge\Gamma_2)$ has grading $+3$. Using (H.2) we get

$$\langle C^\star(X_1\wedge X_2\wedge X_3),\Gamma_1\wedge\Gamma_2\rangle = \langle X_1\wedge X_2\wedge X_3, C(\Gamma_1)\wedge\Gamma_2\rangle - \langle X_1\wedge X_2\wedge X_3,\Gamma_1\wedge C(\Gamma_2)\rangle\,. \tag{H.4}$$

Defining $\langle A\otimes B, C\otimes D\rangle = \langle A,C\rangle\langle B,D\rangle$ we obtain

$$\langle A\wedge B, C\wedge D\rangle = 2\langle A,C\rangle\langle B,D\rangle - 2\langle A,D\rangle\langle B,C\rangle\,. \tag{H.5}$$

[68] We acknowledge illuminating discussions with Hadleigh Frost.

To use this, we need to split the three-fold wedge product democratically into two factors:

$$X_1 \wedge X_2 \wedge X_3 = \frac{1}{3}\Big((X_1 \wedge X_2) \wedge X_3 + X_1 \wedge (X_2 \wedge X_3) + (X_3 \wedge X_1) \wedge X_2\Big) \tag{H.6}$$

which exploits the cyclic symmetry of $X_1 \wedge X_2 \wedge X_3$ and the parenthesis indicates the split. Therefore (H.4) becomes

$$\begin{aligned}\langle C^\star(X_1 \wedge X_2 \wedge X_3), \Gamma_1 \wedge \Gamma_2\rangle &= \frac{1}{3}\langle (X_1 \wedge X_2) \wedge X_3, C(\Gamma_1) \wedge \Gamma_2\rangle - (\Gamma_1 \leftrightarrow \Gamma_2) + \mathrm{cyc}(X_1, X_2, X_3) \\ &= \frac{2}{3}\langle (X_1 \wedge X_2), C(\Gamma_1)\rangle\langle X_3, \Gamma_2\rangle - (\Gamma_1 \leftrightarrow \Gamma_2) + \mathrm{cyc}(X_1, X_2, X_3) \\ &= \frac{2}{3}\langle C^\star(X_1 \wedge X_2), \Gamma_1\rangle\langle X_3, \Gamma_2\rangle - (\Gamma_1 \leftrightarrow \Gamma_2) + \mathrm{cyc}(X_1, X_2, X_3) \\ &= \frac{1}{3}\langle C^\star(X_1 \wedge X_2) \wedge X_3, \Gamma_1 \wedge \Gamma_2\rangle + \mathrm{cyc}(X_1, X_2, X_3)\,,\end{aligned} \tag{H.7}$$

where we used (H.5) in the second line, $\langle X_1 \wedge X_2, \Gamma_2\rangle = 0$, and (H.5) again to identify the last line. Therefore we conclude

$$C^\star(X_1 \wedge X_2 \wedge X_3) = \frac{1}{3}C^\star(X_1 \wedge X_2) \wedge X_3 + \mathrm{cyc}(X_1, X_2, X_3) \tag{H.8}$$

which resembles the action of the (nilpotent) homology operator ∂ of [465] (see also [466]). Noting that $C^\star(X_1 \wedge X_2) = 2\{X_1, X_2\} \in \mathcal{L}^\star$, the right-hand side is in $\mathcal{L}^* \wedge \mathcal{L}^*$ and therefore $C^\star$ can act again,

$$C^\star \circ C^\star(X_1 \wedge X_2 \wedge X_3) = \frac{4}{3}\{\{X_1, X_2\}, X_3\} + \mathrm{cyc}(X_1, X_2, X_3) = 0 \tag{H.9}$$

by virtue of the Jacobi identity of the S bracket [157]. Therefore $C^\star \circ C^\star = 0$ and we conclude

Proposition 23. *The contact-term map is nilpotent*

$$C \circ C = 0\,. \tag{H.10}$$

Proof. Using that $C \circ C(\Gamma_1) \in \mathcal{L} \wedge \mathcal{L} \wedge \mathcal{L}$ we get

$$\langle X_1 \wedge X_2 \wedge X_3, C \circ C(\Gamma_1)\rangle = \langle C^\star(X_1 \wedge X_2 \wedge X_3), C(\Gamma_1)\rangle = \langle C^\star \circ C^\star(X_1 \wedge X_2 \wedge X_3), \Gamma_1\rangle = 0\,. \tag{H.11}$$

Therefore $C^2(\Gamma_1) = 0$ for any Lie polynomial Γ_1. By induction if $C^2(\Gamma_1) = C^2(\Gamma_2) = 0$ we also get $C^2(\Gamma_1 \wedge \Gamma_2) = 0$ as (H.2) implies

$$C \circ C(\Gamma_1 \wedge \Gamma_2) = C^2(\Gamma_1) \wedge \Gamma_2 + \Gamma_1 \wedge C^2(\Gamma_2)\,, \tag{H.12}$$

where we used that $C(\Gamma_1)$ has grading $+2$. □

Appendix I. BRST-invariant permutations at low multiplicities

To help understanding the definition of the Berends–Giele idempotent given in section 8.4.3, the first few permutations of (8.51) read as follows

$$\begin{aligned}\mathcal{E}(1) &= W_1\,, \\ \mathcal{E}(12) &= \frac{1}{2}(W_{12} - W_{21})\,, \\ \mathcal{E}(123) &= \frac{1}{3}W_{123} - \frac{1}{6}W_{132} - \frac{1}{6}W_{213} - \frac{1}{6}W_{231} - \frac{1}{6}W_{312} + \frac{1}{3}W_{321}\,, \\ \mathcal{E}(1234) &= \frac{1}{4}W_{1234} - \frac{1}{12}W_{1243} - \frac{1}{12}W_{1324} - \frac{1}{12}W_{1342} - \frac{1}{12}W_{1423} + \frac{1}{12}W_{1432}\end{aligned} \tag{I.1}$$

$$
\begin{aligned}
&- \frac{1}{12}W_{2134} + \frac{1}{12}W_{2143} - \frac{1}{12}W_{2314} - \frac{1}{12}W_{2341} + \frac{1}{12}W_{2413} + \frac{1}{12}W_{2431} \\
&- \frac{1}{12}W_{3124} - \frac{1}{12}W_{3142} + \frac{1}{12}W_{3214} + \frac{1}{12}W_{3241} - \frac{1}{12}W_{3412} + \frac{1}{12}W_{3421} \\
&- \frac{1}{12}W_{4123} + \frac{1}{12}W_{4132} + \frac{1}{12}W_{4213} + \frac{1}{12}W_{4231} + \frac{1}{12}W_{4312} - \frac{1}{4}W_{4321} \,,
\end{aligned}
$$

where a permutation σ is written as W_σ in order to avoid confusion with the rational coefficients. Using these definitions and examples, it is easy to generate the first few permutations of (8.50). For instance, at multiplicities three and four we have

$$
\begin{aligned}
\gamma_{1|2,3} &= W_{123} + W_{132} \,, \quad \gamma_{1|23} = \frac{1}{2}W_{123} - \frac{1}{2}W_{132} \,, \\
\gamma_{1|2,3,4} &= W_{1234} + W_{1243} + W_{1324} + W_{1342} + W_{1423} + W_{1432} \,, \\
\gamma_{1|23,4} &= \frac{1}{2}W_{1234} + \frac{1}{2}W_{1243} - \frac{1}{2}W_{1324} - \frac{1}{2}W_{1342} + \frac{1}{2}W_{1423} - \frac{1}{2}W_{1432} \,, \\
\gamma_{1|234} &= \frac{1}{3}W_{1234} - \frac{1}{6}W_{1243} - \frac{1}{6}W_{1324} - \frac{1}{6}W_{1342} - \frac{1}{6}W_{1423} + \frac{1}{3}W_{1432} \,,
\end{aligned}
\tag{I.2}
$$

where it suffices to list only the different partitions of labels as other permutations follow from relabeling due to the total symmetry of (8.50) under exchanges of any pair of words $P_i \leftrightarrow P_j$ and the functional form of (8.51). Similarly, at multiplicity five the BRST invariant permutations are given by

$$
\begin{aligned}
\gamma_{1|2,3,4,5} &= W_{1(2\sqcup\!\sqcup 3\sqcup\!\sqcup 4\sqcup\!\sqcup 5)} \,, \\
\gamma_{1|23,4,5} &= \frac{1}{2}W_{12345} + \frac{1}{2}W_{12354} + \frac{1}{2}W_{12435} + \frac{1}{2}W_{12453} + \frac{1}{2}W_{12534} + \frac{1}{2}W_{12543} \\
&- \frac{1}{2}W_{13245} - \frac{1}{2}W_{13254} - \frac{1}{2}W_{13425} - \frac{1}{2}W_{13452} - \frac{1}{2}W_{13524} - \frac{1}{2}W_{13542} \\
&+ \frac{1}{2}W_{14235} + \frac{1}{2}W_{14253} - \frac{1}{2}W_{14325} - \frac{1}{2}W_{14352} + \frac{1}{2}W_{14523} - \frac{1}{2}W_{14532} \\
&+ \frac{1}{2}W_{15234} + \frac{1}{2}W_{15243} - \frac{1}{2}W_{15324} - \frac{1}{2}W_{15342} + \frac{1}{2}W_{15423} - \frac{1}{2}W_{15432} \,, \\
\gamma_{1|234,5} &= \frac{1}{3}W_{12345} + \frac{1}{3}W_{12354} - \frac{1}{6}W_{12435} - \frac{1}{6}W_{12453} + \frac{1}{3}W_{12534} - \frac{1}{6}W_{12543} \\
&- \frac{1}{6}W_{13245} - \frac{1}{6}W_{13254} - \frac{1}{6}W_{13425} - \frac{1}{6}W_{13452} - \frac{1}{6}W_{13524} - \frac{1}{6}W_{13542} \\
&- \frac{1}{6}W_{14235} - \frac{1}{6}W_{14253} + \frac{1}{3}W_{14325} + \frac{1}{3}W_{14352} - \frac{1}{6}W_{14523} + \frac{1}{3}W_{14532} \\
&+ \frac{1}{3}W_{15234} - \frac{1}{6}W_{15243} - \frac{1}{6}W_{15324} - \frac{1}{6}W_{15342} - \frac{1}{6}W_{15423} + \frac{1}{3}W_{15432} \,, \\
\gamma_{1|23,45} &= \frac{1}{4}W_{12345} - \frac{1}{4}W_{12354} + \frac{1}{4}W_{12435} + \frac{1}{4}W_{12453} - \frac{1}{4}W_{12534} - \frac{1}{4}W_{12543} \\
&- \frac{1}{4}W_{13245} + \frac{1}{4}W_{13254} - \frac{1}{4}W_{13425} - \frac{1}{4}W_{13452} + \frac{1}{4}W_{13524} + \frac{1}{4}W_{13542} \\
&+ \frac{1}{4}W_{14235} + \frac{1}{4}W_{14253} - \frac{1}{4}W_{14325} - \frac{1}{4}W_{14352} + \frac{1}{4}W_{14523} - \frac{1}{4}W_{14532} \\
&- \frac{1}{4}W_{15234} - \frac{1}{4}W_{15243} + \frac{1}{4}W_{15324} + \frac{1}{4}W_{15342} - \frac{1}{4}W_{15423} + \frac{1}{4}W_{15432} \,, \\
\gamma_{1|2345} &= \frac{1}{4}W_{12345} - \frac{1}{12}W_{12354} - \frac{1}{12}W_{12435} - \frac{1}{12}W_{12453} - \frac{1}{12}W_{12534} + \frac{1}{12}W_{12543} \\
&- \frac{1}{12}W_{13245} + \frac{1}{12}W_{13254} - \frac{1}{12}W_{13425} - \frac{1}{12}W_{13452} + \frac{1}{12}W_{13524} + \frac{1}{12}W_{13542} \\
&- \frac{1}{12}W_{14235} - \frac{1}{12}W_{14253} + \frac{1}{12}W_{14325} + \frac{1}{12}W_{14352} - \frac{1}{12}W_{14523} + \frac{1}{12}W_{14532}
\end{aligned}
\tag{I.3}
$$

$$- \frac{1}{12}W_{15234} + \frac{1}{12}W_{15243} + \frac{1}{12}W_{15324} + \frac{1}{12}W_{15342} + \frac{1}{12}W_{15423} - \frac{1}{4}W_{15432} \,.$$

References

[1] N. Berkovits, Super Poincaré covariant quantization of the superstring, JHEP 04 (2000) 018. arXiv:hep-th/0001035, doi:10.1088/1126-6708/2000/04/018.

[2] N. Berkovits, Multiloop amplitudes and vanishing theorems using the pure spinor formalism for the superstring, JHEP 09 (2004) 047. arXiv:hep-th/0406055, doi:10.1088/1126-6708/2004/09/047.

[3] N. Berkovits, Pure spinor formalism as an N=2 topological string, JHEP 10 (2005) 089. arXiv:hep-th/0509120, doi:10.1088/1126-6708/2005/10/089.

[4] D. Friedan, E. J. Martinec, S. H. Shenker, Conformal Invariance, Supersymmetry and String Theory, Nucl. Phys. B 271 (1986) 93–165. doi:10.1016/0550-3213(86)90356-1.

[5] E. D'Hoker, D. H. Phong, The Geometry of String Perturbation Theory, Rev. Mod. Phys. 60 (1988) 917. doi:10.1103/RevModPhys.60.917.

[6] E. D'Hoker, D. H. Phong, Lectures on two loop superstrings, Conf. Proc. C 0208124 (2002) 85–123. arXiv:hep-th/0211111.

[7] E. Witten, Superstring Perturbation Theory Revisited (9 2012). arXiv:1209.5461.

[8] M. B. Green, J. H. Schwarz, Covariant Description of Superstrings, Phys. Lett. B 136 (1984) 367–370. doi:10.1016/0370-2693(84)92021-5.

[9] M. B. Green, J. H. Schwarz, Properties of the Covariant Formulation of Superstring Theories, Nucl. Phys. B 243 (1984) 285–306. doi:10.1016/0550-3213(84)90030-0.

[10] M. B. Green, J. H. Schwarz, E. Witten, Superstring Theory. Vol. 1: Introduction, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1988.

[11] M. B. Green, J. H. Schwarz, E. Witten, Superstring Theory. Vol. 2: Loop Amplitudes, Anomalies and Phenomenology, Cambridge University Press, 1988.

[12] J. Polchinski, String theory. Vol. 1: An introduction to the bosonic string, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2007. doi:10.1017/CBO9780511816079.

[13] J. Polchinski, String theory. Vol. 2: Superstring theory and beyond, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2007. doi:10.1017/CBO9780511618123.

[14] B. Zwiebach, A first course in string theory, Cambridge University Press, 2006.

[15] K. Becker, M. Becker, J. H. Schwarz, String theory and M-theory: A modern introduction, Cambridge University Press, 2006. doi:10.1017/CBO9780511816086.

[16] R. Blumenhagen, D. Lüst, S. Theisen, Basic concepts of string theory, Theoretical and Mathematical Physics, Springer, Heidelberg, Germany, 2013. doi:10.1007/978-3-642-29497-6.

[17] E. Kiritsis, String theory in a nutshell, Princeton University Press, USA, 2019.

[18] N. Berkovits, Relating the RNS and pure spinor formalisms for the superstring, JHEP 08 (2001) 026. arXiv:hep-th/0104247, doi:10.1088/1126-6708/2001/08/026.

[19] N. Berkovits, Untwisting the pure spinor formalism to the RNS and twistor string in a flat and $AdS_5 \times S^5$ background, JHEP 06 (2016) 127. arXiv:1604.04617, doi:10.1007/JHEP06(2016)127.

[20] N. Berkovits, Manifest spacetime supersymmetry and the superstring, JHEP 10 (2021) 162. arXiv:2106.04448, doi:10.1007/JHEP10(2021)162.

[21] C. R. Mafra, O. Schlotterer, S. Stieberger, Complete N-Point Superstring Disk Amplitude I. Pure Spinor Computation, Nucl. Phys. B873 (2013) 419–460. arXiv:1106.2645, doi:10.1016/j.nuclphysb.2013.04.023.

[22] W. Siegel, Classical Superstring Mechanics, Nucl. Phys. B 263 (1986) 93–104. doi:10.1016/0550-3213(86)90029-5.

[23] A. A. Rosly, K. G. Selivanov, On amplitudes in selfdual sector of Yang-Mills theory, Phys. Lett. B 399 (1997) 135–140. arXiv:hep-th/9611101, doi:10.1016/S0370-2693(97)00268-2.

[24] A. A. Rosly, K. G. Selivanov, Gravitational SD perturbiner (10 1997). arXiv:hep-th/9710196.

[25] K. G. Selivanov, On tree form-factors in (supersymmetric) Yang-Mills theory, Commun. Math. Phys. 208 (2000) 671–687. arXiv:hep-th/9809046, doi:10.1007/s002200050006.

[26] K. G. Selivanov, Post-classicism in Tree Amplitudes, in: 34th Rencontres de Moriond: Electroweak Interactions and Unified Theories, 1999. arXiv:hep-th/9905128.

[27] F. A. Berends, W. T. Giele, Recursive Calculations for Processes with n Gluons, Nucl. Phys. B 306 (1988) 759–808. doi:10.1016/0550-3213(88)90442-7.

[28] R. Kleiss, H. Kuijf, Multi-Gluon Cross-sections and Five Jet Production at Hadron Colliders, Nucl. Phys. B312 (1989) 616–644. doi:10.1016/0550-3213(89)90574-9.

[29] Z. Bern, J. J. M. Carrasco, H. Johansson, New Relations for Gauge-Theory Amplitudes, Phys. Rev. D 78 (2008) 085011. arXiv:0805.3993, doi:10.1103/PhysRevD.78.085011.

[30] Z. Bern, J. J. Carrasco, M. Chiodaroli, H. Johansson, R. Roiban, The Duality Between Color and Kinematics and its Applications (9 2019). arXiv:1909.01358.

[31] Z. Bern, J. J. Carrasco, M. Chiodaroli, H. Johansson, R. Roiban, The SAGEX Review on Scattering Amplitudes, Chapter 2: An Invitation to Color-Kinematics Duality and the Double Copy (3 2022). arXiv:2203.13013.

[32] T. Adamo, J. J. M. Carrasco, M. Carrillo-González, M. Chiodaroli, H. Elvang, H. Johansson, D. O'Connell, R. Roiban, O. Schlotterer, Snowmass White Paper: the Double Copy and its Applications, in: 2022 Snowmass Summer Study, 2022. arXiv:2204.06547.

[33] F. Cachazo, S. He, E. Y. Yuan, Scattering equations and Kawai-Lewellen-Tye orthogonality, Phys. Rev. D 90 (6) (2014) 065001. arXiv:1306.6575, doi:10.1103/PhysRevD.90.065001.

[34] F. Cachazo, S. He, E. Y. Yuan, Scattering of Massless Particles in Arbitrary Dimensions, Phys. Rev. Lett. 113 (17) (2014) 171601. arXiv:1307.2199, doi:10.1103/PhysRevLett.113.171601.

[35] F. Cachazo, S. He, E. Y. Yuan, Scattering of Massless Particles: Scalars, Gluons and Gravitons, JHEP 07 (2014) 033. arXiv:1309.0885, doi:10.1007/JHEP07(2014)033.

[36] E. Witten, Perturbative gauge theory as a string theory in twistor space, Commun. Math. Phys. 252 (2004) 189–258. arXiv:hep-th/0312171, doi:10.1007/s00220-004-1187-3.

[37] R. Roiban, M. Spradlin, A. Volovich, On the tree level S matrix of Yang-Mills theory, Phys. Rev. D 70 (2004) 026009. arXiv:hep-th/0403190, doi:10.1103/PhysRevD.70.026009.

[38] F. Cachazo, D. Skinner, Gravity from Rational Curves in Twistor Space, Phys. Rev. Lett. 110 (16) (2013) 161301. arXiv:1207.0741, doi:10.1103/PhysRevLett.110.161301.

[39] L. Mason, D. Skinner, Ambitwistor strings and the scattering equations, JHEP 07 (2014) 048. arXiv:1311.2564, doi:10.1007/JHEP07(2014)048.

[40] T. Adamo, E. Casali, D. Skinner, Ambitwistor strings and the scattering equations at one loop, JHEP 04 (2014) 104. arXiv:1312.3828, doi:10.1007/JHEP04(2014)104.

[41] N. Berkovits, Infinite Tension Limit of the Pure Spinor Superstring, JHEP 03 (2014) 017. arXiv:1311.4156, doi:10.1007/JHEP03(2014)017.

[42] T. Adamo, E. Casali, Scattering equations, supergravity integrands, and pure spinors, JHEP 05 (2015) 120. arXiv:1502.06826, doi:10.1007/JHEP05(2015)120.

[43] Y. Geyer, L. Mason, The SAGEX Review on Scattering Amplitudes, Chapter 6: Ambitwistor Strings and Amplitudes from the Worldsheet (3 2022). arXiv:2203.13017.

[44] H. Gomez, E. Y. Yuan, N-point tree-level scattering amplitude in the new Berkovits' string, JHEP 04 (2014) 046. arXiv:1312.5485, doi:10.1007/JHEP04(2014)046.

[45] T. Erler, Four Lectures on Closed String Field Theory, Phys. Rept. 851 (2020) 1–36. arXiv:1905.06785, doi:10.1016/j.physrep.2020.01.003.

[46] T. Erler, Four lectures on analytic solutions in open string field theory, Phys. Rept. 980 (2022) 1–95. arXiv:1912.00521, doi:10.1016/j.physrep.2022.06.004.

[47] H. Erbin, String Field Theory: A Modern Introduction, Vol. 980 of Lecture Notes in Physics, 2021. doi:10.1007/978-3-030-65321-7.

[48] A. Sen, Tachyon condensation on the brane anti-brane system, JHEP 08 (1998) 012. arXiv:hep-th/9805170, doi:10.1088/1126-6708/1998/08/012.

[49] A. Sen, B. Zwiebach, Tachyon condensation in string field theory, JHEP 03 (2000) 002. arXiv:hep-th/9912249, doi:10.1088/1126-6708/2000/03/002.

[50] N. Berkovits, A. Sen, B. Zwiebach, Tachyon condensation in superstring field theory, Nucl. Phys. B 587 (2000) 147–178. arXiv:hep-th/0002211, doi:10.1016/S0550-3213(00)00501-0.

[51] R. Pius, A. Rudra, A. Sen, Mass Renormalization in String Theory: Special States, JHEP 07 (2014) 058. arXiv:1311.1257, doi:10.1007/JHEP07(2014)058.

[52] R. Pius, A. Rudra, A. Sen, Mass Renormalization in String Theory: General States, JHEP 07 (2014) 062. arXiv:1401.7014, doi:10.1007/JHEP07(2014)062.

[53] A. Sen, One Loop Mass Renormalization of Unstable Particles in Superstring Theory, JHEP 11 (2016) 050. arXiv:1607.06500, doi:10.1007/JHEP11(2016)050.

[54] R. Gopakumar, C. Vafa, On the gauge theory / geometry correspondence, Adv. Theor. Math. Phys. 3 (1999) 1415–1443. arXiv:hep-th/9811131, doi:10.4310/ATMP.1999.v3.n5.a5.

[55] D. Gaiotto, L. Rastelli, A Paradigm of open / closed duality: Liouville D-branes and the Kontsevich model, JHEP 07 (2005) 053. arXiv:hep-th/0312196, doi:10.1088/1126-6708/2005/07/053.

[56] N. Berkovits, Perturbative Super-Yang-Mills from the Topological $AdS_5 \times S^5$ Sigma Model, JHEP 09 (2008) 088. arXiv:0806.1960, doi:10.1088/1126-6708/2008/09/088.

[57] Y. Okawa, Nonperturbative definition of closed string theory via open string field theory (6 2020). arXiv:2006.16449.

[58] N. Berkovits, Covariant quantization of the Green-Schwarz superstring in a Calabi-Yau background, Nucl. Phys. B 431 (1994) 258–272. arXiv:hep-th/9404162, doi:10.1016/0550-3213(94)90106-6.

[59] N. Berkovits, C. Vafa, N=4 topological strings, Nucl. Phys. B 433 (1995) 123–180. arXiv:hep-th/9407190, doi:10.1016/0550-3213(94)00419-F.

[60] N. Berkovits, A New description of the superstring, in: 8th Jorge Andre Swieca Summer School: Particles and Fields, 1996, pp. 390–418. arXiv:hep-th/9604123.

[61] N. Berkovits, C. Vafa, E. Witten, Conformal field theory of AdS background with Ramond-Ramond flux, JHEP 03 (1999) 018. arXiv:hep-th/9902098, doi:10.1088/1126-6708/1999/03/018.

[62] N. Berkovits, Quantization of the superstring with manifest U(5) superPoincaré invariance, Phys. Lett. B 457 (1999) 94–100. arXiv:hep-th/9902099, doi:10.1016/S0370-2693(99)00548-1.

[63] N. Berkovits, B. C. Vallilo, One loop N point superstring amplitudes with manifest d = 4 supersymmetry, Nucl. Phys. B 624 (2002) 45–62. arXiv:hep-th/0110168, doi:10.1016/S0550-3213(01)00653-8.

[64] N. Berkovits, M. Bershadsky, T. Hauer, S. Zhukov, B. Zwiebach, Superstring theory on $AdS_2 \times S^2$ as a coset supermanifold, Nucl. Phys. B 567 (2000) 61–86. arXiv:hep-th/9907200, doi:10.1016/S0550-3213(99)00683-5.

[65] L. Dolan, E. Witten, Vertex operators for AdS_3 background with Ramond-Ramond flux, JHEP 11 (1999) 003. arXiv:hep-th/9910205, doi:10.1088/1126-6708/1999/11/003.

[66] M. R. Gaberdiel, S. Gerigk, The massless string spectrum on $AdS_3 \times S^3$ from the supergroup, JHEP 10 (2011) 045. arXiv:1107.2660, doi:10.1007/JHEP10(2011)045.

[67] S. Gerigk, String States on $AdS_3 \times S^3$ from the Supergroup, JHEP 10 (2012) 084. arXiv:1208.0345, doi:10.1007/JHEP10(2012)084.

[68] A. Dei, M. R. Gaberdiel, R. Gopakumar, B. Knighton, Free field world-sheet correlators for AdS_3, JHEP 02 (2021) 081. arXiv:2009.11306, doi:10.1007/JHEP02(2021)081.

[69] M. R. Gaberdiel, K. Naderi, The physical states of the Hybrid Formalism, JHEP 10 (2021) 168. arXiv:2106.06476, doi:10.1007/JHEP10(2021)168.

[70] K. Bobkov, L. Dolan, Three graviton amplitude in Berkovits-Vafa-Witten variables, Phys. Lett. B 537 (2002) 155–164. arXiv:hep-th/0201027, doi:10.1016/S0370-2693(02)01877-4.

[71] L. Eberhardt, M. R. Gaberdiel, R. Gopakumar, The Worldsheet Dual of the Symmetric Product CFT, JHEP 04 (2019) 103. arXiv:1812.01007, doi:10.1007/JHEP04(2019)103.

[72] L. Eberhardt, M. R. Gaberdiel, R. Gopakumar, Deriving the AdS_3/CFT_2 correspondence, JHEP 02 (2020) 136. arXiv:1911.00378, doi:10.1007/JHEP02(2020)136.

[73] B. Knighton, Higher genus correlators for tensionless AdS_3 strings, JHEP 04 (2021) 211. arXiv:2012.01445, doi:10.1007/JHEP04(2021)211.

[74] N. Berkovits, Quantum consistency of the superstring in $AdS_5 \times S^5$ background, JHEP 03 (2005) 041. arXiv:hep-th/0411170, doi:10.1088/1126-6708/2005/03/041.

[75] R. R. Metsaev, A. A. Tseytlin, Type IIB superstring action in $AdS_5 \times S^5$ background, Nucl. Phys. B 533 (1998) 109–126. arXiv:hep-th/9805028, doi:10.1016/S0550-3213(98)00570-7.

[76] L. Mazzucato, Superstrings in AdS, Phys. Rept. 521 (2012) 1–68. arXiv:1104.2604, doi:10.1016/j.physrep.2012.08.001.

[77] N. Berkovits, E. D'Hoker, M. B. Green, H. Johansson, O. Schlotterer, Snowmass White Paper: String Perturbation Theory, in: 2022 Snowmass Summer Study, 2022. arXiv:2203.09099.

[78] R. Gopakumar, E. Perlmutter, S. S. Pufu, X. Yin, Snowmass White Paper: Bootstrapping String Theory (2 2022). arXiv:2202.07163.

[79] L. Brink, J. H. Schwarz, J. Scherk, Supersymmetric Yang-Mills Theories, Nucl. Phys. B 121 (1977) 77–92. doi:10.1016/0550-3213(77)90328-5.

[80] M. B. Green, J. H. Schwarz, L. Brink, N=4 Yang-Mills and N=8 Supergravity as Limits of String Theories, Nucl. Phys. B 198 (1982) 474–492. doi:10.1016/0550-3213(82)90336-4.

[81] W. Siegel, Superfields in Higher Dimensional Space-time, Phys. Lett. B 80 (1979) 220–223. doi:10.1016/0370-2693(79)90202-8.

[82] E. Witten, Twistor - Like Transform in Ten-Dimensions, Nucl. Phys. B 266 (1986) 245–264. doi:10.1016/0550-3213(86)90090-8.

[83] C. R. Mafra, O. Schlotterer, Multiparticle SYM equations of motion and pure spinor BRST blocks, JHEP 07 (2014) 153. arXiv:1404.4986, doi:10.1007/JHEP07(2014)153.

[84] S. Lee, C. R. Mafra, O. Schlotterer, Non-linear gauge transformations in $D = 10$ SYM theory and the BCJ duality, JHEP 03 (2016) 090. arXiv:1510.08843, doi:10.1007/JHEP03(2016)090.

[85] E. Bridges, C. R. Mafra, Algorithmic construction of SYM multiparticle superfields in the BCJ gauge, JHEP 10 (2019) 022. arXiv:1906.12252, doi:10.1007/JHEP10(2019)022.

[86] H. Gomez, C. R. Mafra, The closed-string 3-loop amplitude and S-duality, JHEP 10 (2013) 217. arXiv:1308.6567, doi:10.1007/JHEP10(2013)217.

[87] C. R. Mafra, O. Schlotterer, Solution to the nonlinear field equations of ten dimensional supersymmetric Yang-Mills theory, Phys. Rev. D 92 (6) (2015) 066001. arXiv:1501.05562, doi:10.1103/PhysRevD.92.066001.

[88] J. P. Harnad, S. Shnider, Constraints and field equations for ten-dimensional Super-Yang-Mills Theory, Commun. Math. Phys. 106 (1986) 183. doi:10.1007/BF01454971.

[89] G. Policastro, D. Tsimpis, R^4, purified, Class. Quant. Grav. 23 (2006) 4753–4780. arXiv:hep-th/0603165, doi:10.1088/0264-9381/23/14/012.

[90] H. Ooguri, J. Rahmfeld, H. Robins, J. Tannenhauser, Holography in superspace, JHEP 07 (2000) 045. arXiv:hep-th/0007104, doi:10.1088/1126-6708/2000/07/045.

[91] N. Berkovits, Super-Poincaré covariant two-loop superstring amplitudes, JHEP 01 (2006) 005. arXiv:hep-th/0503197, doi:10.1088/1126-6708/2006/01/005.

[92] N. Berkovits, ICTP lectures on covariant quantization of the superstring, ICTP Lect. Notes Ser. 13 (2003) 57–107. arXiv:hep-th/0209059.

[93] C. R. Mafra, Superstring Scattering Amplitudes with the Pure Spinor Formalism, Ph.D. thesis, São Paulo, IFT (2008). arXiv:0902.1552.

[94] O. Schlotterer, Scattering amplitudes in open superstring theory, Fortsch. Phys. 60 (2012) 373–691. doi:10.1002/prop.201100084.

[95] M. B. Green, J. H. Schwarz, Supersymmetrical Dual String Theory, Nucl. Phys. B 181 (1981) 502–530. doi:10.1016/0550-3213(81)90538-1.

[96] M. B. Green, J. H. Schwarz, Supersymmetrical String Theories, Phys. Lett. B 109 (1982) 444–448. doi:10.1016/0370-2693(82)91110-8.

[97] D. J. Gross, J. A. Harvey, E. J. Martinec, R. Rohm, The Heterotic String, Phys. Rev. Lett. 54 (1985) 502–505. doi:10.1103/PhysRevLett.54.502.

[98] P. Dirac, Lectures on Quantum Mechanics, Belfer Graduate School of Science, monograph series, Dover Publications,

2001.
URL https://books.google.co.uk/books?id=GVwzb1rZW9kC

[99] M. B. Green, J. H. Schwarz, Supersymmetrical Dual String Theory. 2. Vertices and Trees, Nucl. Phys. B 198 (1982) 252–268. doi:10.1016/0550-3213(82)90556-9.

[100] M. B. Green, J. H. Schwarz, Supersymmetrical Dual String Theory. 3. Loops and Renormalization, Nucl. Phys. B 198 (1982) 441–460. doi:10.1016/0550-3213(82)90334-0.

[101] P. Francesco, P. Mathieu, D. Sénéchal, Conformal Field Theory, Springer New York, NY, 1997. doi:10.1007/978-1-4612-2256-9.

[102] F. Essler, M. Hatsuda, E. Laenen, W. Siegel, J. P. Yamron, T. Kimura, A. R. Mikovic, Covariant quantization of the first ilk superparticle, Nucl. Phys. B 364 (1991) 67–84. doi:10.1016/0550-3213(91)90578-L.

[103] F. Essler, E. Laenen, W. Siegel, J. P. Yamron, BRST operator for the first ilk superparticle, Phys. Lett. B 254 (1991) 411–416. doi:10.1016/0370-2693(91)91176-V.

[104] E. Cartan, The Theory of Spinors, Hermann, 1966.

[105] P. Furlan, R. Raczka, Nonlinear Spinor Representations, J. Math. Phys. 26 (1985) 3021. doi:10.1063/1.526678.

[106] N. Berkovits, S. A. Cherkis, Higher-dimensional twistor transforms using pure spinors, JHEP 12 (2004) 049. arXiv:hep-th/0409243, doi:10.1088/1126-6708/2004/12/049.

[107] C. R. Mafra, Introdução aos Formalismos de Green-Schwarz e Espinores Puros da Supercorda, Master's thesis, São Paulo, IFT (2005).

[108] J. Hoogeveen, Fundamentals of the pure spinor formalism, Ph.D. thesis, Amsterdam U., Inst. Math. (2010).

[109] C. R. Mafra, Simplifying the Tree-level Superstring Massless Five-point Amplitude, JHEP 01 (2010) 007. arXiv:0909.5206, doi:10.1007/JHEP01(2010)007.

[110] H. Gomez, One-loop Superstring Amplitude From Integrals on Pure Spinors Space, JHEP 12 (2009) 034. arXiv:0910.3405, doi:10.1088/1126-6708/2009/12/034.

[111] H. Gomez, C. R. Mafra, The Overall Coefficient of the Two-loop Superstring Amplitude Using Pure Spinors, JHEP 05 (2010) 017. arXiv:1003.0678, doi:10.1007/JHEP05(2010)017.

[112] N. Berkovits, Twistor Origin of the Superstring, JHEP 03 (2015) 122. arXiv:1409.2510, doi:10.1007/JHEP03(2015)122.

[113] M. Matone, L. Mazzucato, I. Oda, D. Sorokin, M. Tonin, The Superembedding origin of the Berkovits pure spinor covariant quantization of superstrings, Nucl. Phys. B 639 (2002) 182–202. arXiv:hep-th/0206104, doi:10.1016/S0550-3213(02)00562-X.

[114] Y. Aisaka, Y. Kazama, Origin of pure spinor superstring, JHEP 05 (2005) 046. arXiv:hep-th/0502208, doi:10.1088/1126-6708/2005/05/046.

[115] N. Berkovits, Explaining the Pure Spinor Formalism for the Superstring, JHEP 01 (2008) 065. arXiv:0712.0324, doi:10.1088/1126-6708/2008/01/065.

[116] N. Berkovits, Pure spinors, twistors, and emergent supersymmetry, JHEP 12 (2012) 006. arXiv:1105.1147, doi:10.1007/JHEP12(2012)006.

[117] P. S. Howe, Pure spinors lines in superspace and ten-dimensional supersymmetric theories, Phys. Lett. B 258 (1991) 141–144, [Addendum: Phys.Lett.B 259, 511 (1991)]. doi:10.1016/0370-2693(91)91221-G.

[118] B. E. Nilsson, Pure spinors as auxiliary fields in the ten-dimensional supersymmetric yang-mills theory, Classical and Quantum Gravity 3 (2) (1986) L41.

[119] M. Cederwall, Pure spinor superfields – an overview, Springer Proc. Phys. 153 (2014) 61–93. arXiv:1307.1762, doi:10.1007/978-3-319-03774-5_4.

[120] M. Cederwall, Pure spinors in classical and quantum supergravity (10 2022). arXiv:2210.06141.

[121] L. P. Hughston, W. T. Shaw, Classical strings in ten dimensions, Proc. Roy. Soc. Lond. A 414 (1987) 423. doi:10.1098/rspa.1987.0152.

[122] N. Berkovits, M. T. Hatsuda, W. Siegel, The Big picture, Nucl. Phys. B 371 (1992) 434–466. arXiv:hep-th/9108021, doi:10.1016/0550-3213(92)90242-4.

[123] I. Oda, M. Tonin, Y-formalism and b ghost in the non-minimal pure spinor formalism of superstrings, Nucl. Phys. B 779 (2007) 63–100. arXiv:0704.1219, doi:10.1016/j.nuclphysb.2007.04.032.

[124] R. Lipinski Jusinskas, Nilpotency of the b ghost in the non-minimal pure spinor formalism, JHEP 05 (2013) 048. arXiv:1303.3966, doi:10.1007/JHEP05(2013)048.

[125] O. Chandia, The b Ghost of the Pure Spinor Formalism is Nilpotent, Phys. Lett. B 695 (2011) 312–316. arXiv:1008.1778, doi:10.1016/j.physletb.2010.10.058.

[126] R. L. Jusinskas, Notes on the pure spinor b ghost, JHEP 07 (2013) 142. arXiv:1306.1963, doi:10.1007/JHEP07(2013)142.

[127] O. Chandia, B. C. Vallilo, Relating the b ghost and the vertex operators of the pure spinor superstring, JHEP 03 (2021) 165. arXiv:2101.01129, doi:10.1007/JHEP03(2021)165.

[128] N. Berkovits, Cohomology in the pure spinor formalism for the superstring, JHEP 09 (2000) 046. arXiv:hep-th/0006003, doi:10.1088/1126-6708/2000/09/046.

[129] N. Berkovits, O. Chandia, Lorentz invariance of the pure spinor BRST cohomology for the superstring, Phys. Lett. B 514 (2001) 394–400. arXiv:hep-th/0105149, doi:10.1016/S0370-2693(01)00789-4.

[130] Y. Aisaka, E. A. Arroyo, N. Berkovits, N. Nekrasov, Pure Spinor Partition Function and the Massive Superstring Spectrum, JHEP 08 (2008) 050. arXiv:0806.0584, doi:10.1088/1126-6708/2008/08/050.

[131] N. Berkovits, O. Chandia, Massive superstring vertex operator in D = 10 superspace, JHEP 08 (2002) 040. arXiv:hep-th/0204121, doi:10.1088/1126-6708/2002/08/040.

[132] S. Chakrabarti, S. P. Kashyap, M. Verma, Integrated Massive Vertex Operator in Pure Spinor Formalism, JHEP 10 (2018) 147. arXiv:1802.04486, doi:10.1007/JHEP10(2018)147.

[133] S. Chakrabarti, S. P. Kashyap, M. Verma, Amplitudes Involving Massive States Using Pure Spinor Formalism, JHEP 12 (2018) 071. arXiv:1808.08735, doi:10.1007/JHEP12(2018)071.

[134] N. E. J. Bjerrum-Bohr, P. H. Damgaard, P. Vanhove, Minimal Basis for Gauge Theory Amplitudes, Phys. Rev. Lett. 103 (2009) 161602. arXiv:0907.1425, doi:10.1103/PhysRevLett.103.161602.

[135] S. Stieberger, Open & Closed vs. Pure Open String Disk Amplitudes (2009). arXiv:0907.2211.

[136] J. Broedel, L. J. Dixon, Color-kinematics duality and double-copy construction for amplitudes from higher-dimension operators, JHEP 10 (2012) 091. arXiv:1208.0876, doi:10.1007/JHEP10(2012)091.

[137] S. Stieberger, T. R. Taylor, Closed String Amplitudes as Single-Valued Open String Amplitudes, Nucl. Phys. B881 (2014) 269–287. arXiv:1401.1218, doi:10.1016/j.nuclphysb.2014.02.005.

[138] C. R. Mafra, KK-like relations of α' corrections to disk amplitudes, JHEP 03 (2022) 012. arXiv:2108.01081, doi:10.1007/JHEP03(2022)012.

[139] L. Solomon, A mackey formula in the group ring of a coxeter group, Journal of Algebra 41 (2) (1976) 255–264.

[140] A. M. Garsia, C. Reutenauer, A decomposition of solomon's descent algebra, Advances in mathematics 77 (2) (1989) 189–262.

[141] M. Schocker, The descent algebra of the symmetric group, 2004, pp. 145–161. doi:10.1090/fic/040/07.

[142] J.-Y. Thibon, Lie idempotents in descent algebras (2016).
URL https://igm.univ-mlv.fr/~jyt/TALKS/lieids.ps

[143] C. Reutenauer, Free Lie algebras, in: Handbook of algebra, Vol. 3, Handb. Algebr., Elsevier/North-Holland, Amsterdam, 2003, pp. 887–903. doi:10.1016/S1570-7954(03)80075-X.
URL https://doi.org/10.1016/S1570-7954(03)80075-X

[144] T. Azevedo, On the $\mathcal{N}$ = 4, d = 4 pure spinor measure factor, JHEP 03 (2015) 136. arXiv:1412.5927, doi:10.1007/JHEP03(2015)136.

[145] N. Berkovits, B. C. Vallilo, Consistency of superPoincaré covariant superstring tree amplitudes, JHEP 07 (2000) 015. arXiv:hep-th/0004171, doi:10.1088/1126-6708/2000/07/015.

[146] A. Leeuwen, M. Cohen, B. Lisser, LiE, A Package for Lie Group Computations, Computer Algebra Nederland, Amsterdam, ISBN 90-74116-02-7 (1992).
URL http://www-math.univ-poitiers.fr/~maavl/LiE/

[147] J. Scherk, Zero-slope limit of the dual resonance model, Nucl. Phys. B 31 (1971) 222–234. doi:10.1016/0550-3213(71)90227-6.

[148] A. Neveu, J. Scherk, Connection between Yang-Mills fields and dual models, Nucl. Phys. B 36 (1972) 155–161. doi:10.1016/0550-3213(72)90301-X.

[149] F. Gliozzi, J. Scherk, D. I. Olive, Supergravity and the Spinor Dual Model, Phys. Lett. B 65 (1976) 282–286. doi:10.1016/0370-2693(76)90183-0.

[150] F. Gliozzi, J. Scherk, D. I. Olive, Supersymmetry, Supergravity Theories and the Dual Spinor Model, Nucl. Phys. B 122 (1977) 253–290. doi:10.1016/0550-3213(77)90206-1.

[151] N. Berkovits, Explaining Pure Spinor Superspace (12 2006). arXiv:hep-th/0612021.

[152] J. A. M. Vermaseren, New features of FORM (2000). arXiv:math-ph/0010025.

[153] C. R. Mafra, PSS: A FORM Program to Evaluate Pure Spinor Superspace Expressions (7 2010). arXiv:1007.4999.

[154] K.-S. Sun, X.-M. Ding, F. Sun, H.-B. Zhang, Computations of superstring amplitudes in pure spinor formalism via Cadabra (7 2016). arXiv:1607.00463.

[155] W. Giele, G. Stavenga, J.-C. Winter, Thread-Scalable Evaluation of Multi-Jet Observables, Eur. Phys. J. C 71 (2011) 1703. arXiv:1002.3446, doi:10.1140/epjc/s10052-011-1703-5.

[156] C. R. Mafra, Planar binary trees in scattering amplitudes, 2020. arXiv:2011.14413, doi:10.4171/205-1/6.

[157] H. Frost, C. R. Mafra, L. Mason, A Lie bracket for the momentum kernel (12 2020). arXiv:2012.00519.

[158] C. R. Mafra, Pure Spinor Superspace Identities for Massless Four-point Kinematic Factors, JHEP 04 (2008) 093. arXiv:0801.0580, doi:10.1088/1126-6708/2008/04/093.

[159] C. R. Mafra, Towards Field Theory Amplitudes From the Cohomology of Pure Spinor Superspace, JHEP 11 (2010) 096. arXiv:1007.3639, doi:10.1007/JHEP11(2010)096.

[160] C. R. Mafra, O. Schlotterer, S. Stieberger, D. Tsimpis, A recursive method for SYM n-point tree amplitudes, Phys. Rev. D83 (2011) 126012. arXiv:1012.3981, doi:10.1103/PhysRevD.83.126012.

[161] C. R. Mafra, O. Schlotterer, S. Stieberger, D. Tsimpis, Six Open String Disk Amplitude in Pure Spinor Superspace, Nucl. Phys. B846 (2011) 359–393. arXiv:1011.0994, doi:10.1016/j.nuclphysb.2011.01.008.

[162] D. Blessenohl, H. Laue, Generalized jacobi identities, Note di Matematica 8 (1) (1988) 111–121.

[163] C. R. Mafra, O. Schlotterer, The Structure of n-Point One-Loop Open Superstring Amplitudes, JHEP 08 (2014) 099. arXiv:1203.6215, doi:10.1007/JHEP08(2014)099.

[164] N. Ahmadiniaz, F. M. Balli, C. Lopez-Arcos, A. Q. Velez, C. Schubert, Color-kinematics duality from the Bern-Kosower formalism, Phys. Rev. D 104 (4) (2021) L041702. arXiv:2105.06745, doi:10.1103/PhysRevD.104.L041702.

[165] C. R. Mafra, O. Schlotterer, Berends-Giele recursions and the BCJ duality in superspace and components, JHEP 03 (2016) 097. arXiv:1510.08846, doi:10.1007/JHEP03(2016)097.

[166] J. Broedel, O. Schlotterer, S. Stieberger, Polylogarithms, Multiple Zeta Values and Superstring Amplitudes, Fortsch. Phys. 61 (2013) 812–870. arXiv:1304.7267, doi:10.1002/prop.201300019.

[167] R. Ree, Lie elements and an algebra associated with shuffles, Annals of Mathematics (1958) 210–220.

[168] M. B. Green, C. R. Mafra, O. Schlotterer, Multiparticle one-loop amplitudes and S-duality in closed superstring theory, JHEP 10 (2013) 188. arXiv:1307.3534, doi:10.1007/JHEP10(2013)188.

[169] C. R. Mafra, O. Schlotterer, Towards one-loop SYM amplitudes from the pure spinor BRST cohomology, Fortsch. Phys.

63 (2) (2015) 105–131. arXiv:1410.0668, doi:10.1002/prop.201400076.
[170] C. R. Mafra, O. Schlotterer, Two-loop five-point amplitudes of super Yang-Mills and supergravity in pure spinor superspace, JHEP 10 (2015) 124. arXiv:1505.02746, doi:10.1007/JHEP10(2015)124.
[171] C. R. Mafra, O. Schlotterer, Towards the n-point one-loop superstring amplitude. Part III. One-loop correlators and their double-copy structure, JHEP 08 (2019) 092. arXiv:1812.10971, doi:10.1007/JHEP08(2019)092.
[172] E. D'Hoker, C. R. Mafra, B. Pioline, O. Schlotterer, Two-loop superstring five-point amplitudes. Part I. Construction via chiral splitting and pure spinors, JHEP 08 (2020) 135. arXiv:2006.05270, doi:10.1007/JHEP08(2020)135.
[173] J. Bjornsson, Multi-loop amplitudes in maximally supersymmetric pure spinor field theory, JHEP 01 (2011) 002. arXiv:1009.5906, doi:10.1007/JHEP01(2011)002.
[174] J. Bjornsson, M. B. Green, 5 loops in 24/5 dimensions, JHEP 08 (2010) 132. arXiv:1004.2692, doi:10.1007/JHEP08(2010)132.
[175] F. A. Berends, W. T. Giele, Multiple Soft Gluon Radiation in Parton Processes, Nucl. Phys. B 313 (1989) 595–633. doi:10.1016/0550-3213(89)90398-2.
[176] M. Schocker, Lie elements and knuth relations, Canadian Journal of Mathematics 56 (4) (2004) 871–882.
[177] C. Schubert, The Finite gauge transformations in closed string field theory, Lett. Math. Phys. 26 (1992) 259–264. doi:10.1007/BF00420234.
[178] N. B. Elliot, BCJ Identities in Pure Spinor Superspace, Ph.D. thesis, Southampton U. (2021).
[179] A. M. Garsia, Combinatorics of the free lie algebra and the symmetric group, in: Analysis, et cetera, Elsevier, 1990, pp. 309–382.
[180] H. Frost, Universal aspects of perturbative gauge theory, Ph.D. thesis, University of Oxford (2020).
[181] W. Michaelis, Lie coalgebras, Advances in mathematics 38 (1) (1980) 1–54.
[182] H. Kawai, D. C. Lewellen, S. H. H. Tye, A Relation Between Tree Amplitudes of Closed and Open Strings, Nucl. Phys. B 269 (1986) 1–23. doi:10.1016/0550-3213(86)90362-7.
[183] Z. Bern, L. J. Dixon, M. Perelstein, J. S. Rozowsky, Multileg one loop gravity amplitudes from gauge theory, Nucl. Phys. B 546 (1999) 423–479. arXiv:hep-th/9811140, doi:10.1016/S0550-3213(99)00029-2.
[184] N. E. J. Bjerrum-Bohr, P. H. Damgaard, B. Feng, T. Sondergaard, Gravity and Yang-Mills Amplitude Relations, Phys. Rev. D 82 (2010) 107702. arXiv:1005.4367, doi:10.1103/PhysRevD.82.107702.
[185] N. E. J. Bjerrum-Bohr, P. H. Damgaard, T. Sondergaard, P. Vanhove, The Momentum Kernel of Gauge and Gravity Theories, JHEP 01 (2011) 001. arXiv:1010.3933, doi:10.1007/JHEP01(2011)001.
[186] Y.-J. Du, C.-H. Fu, Explicit BCJ numerators of nonlinear simga model, JHEP 09 (2016) 174. arXiv:1606.05846, doi:10.1007/JHEP09(2016)174.
[187] J. J. M. Carrasco, C. R. Mafra, O. Schlotterer, Abelian Z-theory: NLSM amplitudes and α'-corrections from the open string, JHEP 06 (2017) 093. arXiv:1608.02569, doi:10.1007/JHEP06(2017)093.
[188] S. Mizera, Combinatorics and Topology of Kawai-Lewellen-Tye Relations, JHEP 08 (2017) 097. arXiv:1706.08527, doi:10.1007/JHEP08(2017)097.
[189] S. Mizera, Scattering Amplitudes from Intersection Theory, Phys. Rev. Lett. 120 (14) (2018) 141602. arXiv:1711.00469, doi:10.1103/PhysRevLett.120.141602.
[190] S. Mizera, Aspects of Scattering Amplitudes and Moduli Space Localization, Ph.D. thesis, Princeton, Inst. Advanced Study (2020). arXiv:1906.02099, doi:10.1007/978-3-030-53010-5.
[191] H. Frost, The Algebraic Structure of the KLT Relations for Gauge and Gravity Tree Amplitudes, SIGMA 17 (2021) 101. arXiv:2111.07257, doi:10.3842/SIGMA.2021.101.
[192] C. R. Mafra, O. Schlotterer, S. Stieberger, Complete N-Point Superstring Disk Amplitude II. Amplitude and Hypergeometric Function Structure, Nucl. Phys. B873 (2013) 461–513. arXiv:1106.2646, doi:10.1016/j.nuclphysb.2013.04.022.
[193] F. A. Berends, W. Giele, The Six Gluon Process as an Example of Weyl-Van Der Waerden Spinor Calculus, Nucl. Phys. B 294 (1987) 700–732. doi:10.1016/0550-3213(87)90604-3.
[194] M. L. Mangano, S. J. Parke, Z. Xu, Duality and Multi - Gluon Scattering, Nucl. Phys. B 298 (1988) 653–672. doi:10.1016/0550-3213(88)90001-6.
[195] V. Del Duca, L. J. Dixon, F. Maltoni, New color decompositions for gauge amplitudes at tree and loop level, Nucl. Phys. B 571 (2000) 51–70. arXiv:hep-ph/9910563, doi:10.1016/S0550-3213(99)00809-3.
[196] S. Badger, B. Biedermann, L. Hackl, J. Plefka, T. Schuster, P. Uwer, Comparing efficient computation methods for massless QCD tree amplitudes: Closed analytic formulas versus Berends-Giele recursion, Phys. Rev. D 87 (3) (2013) 034011. arXiv:1206.2381, doi:10.1103/PhysRevD.87.034011.
[197] C. Duhr, S. Hoeche, F. Maltoni, Color-dressed recursive relations for multi-parton amplitudes, JHEP 08 (2006) 062. arXiv:hep-ph/0607057, doi:10.1088/1126-6708/2006/08/062.
[198] R. Medina, F. T. Brandt, F. R. Machado, The Open superstring five point amplitude revisited, JHEP 07 (2002) 071. arXiv:hep-th/0208121, doi:10.1088/1126-6708/2002/07/071.
[199] D. Oprisa, S. Stieberger, Six gluon open superstring disk amplitude, multiple hypergeometric series and Euler-Zagier sums (2005). arXiv:hep-th/0509042.
[200] C. Mafra, O. Schlotterer, PSS: From pure spinor superspace to components.
URL http://www.southampton.ac.uk/~crm1n16/pss.html
[201] L. M. Garozzo, L. Queimada, O. Schlotterer, Berends-Giele currents in Bern-Carrasco-Johansson gauge for F^3- and F^4-deformed Yang-Mills amplitudes, JHEP 02 (2019) 078. arXiv:1809.08103, doi:10.1007/JHEP02(2019)078.
[202] F. A. Berends, W. T. Giele, H. Kuijf, Exact and Approximate Expressions for Multi - Gluon Scattering, Nucl. Phys. B 333 (1990) 120–159. doi:10.1016/0550-3213(90)90225-3.
[203] R. Bandiera, private communication (2020).

[204] J.-C. Novelli, J.-Y. Thibon, F. Toumazet, A noncommutative cycle index and new bases of quasi-symmetric functions and noncommutative symmetric functions, Annals of Combinatorics 24 (3) (2020) 557–576. arXiv:1804.01762.

[205] B. Feng, R. Huang, Y. Jia, Gauge Amplitude Identities by On-shell Recursion Relation in S-matrix Program, Phys. Lett. B 695 (2011) 350–353. arXiv:1004.3417, doi:10.1016/j.physletb.2010.11.011.

[206] Y.-X. Chen, Y.-J. Du, B. Feng, A Proof of the Explicit Minimal-basis Expansion of Tree Amplitudes in Gauge Field Theory, JHEP 02 (2011) 112. arXiv:1101.0009, doi:10.1007/JHEP02(2011)112.

[207] N. Berkovits, Covariant quantization of the superparticle using pure spinors, JHEP 09 (2001) 016. arXiv:hep-th/0105050, doi:10.1088/1126-6708/2001/09/016.

[208] L. S. Brown, Summing tree graphs at threshold, Phys. Rev. D 46 (1992) R4125–R4127. arXiv:hep-ph/9209203, doi:10.1103/PhysRevD.46.R4125.

[209] K. Aomoto, Gauss-Manin connection of integral of difference products, J. Math. Soc. Japan 39 (2) (1987) 191–208. doi:10.2969/jmsj/03920191.
URL https://doi.org/10.2969/jmsj/03920191

[210] Y.-t. Huang, O. Schlotterer, C. Wen, Universality in string interactions, JHEP 09 (2016) 155. arXiv:1602.01674, doi:10.1007/JHEP09(2016)155.

[211] T. Azevedo, M. Chiodaroli, H. Johansson, O. Schlotterer, Heterotic and bosonic string amplitudes via field theory, JHEP 10 (2018) 012. arXiv:1803.05452, doi:10.1007/JHEP10(2018)012.

[212] G. Veneziano, Construction of a crossing - symmetric, Regge behaved amplitude for linearly rising trajectories, Nuovo Cim. A 57 (1968) 190–197. doi:10.1007/BF02824451.

[213] Y. Kitazawa, Effective Lagrangian for Open Superstring From Five Point Function, Nucl. Phys. B 289 (1987) 599–608. doi:10.1016/0550-3213(87)90396-8.

[214] L. A. Barreiro, R. Medina, 5-field terms in the open superstring effective action, JHEP 03 (2005) 055. arXiv:hep-th/0503182, doi:10.1088/1126-6708/2005/03/055.

[215] S. Stieberger, T. R. Taylor, Amplitude for N-Gluon Superstring Scattering, Phys. Rev. Lett. 97 (2006) 211601. arXiv:hep-th/0607184, doi:10.1103/PhysRevLett.97.211601.

[216] S. Stieberger, T. R. Taylor, Multi-Gluon Scattering in Open Superstring Theory, Phys. Rev. D 74 (2006) 126007. arXiv:hep-th/0609175, doi:10.1103/PhysRevD.74.126007.

[217] S. Stieberger, T. R. Taylor, Supersymmetry Relations and MHV Amplitudes in Superstring Theory, Nucl. Phys. B 793 (2008) 83–113. arXiv:0708.0574, doi:10.1016/j.nuclphysb.2007.09.033.

[218] S. Stieberger, T. R. Taylor, Complete Six-Gluon Disk Amplitude in Superstring Theory, Nucl. Phys. B 801 (2008) 128–152. arXiv:0711.4354, doi:10.1016/j.nuclphysb.2008.04.026.

[219] K. Aomoto, On the complex Selberg integral, The Quarterly Journal of Mathematics 38 (4) (1987) 385–399. doi:10.1093/qmath/38.4.385.
URL https://dx.doi.org/10.1093/qmath/38.4.385

[220] S. Stieberger, T. R. Taylor, Maximally Helicity Violating Disk Amplitudes, Twistors and Transcendental Integrals, Phys. Lett. B 716 (2012) 236–239. arXiv:1204.3848, doi:10.1016/j.physletb.2012.08.018.

[221] S. J. Parke, T. R. Taylor, An Amplitude for n Gluon Scattering, Phys. Rev. Lett. 56 (1986) 2459. doi:10.1103/PhysRevLett.56.2459.

[222] P. Frampton, Dual Resonance Models, Advance book program, W. A. Benjamin, 1974.
URL https://books.google.co.uk/books?id=lp6yAAAAIAAJ

[223] F. Brown, C. Dupont, Single-valued integration and superstring amplitudes in genus zero, Commun. Math. Phys. 382 (2) (2021) 815–874. arXiv:1910.01107, doi:10.1007/s00220-021-03969-4.

[224] C. R. Mafra, Berends-Giele recursion for double-color-ordered amplitudes, JHEP 07 (2016) 080. arXiv:1603.09731, doi:10.1007/JHEP07(2016)080.

[225] C. R. Mafra, O. Schlotterer, Non-abelian Z-theory: Berends-Giele recursion for the α'-expansion of disk integrals, JHEP 01 (2017) 031. arXiv:1609.07078, doi:10.1007/JHEP01(2017)031.

[226] C. R. Mafra, O. Schlotterer, S. Stieberger, Explicit BCJ Numerators from Pure Spinors, JHEP 07 (2011) 092. arXiv:1104.5224, doi:10.1007/JHEP07(2011)092.

[227] Z. Bern, T. Dennen, Y.-t. Huang, M. Kiermaier, Gravity as the Square of Gauge Theory, Phys. Rev. D82 (2010) 065003. arXiv:1004.0693, doi:10.1103/PhysRevD.82.065003.

[228] Z. Bern, J. J. M. Carrasco, H. Johansson, Perturbative Quantum Gravity as a Double Copy of Gauge Theory, Phys. Rev. Lett. 105 (2010) 061602. arXiv:1004.0476, doi:10.1103/PhysRevLett.105.061602.

[229] N. E. J. Bjerrum-Bohr, P. H. Damgaard, T. Sondergaard, P. Vanhove, Monodromy and Jacobi-like Relations for Color-Ordered Amplitudes, JHEP 06 (2010) 003. arXiv:1003.2403, doi:10.1007/JHEP06(2010)003.

[230] L. J. Dixon, Calculating scattering amplitudes efficiently, in: QCD and beyond. Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics, TASI-95, Boulder, USA, June 4-30, 1995, 1996, pp. 539–584. arXiv:hep-ph/9601359.
URL http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=SLAC-PUB-7106

[231] A. Brandhuber, G. R. Brown, G. Chen, J. Gowdy, G. Travaglini, C. Wen, Amplitudes, Hopf algebras and the colour-kinematics duality (8 2022). arXiv:2208.05886.

[232] G. Chen, G. Lin, C. Wen, Kinematic Hopf algebra for amplitudes and form factors (8 2022). arXiv:2208.05519.

[233] A. Brandhuber, G. Chen, H. Johansson, G. Travaglini, C. Wen, Kinematic Hopf Algebra for Bern-Carrasco-Johansson Numerators in Heavy-Mass Effective Field Theory and Yang-Mills Theory, Phys. Rev. Lett. 128 (12) (2022) 121601. arXiv:2111.15649, doi:10.1103/PhysRevLett.128.121601.

[234] Q. Ma, Y.-J. Du, Y.-X. Chen, On Primary Relations at Tree-level in String Theory and Field Theory, JHEP 02 (2012)

061. arXiv:1109.0685, doi:10.1007/JHEP02(2012)061.
[235] S. Mizera, Inverse of the String Theory KLT Kernel, JHEP 06 (2017) 084. arXiv:1610.04230, doi:10.1007/JHEP06(2017)084.
[236] K. Cho, K. Matsumoto, Intersection theory for twisted cohomologies and twisted riemann's period relations. i, Nagoya Math. J. 139 (1995) 67–86 [https://projecteuclid.org:443/euclid.nmj/1118775097].
[237] M. B. Green, M. Gutperle, Symmetry breaking at enhanced symmetry points, Nucl. Phys. B460 (1996) 77–108. arXiv:hep-th/9509171, doi:10.1016/0550-3213(95)00608-7.
[238] R. R. Metsaev, M. Rakhmanov, A. A. Tseytlin, The Born-Infeld Action as the Effective Action in the Open Superstring Theory, Phys. Lett. B193 (1987) 207–212. doi:10.1016/0370-2693(87)91223-8.
[239] A. A. Tseytlin, Born-Infeld action, supersymmetry and string theory (1999) 417–452arXiv:hep-th/9908105, doi:10.1142/9789812793850_0025.
[240] R. Kallosh, Volkov-Akulov theory and D-branes, Lect. Notes Phys. 509 (1998) 49. arXiv:hep-th/9705118, doi:10.1007/BFb0105228.
[241] E. Bergshoeff, F. Coomans, R. Kallosh, C. S. Shahbazi, A. Van Proeyen, Dirac-Born-Infeld-Volkov-Akulov and Deformation of Supersymmetry, JHEP 08 (2013) 100. arXiv:1303.5662, doi:10.1007/JHEP08(2013)100.
[242] F. Cachazo, S. He, E. Y. Yuan, Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM, JHEP 07 (2015) 149. arXiv:1412.3479, doi:10.1007/JHEP07(2015)149.
[243] J. A. Cronin, Phenomenological model of strong and weak interactions in chiral $U(3) \times U(3)$, Phys. Rev. 161 (1967) 1483–1494. doi:10.1103/PhysRev.161.1483.
[244] S. Weinberg, Dynamical approach to current algebra, Phys. Rev. Lett. 18 (1967) 188–191. doi:10.1103/PhysRevLett.18.188.
[245] S. Weinberg, Nonlinear realizations of chiral symmetry, Phys. Rev. 166 (1968) 1568–1577. doi:10.1103/PhysRev.166.1568.
[246] L. S. Brown, Field theory of chiral symmetry, Phys. Rev. 163 (1967) 1802–1807. doi:10.1103/PhysRev.163.1802.
[247] P. Chang, F. Gursey, Unified Formulation of Effective Nonlinear Pion-Nucleon Lagrangians, Phys. Rev. 164 (1967) 1752–1761. doi:10.1103/PhysRev.164.1752.
[248] H. Osborn, Implications of adler zeros for multipion processes, Lett. Nuovo Cim. 2S1 (1969) 717–723, [Lett. Nuovo Cim.2,717(1969)]. doi:10.1007/BF02755724.
[249] L. Susskind, G. Frye, Algebraic aspects of pionic duality diagrams, Phys. Rev. D1 (1970) 1682–1686. doi:10.1103/PhysRevD.1.1682.
[250] J. R. Ellis, B. Renner, On the relationship between chiral and dual models, Nucl. Phys. B21 (1970) 205–216. doi:10.1016/0550-3213(70)90515-8.
[251] K. Kampf, J. Novotny, J. Trnka, Tree-level Amplitudes in the Nonlinear Sigma Model, JHEP 05 (2013) 032. arXiv:1304.3048, doi:10.1007/JHEP05(2013)032.
[252] G. Chen, Y.-J. Du, Amplitude Relations in Non-linear Sigma Model, JHEP 01 (2014) 061. arXiv:1311.1133, doi:10.1007/JHEP01(2014)061.
[253] J. J. M. Carrasco, C. R. Mafra, O. Schlotterer, Semi-abelian Z-theory: NLSM+ϕ^3 from the open string, JHEP 08 (2017) 135. arXiv:1612.06446, doi:10.1007/JHEP08(2017)135.
[254] C. Cheung, C.-H. Shen, Symmetry for Flavor-Kinematics Duality from an Action, Phys. Rev. Lett. 118 (12) (2017) 121601. arXiv:1612.00868, doi:10.1103/PhysRevLett.118.121601.
[255] F. Cachazo, P. Cha, S. Mizera, Extensions of Theories from Soft Limits, JHEP 06 (2016) 170. arXiv:1604.03893, doi:10.1007/JHEP06(2016)170.
[256] D. J. Gross, J. A. Harvey, E. J. Martinec, R. Rohm, Heterotic String Theory. 2. The Interacting Heterotic String, Nucl. Phys. B 267 (1986) 75–124. doi:10.1016/0550-3213(86)90146-X.
[257] I. Frenkel, M. Zhu, Vertex operator algebras associated to representations of affine and Virasoro algebras, Duke Math J. 66 (1992) 123.
[258] M. Chiodaroli, M. Günaydin, H. Johansson, R. Roiban, Scattering amplitudes in $\mathcal{N} = 2$ Maxwell-Einstein and Yang-Mills/Einstein supergravity, JHEP 01 (2015) 081. arXiv:1408.0764, doi:10.1007/JHEP01(2015)081.
[259] O. Schlotterer, Amplitude relations in heterotic string theory and Einstein-Yang-Mills, JHEP 11 (2016) 074. arXiv:1608.00130, doi:10.1007/JHEP11(2016)074.
[260] S. He, F. Teng, Y. Zhang, String amplitudes from field-theory amplitudes and vice versa, Phys. Rev. Lett. 122 (21) (2019) 211603. arXiv:1812.03369, doi:10.1103/PhysRevLett.122.211603.
[261] S. He, F. Teng, Y. Zhang, String Correlators: Recursive Expansion, Integration-by-Parts and Scattering Equations, JHEP 09 (2019) 085. arXiv:1907.06041, doi:10.1007/JHEP09(2019)085.
[262] H. Johansson, J. Nohle, Conformal Gravity from Gauge Theory (7 2017). arXiv:1707.02965.
[263] M. Chiodaroli, M. Gunaydin, H. Johansson, R. Roiban, Explicit Formulae for Yang-Mills-Einstein Amplitudes from the Double Copy, JHEP 07 (2017) 002. arXiv:1703.00421, doi:10.1007/JHEP07(2017)002.
[264] S. Stieberger, T. R. Taylor, New relations for Einstein–Yang–Mills amplitudes, Nucl. Phys. B 913 (2016) 151–162. arXiv:1606.09616, doi:10.1016/j.nuclphysb.2016.09.014.
[265] F. Cachazo, S. He, E. Y. Yuan, Einstein-Yang-Mills Scattering Amplitudes From Scattering Equations, JHEP 01 (2015) 121. arXiv:1409.8256, doi:10.1007/JHEP01(2015)121.
[266] D. Nandan, J. Plefka, O. Schlotterer, C. Wen, Einstein-Yang-Mills from pure Yang-Mills amplitudes, JHEP 10 (2016) 070. arXiv:1607.05701, doi:10.1007/JHEP10(2016)070.
[267] P. Mazloumi, S. Stieberger, Einstein Yang-Mills amplitudes from intersections of twisted forms, JHEP 06 (2022) 125. arXiv:2201.00837, doi:10.1007/JHEP06(2022)125.

[268] Y.-J. Du, B. Feng, F. Teng, Expansion of All Multitrace Tree Level EYM Amplitudes, JHEP 12 (2017) 038. arXiv:1708.04514, doi:10.1007/JHEP12(2017)038.

[269] J. Broedel, O. Schlotterer, S. Stieberger, T. Terasoma, All order α'-expansion of superstring trees from the Drinfeld associator, Phys. Rev. D89 (6) (2014) 066014. arXiv:1304.7304, doi:10.1103/PhysRevD.89.066014.

[270] S. Stieberger, Constraints on Tree-Level Higher Order Gravitational Couplings in Superstring Theory, Phys. Rev. Lett. 106 (2011) 111601. arXiv:0910.0180, doi:10.1103/PhysRevLett.106.111601.

[271] O. Schlotterer, S. Stieberger, Motivic Multiple Zeta Values and Superstring Amplitudes, J. Phys. A46 (2013) 475401. arXiv:1205.1516, doi:10.1088/1751-8113/46/47/475401.

[272] R. H. Boels, On the field theory expansion of superstring five point amplitudes, Nucl. Phys. B 876 (2013) 215–233. arXiv:1304.7918, doi:10.1016/j.nuclphysb.2013.08.009.

[273] G. Puhlfürst, S. Stieberger, Differential Equations, Associators, and Recurrences for Amplitudes, Nucl. Phys. B 902 (2016) 186–245. arXiv:1507.01582, doi:10.1016/j.nuclphysb.2015.11.005.

[274] T. Terasoma, Selberg integrals and multiple zeta values, Compositio Math. 133 (1) (2002) 1–24. doi:10.1023/A:1016377828316.

[275] A. B. Goncharov, Galois symmetries of fundamental groupoids and noncommutative geometry, Duke Math. J. 128 (2005) 209. arXiv:math/0208144, doi:10.1215/S0012-7094-04-12822-2.

[276] F. Brown, Multiple zeta values and periods of moduli spaces $\mathcal{M}_{0,n}(\mathbb{R})$, Annales Sci. Ecole Norm. Sup. 42 (2009) 371. arXiv:math/0606419.

[277] F. Brown, On the decomposition of motivic multiple zeta values, in: Galois-Teichmüller theory and arithmetic geometry, Vol. 63 of Adv. Stud. Pure Math., Math. Soc. Japan, Tokyo, 2012, pp. 31–58. arXiv:1102.1310.

[278] O. Schnetz, Graphical functions and single-valued multiple polylogarithms, Commun. Num. Theor. Phys. 08 (2014) 589–675. arXiv:1302.6445, doi:10.4310/CNTP.2014.v8.n4.a1.

[279] F. Brown, Single-valued Motivic Periods and Multiple Zeta Values, SIGMA 2 (2014) e25. arXiv:1309.5309, doi:10.1017/fms.2014.18.

[280] F. Brown, C. Dupont, Single-valued integration and double copy, J. Reine Angew. Math. 2021 (775) (2021) 145–196. arXiv:1810.07682, doi:10.1515/crelle-2020-0042.

[281] O. Schlotterer, The number theory of superstring amplitudes, in: J. I. Burgos Gil, K. Ebrahimi-Fard, H. Gangl (Eds.), Periods in Quantum Field Theory and Arithmetic. doi:10.1007/978-3-030-37031-2.

[282] S. Stieberger, Periods and superstring amplitudes, in: J. I. Burgos Gil, K. Ebrahimi-Fard, H. Gangl (Eds.), Periods in Quantum Field Theory and Arithmetic. arXiv:1605.03630.

[283] C. Duhr, Hopf algebras, coproducts and symbols: an application to Higgs boson amplitudes, JHEP 08 (2012) 043. arXiv:1203.0454, doi:10.1007/JHEP08(2012)043.

[284] D. M. Richards, The One-Loop Five-Graviton Amplitude and the Effective Action, JHEP 10 (2008) 042. arXiv:0807.2421, doi:10.1088/1126-6708/2008/10/042.

[285] L. A. Barreiro, R. Medina, Revisiting the S-matrix approach to the open superstring low energy effective lagrangian, JHEP 10 (2012) 108. arXiv:1208.6066, doi:10.1007/JHEP10(2012)108.

[286] L. A. Barreiro, R. Medina, RNS derivation of N-point disk amplitudes from the revisited S-matrix approach, Nucl. Phys. B 886 (2014) 870–951. arXiv:1310.5942, doi:10.1016/j.nuclphysb.2014.07.015.

[287] P. Koerber, A. Sevrin, The NonAbelian D-brane effective action through order α'^4, JHEP 10 (2002) 046. arXiv:hep-th/0208044, doi:10.1088/1126-6708/2002/10/046.

[288] P. S. Howe, U. Lindstrom, L. Wulff, D=10 supersymmetric Yang-Mills theory at α'^4, JHEP 07 (2010) 028. arXiv:1004.3466, doi:10.1007/JHEP07(2010)028.

[289] A. A. Tseytlin, Vector Field Effective Action in the Open Superstring Theory, Nucl. Phys. B 276 (1986) 391, [Erratum: Nucl.Phys.B 291, 876 (1987)]. doi:10.1016/0550-3213(86)90303-2.

[290] A. A. Tseytlin, On nonAbelian generalization of Born-Infeld action in string theory, Nucl. Phys. B 501 (1997) 41–52. arXiv:hep-th/9701125, doi:10.1016/S0550-3213(97)00354-4.

[291] M. Cederwall, B. E. W. Nilsson, D. Tsimpis, Spinorial cohomology of Abelian D=10 superYang-Mills at O(α'^3), JHEP 11 (2002) 023. arXiv:hep-th/0205165, doi:10.1088/1126-6708/2002/11/023.

[292] D. J. Gross, E. Witten, Superstring Modifications of Einstein's Equations, Nucl. Phys. B277 (1986) 1. doi:10.1016/0550-3213(86)90429-3.

[293] K. Peeters, J. Plefka, S. Stern, Higher-derivative gauge field terms in the M-theory action, JHEP 08 (2005) 095. arXiv:hep-th/0507178, doi:10.1088/1126-6708/2005/08/095.

[294] J. T. Liu, R. Minasian, R. Savelli, A. Schachner, Type IIB at eight derivatives: insights from Superstrings, Superfields and Superparticles, JHEP 08 (2022) 267. arXiv:2205.11530, doi:10.1007/JHEP08(2022)267.

[295] M. T. Grisaru, A. E. M. van de Ven, D. Zanon, Four Loop beta Function for the N=1 and N=2 Supersymmetric Nonlinear Sigma Model in Two-Dimensions, Phys. Lett. B 173 (1986) 423–428. doi:10.1016/0370-2693(86)90408-9.

[296] M. T. Grisaru, D. Zanon, σ Model Superstring Corrections to the Einstein-hilbert Action, Phys. Lett. B 177 (1986) 347–351. doi:10.1016/0370-2693(86)90765-3.

[297] M. D. Freeman, C. N. Pope, M. F. Sohnius, K. S. Stelle, Higher Order σ Model Counterterms and the Effective Action for Superstrings, Phys. Lett. B 178 (1986) 199–204. doi:10.1016/0370-2693(86)91495-4.

[298] M. B. Green, M. Gutperle, H.-h. Kwon, Sixteen fermion and related terms in M theory on T^2, Phys. Lett. B 421 (1998) 149–161. arXiv:hep-th/9710151, doi:10.1016/S0370-2693(97)01551-7.

[299] E. Bergshoeff, M. Rakowski, E. Sezgin, Higher Derivative Super Yang-Mills Theories, Phys. Lett. B 185 (1987) 371–376. doi:10.1016/0370-2693(87)91017-3.

[300] S. J. Gates, Jr., S. Vashakidze, On $D = 10$, $N = 1$ Supersymmetry, Superspace Geometry and Superstring Effects, Nucl.

Phys. B 291 (1987) 172. doi:10.1016/0550-3213(87)90470-6.
[301] M. Cederwall, B. E. W. Nilsson, D. Tsimpis, The Structure of maximally supersymmetric Yang-Mills theory: Constraining higher order corrections, JHEP 06 (2001) 034. arXiv:hep-th/0102009, doi:10.1088/1126-6708/2001/06/034.
[302] M. Cederwall, B. E. W. Nilsson, D. Tsimpis, D = 10 superYang-Mills at $O(\alpha'^2)$, JHEP 07 (2001) 042. arXiv:hep-th/0104236, doi:10.1088/1126-6708/2001/07/042.
[303] M. B. Green, S. Sethi, Supersymmetry constraints on type IIB supergravity, Phys. Rev. D 59 (1999) 046006. arXiv:hep-th/9808061, doi:10.1103/PhysRevD.59.046006.
[304] P. Koerber, A. Sevrin, The NonAbelian Born-Infeld action through order α'^3, JHEP 10 (2001) 003. arXiv:hep-th/0108169, doi:10.1088/1126-6708/2001/10/003.
[305] A. Collinucci, M. De Roo, M. G. C. Eenink, Supersymmetric Yang-Mills theory at order α'^3, JHEP 06 (2002) 024. arXiv:hep-th/0205150, doi:10.1088/1126-6708/2002/06/024.
[306] N. Berkovits, V. Pershin, Supersymmetric Born-Infeld from the pure spinor formalism of the open superstring, JHEP 01 (2003) 023. arXiv:hep-th/0205154, doi:10.1088/1126-6708/2003/01/023.
[307] J. M. Drummond, P. J. Heslop, P. S. Howe, S. F. Kerstan, Integral invariants in N=4 SYM and the effective action for coincident D-branes, JHEP 08 (2003) 016. arXiv:hep-th/0305202, doi:10.1088/1126-6708/2003/08/016.
[308] M. Cederwall, A. Karlsson, Pure spinor superfields and Born-Infeld theory, JHEP 11 (2011) 134. arXiv:1109.0809, doi:10.1007/JHEP11(2011)134.
[309] Y. Wang, X. Yin, Constraining Higher Derivative Supergravity with Scattering Amplitudes, Phys. Rev. D 92 (4) (2015) 041701. arXiv:1502.03810, doi:10.1103/PhysRevD.92.041701.
[310] J. Zhao, Multiple zeta functions, multiple polylogarithms, and their special values, World Scientific, New Jersey, 2016.
[311] J. I. B. Gil, J. Fresan, Multiple zeta values: from numbers to motives.
URL http://javier.fresan.perso.math.cnrs.fr/mzv.pdf
[312] J. Bluemlein, D. J. Broadhurst, J. A. M. Vermaseren, The Multiple Zeta Value Data Mine, Comput. Phys. Commun. 181 (2010) 582–625. arXiv:0907.2557, doi:10.1016/j.cpc.2009.11.007.
[313] D. Zagier, Values of zeta functions and their applications, First European Congress of Mathematics II (1992).
[314] F. Brown, Mixed Tate motives over $\mathbb{Z}$, Ann. Math. 175 (2012) 949–976. arXiv:1102.1312.
[315] F. Brown, Notes on motivic periods, Communications in Number Theory and Physics 11 (3) (2015) 557–655. arXiv:1512.06410.
[316] E. Panzer, Feynman integrals and hyperlogarithms, Ph.D. thesis, Humboldt U. (2015). arXiv:1506.07243, doi:10.18452/17157.
[317] V. Drinfeld, Quasi Hopf algebras, Leningrad Math. J. 1 (1989) 1419–1457.
[318] V. Drinfeld, On quasitriangular quasi-Hopf algebras and on a group that is closely connected with Gal$(\overline{\mathbb{Q}}/\mathbb{Q})$, Leningrad Math. J. 2 (4) (1991) 829–860.
[319] T. Le, J. Murakami, Kontsevich's integral for the Kauffman polynomial, Nagoya Math J. 142 (1996) 93–65.
[320] F. Brown, Polylogarithmes multiples uniformes en une variable, C. R. Acad. Sci. Paris Ser. I 338 (2004) 527–532.
[321] J. M. Drummond, E. Ragoucy, Superstring amplitudes and the associator, JHEP 08 (2013) 135. arXiv:1301.0794, doi:10.1007/JHEP08(2013)135.
[322] S. Stieberger, Closed superstring amplitudes, single-valued multiple zeta values and the Deligne associator, J. Phys. A 47 (2014) 155401. arXiv:1310.3259, doi:10.1088/1751-8113/47/15/155401.
[323] S. Abreu, R. Britto, C. Duhr, E. Gardi, Algebraic Structure of Cut Feynman Integrals and the Diagrammatic Coaction, Phys. Rev. Lett. 119 (5) (2017) 051601. arXiv:1703.05064, doi:10.1103/PhysRevLett.119.051601.
[324] S. Abreu, R. Britto, C. Duhr, E. Gardi, Diagrammatic Hopf algebra of cut Feynman integrals: the one-loop case, JHEP 12 (2017) 090. arXiv:1704.07931, doi:10.1007/JHEP12(2017)090.
[325] S. Abreu, R. Britto, C. Duhr, E. Gardi, J. Matthew, From positive geometries to a coaction on hypergeometric functions, JHEP 02 (2020) 122. arXiv:1910.08358, doi:10.1007/JHEP02(2020)122.
[326] S. Abreu, R. Britto, C. Duhr, E. Gardi, J. Matthew, The diagrammatic coaction beyond one loop, JHEP 10 (2021) 131. arXiv:2106.01280, doi:10.1007/JHEP10(2021)131.
[327] F. Brown, C. Dupont, Lauricella hypergeometric functions, unipotent fundamental groups of the punctured Riemann sphere, and their motivic coactions, Nagoya Mathematical Journal (2022) 1–73arXiv:1907.06603, doi:10.1017/nmj.2022.27.
[328] J. Broedel, O. Schlotterer, S. Stieberger, α'-expansion of open superstring amplitudes.
URL https://wwwth.mpp.mpg.de/members/stieberg/mzv/index.html
[329] N. E. J. Bjerrum-Bohr, P. H. Damgaard, H. Johansson, T. Sondergaard, Monodromy–like Relations for Finite Loop Amplitudes, JHEP 05 (2011) 039. arXiv:1103.6190, doi:10.1007/JHEP05(2011)039.
[330] L. Solomon, On the poincaré-birkhoff-witt theorem, Journal of Combinatorial Theory 4 (4) (1968) 363–375.
[331] C. Reutenauer, Theorem of poincaré-birkhoff-witt, logarithm and symmetric group representations of degrees equal to stirling numbers, in: Combinatoire énumérative, Springer, 1986, pp. 267–284.
[332] J.-L. Loday, Série de Hausdorff, idempotents eulériens et algebres de Hopf, Université Louis Pasteur, Département de mathématique, 1992.
[333] R. L. Graham, D. E. Knuth, O. Patashnik, S. Liu, Concrete mathematics: a foundation for computer science, Computers in Physics 3 (5) (1989) 106–107.
[334] D. E. Knuth, Two notes on notation, The American Mathematical Monthly 99 (5) (1992) 403–422.
[335] S. Mizera, B. Skrzypek, Perturbiner Methods for Effective Field Theories and the Double Copy, JHEP 10 (2018) 018. arXiv:1809.02096, doi:10.1007/JHEP10(2018)018.
[336] C. Mafra, O. Schlotterer, BGap: α'-expansion of disk integrals.

URL https://repo.or.cz/BGap.git
[337] A. Kaderli, A note on the Drinfeld associator for genus-zero superstring amplitudes in twisted de Rham theory, J. Phys. A 53 (41) (2020) 415401. arXiv:1912.09406, doi:10.1088/1751-8121/ab9462.
[338] J. M. Henn, Multiloop integrals in dimensional regularization made simple, Phys. Rev. Lett. 110 (2013) 251601. arXiv:1304.1806, doi:10.1103/PhysRevLett.110.251601.
[339] L. Adams, S. Weinzierl, The ε-form of the differential equations for Feynman integrals in the elliptic case, Phys. Lett. B 781 (2018) 270–278. arXiv:1802.05020, doi:10.1016/j.physletb.2018.04.002.
[340] A. V. Kotikov, Differential equations method: New technique for massive Feynman diagrams calculation, Phys. Lett. B 254 (1991) 158–164. doi:10.1016/0370-2693(91)90413-K.
[341] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, J. Trnka, Local Integrals for Planar Scattering Amplitudes, JHEP 06 (2012) 125. arXiv:1012.6032, doi:10.1007/JHEP06(2012)125.
[342] J. Broedel, C. Duhr, F. Dulat, B. Penante, L. Tancredi, Elliptic Feynman integrals and pure functions, JHEP 01 (2019) 023. arXiv:1809.10698, doi:10.1007/JHEP01(2019)023.
[343] F. Caola, W. Chen, C. Duhr, X. Liu, B. Mistlberger, F. Petriello, G. Vita, S. Weinzierl, The Path forward to N^3LO, in: 2022 Snowmass Summer Study, 2022. arXiv:2203.06730.
[344] J. L. Bourjaily, et al., Functions Beyond Multiple Polylogarithms for Precision Collider Physics, in: 2022 Snowmass Summer Study, 2022. arXiv:2203.07088.
[345] S. Weinzierl, Feynman Integrals (1 2022). arXiv:2201.03593.
[346] S. Abreu, R. Britto, C. Duhr, The SAGEX Review on Scattering Amplitudes, Chapter 3: Mathematical structures in Feynman integrals (3 2022). arXiv:2203.13014.
[347] J. Blümlein, C. Schneider, The SAGEX Review on Scattering Amplitudes, Chapter 4: Multi-loop Feynman Integrals (3 2022). arXiv:2203.13015.
[348] A. Selberg, Remarks on a multiple integral, Norsk Matematisk Tidsskrift 26 (1944) 71.
[349] R. Britto, S. Mizera, C. Rodriguez, O. Schlotterer, Coaction and double-copy properties of configuration-space integrals at genus zero, JHEP 05 (2021) 053. arXiv:2102.06206, doi:10.1007/JHEP05(2021)053.
[350] G. Felder, A. Varchenko, Integral representation of solutions of the elliptic Knizhnik-Zamolodchikov-Bernard equations, International Mathematics Research Notices (1995) 221arXiv:9502165.
[351] T. Mano, H. Watanabe, Twisted cohomology and homology groups associated to the Riemann-Wirtinger integral, Proceedings of the American Mathematical Society 140 (11) (2012) 3867.
[352] C. R. Mafra, O. Schlotterer, All Order α' Expansion of One-Loop Open-String Integrals, Phys. Rev. Lett. 124 (10) (2020) 101603. arXiv:1908.09848, doi:10.1103/PhysRevLett.124.101603.
[353] C. R. Mafra, O. Schlotterer, One-loop open-string integrals from differential equations: all-order α'-expansions at n points, JHEP 03 (2020) 007. arXiv:1908.10830, doi:10.1007/JHEP03(2020)007.
[354] J. Broedel, A. Kaderli, Amplitude recursions with an extra marked point, Commun. Num. Theor. Phys. 16 (1) (2022) 75–158. arXiv:1912.09927, doi:10.4310/CNTP.2022.v16.n1.a3.
[355] J. Broedel, A. Kaderli, O. Schlotterer, Two dialects for KZB equations: generating one-loop open-string integrals, JHEP 12 (2020) 036. arXiv:2007.03712, doi:10.1007/JHEP12(2020)036.
[356] A. Kaderli, C. Rodriguez, Open-string integrals with multiple unintegrated punctures at genus one (3 2022). arXiv:2203.09649.
[357] A. Beilinson, A. Varchenko, A. Goncharov, V. Shekhtman, Projective Geometry and K-theory, Leningrad Math J. 2 3 (1991) 523–575.
[358] O. Schnetz, Graphical hyperlogarithms, talk within the trimester “Periods in Number Theory, Algebraic Geometry and Physics”, HIM, Bonn, Feb. 27, 2018.
[359] O. Schlotterer, O. Schnetz, Closed strings as single-valued open strings: A genus-zero derivation, J. Phys. A 52 (4) (2019) 045401. arXiv:1808.00713, doi:10.1088/1751-8121/aaea14.
[360] P. Vanhove, F. Zerbini, Single-valued hyperlogarithms, correlation functions and closed string amplitudes, Adv. Theor. Math. Phys. 26 (2) (12 2022). arXiv:1812.03018.
[361] O. Hohm, W. Siegel, B. Zwiebach, Doubled α'-geometry, JHEP 02 (2014) 065. arXiv:1306.2970, doi:10.1007/JHEP02(2014)065.
[362] W. Siegel, Amplitudes for left-handed strings (12 2015). arXiv:1512.02569.
[363] Y.-t. Huang, W. Siegel, E. Y. Yuan, Factorization of Chiral String Amplitudes, JHEP 09 (2016) 101. arXiv:1603.02588, doi:10.1007/JHEP09(2016)101.
[364] R. Lipinski Jusinskas, Chiral strings, the sectorized description and their integrated vertex operators, JHEP 12 (2019) 143. arXiv:1909.04069, doi:10.1007/JHEP12(2019)143.
[365] T. Azevedo, R. L. Jusinskas, M. Lize, Bosonic sectorized strings and the $(DF)^2$ theory, JHEP 01 (2020) 082. arXiv:1908.11371, doi:10.1007/JHEP01(2020)082.
[366] M. Guillen, H. Johansson, R. L. Jusinskas, O. Schlotterer, Scattering Massive String Resonances through Field-Theory Methods, Phys. Rev. Lett. 127 (5) (2021) 051601. arXiv:2104.03314, doi:10.1103/PhysRevLett.127.051601.
[367] N. Berkovits, C. R. Mafra, Pure spinor formulation of the superstring and its applications (10 2022). arXiv:2210.10510.
[368] E. D’Hoker, D. H. Phong, Conformal Scalar Fields and Chiral Splitting on Superriemann Surfaces, Commun. Math. Phys. 125 (1989) 469. doi:10.1007/BF01218413.
[369] N. Berkovits, New higher-derivative R^4 theorems, Phys. Rev. Lett. 98 (2007) 211601. arXiv:hep-th/0609006, doi:10.1103/PhysRevLett.98.211601.
[370] N. Berkovits, M. B. Green, J. G. Russo, P. Vanhove, Non-renormalization conditions for four-gluon scattering in supersymmetric string and field theory, JHEP 11 (2009) 063. arXiv:0908.1923, doi:10.1088/1126-6708/2009/11/063.

[371] N. Berkovits, C. R. Mafra, Equivalence of two-loop superstring amplitudes in the pure spinor and RNS formalisms, Phys. Rev. Lett. 96 (2006) 011602. arXiv:hep-th/0509234, doi:10.1103/PhysRevLett.96.011602.
[372] E. D'Hoker, D. H. Phong, Two-loop superstrings VI: Non-renormalization theorems and the 4-point function, Nucl. Phys. B 715 (2005) 3–90. arXiv:hep-th/0501197, doi:10.1016/j.nuclphysb.2005.02.043.
[373] C. R. Mafra, C. Stahn, The One-loop Open Superstring Massless Five-point Amplitude with the Non-Minimal Pure Spinor Formalism, JHEP 03 (2009) 126. arXiv:0902.1539, doi:10.1088/1126-6708/2009/03/126.
[374] H. Gomez, C. R. Mafra, O. Schlotterer, Two-loop superstring five-point amplitude and S-duality, Phys. Rev. D 93 (4) (2016) 045030. arXiv:1504.02759, doi:10.1103/PhysRevD.93.045030.
[375] C. R. Mafra, O. Schlotterer, One-loop superstring six-point amplitudes and anomalies in pure spinor superspace, JHEP 04 (2016) 148. arXiv:1603.04790, doi:10.1007/JHEP04(2016)148.
[376] M. B. Green, J. H. Schwarz, The Hexagon Gauge Anomaly in Type I Superstring Theory, Nucl. Phys. B 255 (1985) 93–114. doi:10.1016/0550-3213(85)90130-0.
[377] M. B. Green, J. H. Schwarz, Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory, Phys. Lett. B 149 (1984) 117–122. doi:10.1016/0370-2693(84)91565-X.
[378] N. Berkovits, C. R. Mafra, Some Superstring Amplitude Computations with the Non-Minimal Pure Spinor Formalism, JHEP 11 (2006) 079. arXiv:hep-th/0607187, doi:10.1088/1126-6708/2006/11/079.
[379] C. R. Mafra, O. Schlotterer, Towards the n-point one-loop superstring amplitude. Part I. Pure spinors and superfield kinematics, JHEP 08 (2019) 090. arXiv:1812.10969, doi:10.1007/JHEP08(2019)090.
[380] C. R. Mafra, O. Schlotterer, Towards the n-point one-loop superstring amplitude. Part II. Worldsheet functions and their duality to kinematics, JHEP 08 (2019) 091. arXiv:1812.10970, doi:10.1007/JHEP08(2019)091.
[381] C. R. Mafra, O. Schlotterer, Double-Copy Structure of One-Loop Open-String Amplitudes, Phys. Rev. Lett. 121 (1) (2018) 011601. arXiv:1711.09104, doi:10.1103/PhysRevLett.121.011601.
[382] E. D'Hoker, O. Schlotterer, Two-loop superstring five-point amplitudes. Part III. Construction via the RNS formulation: even spin structures, JHEP 12 (2021) 063. arXiv:2108.01104, doi:10.1007/JHEP12(2021)063.
[383] Y. Geyer, R. Monteiro, R. Stark-Muchão, Superstring Loop Amplitudes from the Field Theory Limit, Phys. Rev. Lett. 127 (21) (2021) 211603. arXiv:2106.03968, doi:10.1103/PhysRevLett.127.211603.
[384] J. J. M. Carrasco, H. Johansson, Five-Point Amplitudes in N=4 Super-Yang-Mills Theory and N=8 Supergravity, Phys. Rev. D 85 (2012) 025006. arXiv:1106.4711, doi:10.1103/PhysRevD.85.025006.
[385] P. Tourkine, Tropical Amplitudes, Annales Henri Poincaré 18 (6) (2017) 2199–2249. arXiv:1309.3551, doi:10.1007/s00023-017-0560-7.
[386] S. He, R. Monteiro, O. Schlotterer, String-inspired BCJ numerators for one-loop MHV amplitudes, JHEP 01 (2016) 171. arXiv:1507.06288, doi:10.1007/JHEP01(2016)171.
[387] S. He, O. Schlotterer, Y. Zhang, New BCJ representations for one-loop amplitudes in gauge theories and gravity, Nucl. Phys. B 930 (2018) 328–383. arXiv:1706.00640, doi:10.1016/j.nuclphysb.2018.03.003.
[388] Y. Geyer, L. Mason, R. Monteiro, P. Tourkine, Loop Integrands for Scattering Amplitudes from the Riemann Sphere, Phys. Rev. Lett. 115 (12) (2015) 121603. arXiv:1507.00321, doi:10.1103/PhysRevLett.115.121603.
[389] S. He, O. Schlotterer, New Relations for Gauge-Theory and Gravity Amplitudes at Loop Level, Phys. Rev. Lett. 118 (16) (2017) 161601. arXiv:1612.00417, doi:10.1103/PhysRevLett.118.161601.
[390] E. Bridges, C. R. Mafra, Local BCJ numerators for ten-dimensional SYM at one loop, JHEP 07 (2021) 031. arXiv:2102.12943, doi:10.1007/JHEP07(2021)031.
[391] M. Ben-Shahar, M. Guillen, 10D super-Yang-Mills scattering amplitudes from its pure spinor action, JHEP 12 (2021) 014. arXiv:2108.11708, doi:10.1007/JHEP12(2021)014.
[392] Y. Aisaka, N. Berkovits, Pure Spinor Vertex Operators in Siegel Gauge and Loop Amplitude Regularization, JHEP 07 (2009) 062. arXiv:0903.3443, doi:10.1088/1126-6708/2009/07/062.
[393] P. A. Grassi, P. Vanhove, Higher-loop amplitudes in the non-minimal pure spinor formalism, JHEP 05 (2009) 089. arXiv:0903.3903, doi:10.1088/1126-6708/2009/05/089.
[394] J. Broedel, C. R. Mafra, N. Matthes, O. Schlotterer, Elliptic multiple zeta values and one-loop superstring amplitudes, JHEP 07 (2015) 112. arXiv:1412.5535, doi:10.1007/JHEP07(2015)112.
[395] P. Tourkine, P. Vanhove, Higher-loop amplitude monodromy relations in string and gauge theory, Phys. Rev. Lett. 117 (21) (2016) 211601. arXiv:1608.01665, doi:10.1103/PhysRevLett.117.211601.
[396] S. Hohenegger, S. Stieberger, Monodromy Relations in Higher-Loop String Amplitudes, Nucl. Phys. B 925 (2017) 63–134. arXiv:1702.04963, doi:10.1016/j.nuclphysb.2017.09.020.
[397] P. Tourkine, Integrands and loop momentum in string and field theory, Phys. Rev. D 102 (2) (2020) 026006. arXiv:1901.02432, doi:10.1103/PhysRevD.102.026006.
[398] E. Casali, S. Mizera, P. Tourkine, Monodromy relations from twisted homology, JHEP 12 (2019) 087. arXiv:1910.08514, doi:10.1007/JHEP12(2019)087.
[399] E. Casali, S. Mizera, P. Tourkine, Loop amplitudes monodromy relations and color-kinematics duality, JHEP 03 (2021) 048. arXiv:2005.05329, doi:10.1007/JHEP03(2021)048.
[400] S. Stieberger, Open & Closed vs. Pure Open String One-Loop Amplitudes (5 2021). arXiv:2105.06888.
[401] E. D'Hoker, D. H. Phong, Momentum analyticity and finiteness of the one loop superstring amplitude, Phys. Rev. Lett. 70 (1993) 3692–3695. arXiv:hep-th/9302003, doi:10.1103/PhysRevLett.70.3692.
[402] E. D'Hoker, D. H. Phong, Dispersion relations in string theory, Theor. Math. Phys. 98 (1994) 306–316. arXiv:hep-th/9404128, doi:10.1007/BF01102207.
[403] E. D'Hoker, D. H. Phong, The Box graph in superstring theory, Nucl. Phys. B 440 (1995) 24–94. arXiv:hep-th/9410152, doi:10.1016/0550-3213(94)00526-K.

[404] M. B. Green, J. G. Russo, P. Vanhove, Low energy expansion of the four-particle genus-one amplitude in type II superstring theory, JHEP 02 (2008) 020. arXiv:0801.0322, doi:10.1088/1126-6708/2008/02/020.

[405] E. D'Hoker, M. B. Green, P. Vanhove, On the modular structure of the genus-one Type II superstring low energy expansion, JHEP 08 (2015) 041. arXiv:1502.06698, doi:10.1007/JHEP08(2015)041.

[406] E. D'Hoker, M. B. Green, Exploring transcendentality in superstring amplitudes, JHEP 07 (2019) 149. arXiv:1906.01652, doi:10.1007/JHEP07(2019)149.

[407] A. Edison, M. Guillen, H. Johansson, O. Schlotterer, F. Teng, One-loop matrix elements of effective superstring interactions: α'-expanding loop integrands, JHEP 12 (2021) 007. arXiv:2107.08009, doi:10.1007/JHEP12(2021)007.

[408] L. Eberhardt, S. Mizera, Unitarity Cuts of the Worldsheet (8 2022). arXiv:2208.12233.

[409] C. M. Hull, P. K. Townsend, Unity of superstring dualities, Nucl. Phys. B 438 (1995) 109–137. arXiv:hep-th/9410167, doi:10.1016/0550-3213(94)00559-W.

[410] M. B. Green, M. Gutperle, Effects of D instantons, Nucl. Phys. B 498 (1997) 195–227. arXiv:hep-th/9701093, doi:10.1016/S0550-3213(97)00269-1.

[411] M. B. Green, H.-h. Kwon, P. Vanhove, Two loops in eleven-dimensions, Phys. Rev. D 61 (2000) 104010. arXiv:hep-th/9910055, doi:10.1103/PhysRevD.61.104010.

[412] E. D'Hoker, M. Gutperle, D. H. Phong, Two-loop superstrings and S-duality, Nucl. Phys. B 722 (2005) 81–118. arXiv:hep-th/0503180, doi:10.1016/j.nuclphysb.2005.06.010.

[413] M. B. Green, P. Vanhove, Duality and higher derivative terms in M theory, JHEP 01 (2006) 093. arXiv:hep-th/0510027, doi:10.1088/1126-6708/2006/01/093.

[414] E. D'Hoker, M. B. Green, B. Pioline, R. Russo, Matching the D^6R^4 interaction at two-loops, JHEP 01 (2015) 031. arXiv:1405.6226, doi:10.1007/JHEP01(2015)031.

[415] E. D'Hoker, C. R. Mafra, B. Pioline, O. Schlotterer, Two-loop superstring five-point amplitudes. Part II. Low energy expansion and S-duality, JHEP 02 (2021) 139. arXiv:2008.08687, doi:10.1007/JHEP02(2021)139.

[416] R. H. Boels, Maximal R-symmetry violating amplitudes in type IIB superstring theory, Phys. Rev. Lett. 109 (2012) 081602. arXiv:1204.4208, doi:10.1103/PhysRevLett.109.081602.

[417] M. B. Green, C. Wen, Modular Forms and $SL(2,\mathbb{Z})$-covariance of type IIB superstring theory, JHEP 06 (2019) 087. arXiv:1904.13394, doi:10.1007/JHEP06(2019)087.

[418] M. B. Green, Interconnections between type II superstrings, M theory and N=4 supersymmetric Yang-Mills, Lect. Notes Phys. 525 (1999) 22. arXiv:hep-th/9903124, doi:10.1007/BFb0104240.

[419] E. D'Hoker, M. B. Green, Ö. Gürdogan, P. Vanhove, Modular graph functions, Commun. Num. Theor. Phys. 11 (2017) 165–218. arXiv:1512.06779, doi:10.4310/CNTP.2017.v11.n1.a4.

[420] E. D'Hoker, M. B. Green, Identities between modular graph forms, J. Number Theory 189 (2018) 25–80. arXiv:1603.00839, doi:10.1016/j.jnt.2017.11.015.

[421] M. B. Green, P. Vanhove, The Low-energy expansion of the one loop type II superstring amplitude, Phys.Rev. D61 (2000) 104011. arXiv:hep-th/9910056, doi:10.1103/PhysRevD.61.104011.

[422] J. E. Gerken, Modular Graph Forms and Scattering Amplitudes in String Theory (11 2020). arXiv:2011.08647.

[423] E. D'Hoker, J. Kaidi, Lectures on modular forms and strings (8 2022). arXiv:2208.07242.

[424] D. Dorigoni, M. Doroudiani, J. Drewitt, M. Hidding, A. Kleinschmidt, N. Matthes, O. Schlotterer, B. Verbeek, Modular graph forms from equivariant iterated Eisenstein integrals (9 2022). arXiv:2209.06772.

[425] F. Brown, A class of non-holomorphic modular forms I, Res. Math. Sci. 5 (2018) 5:7. arXiv:1707.01230, doi:10.1007/s40687-018-0130-8.

[426] F. Brown, A class of non-holomorphic modular forms II : equivariant iterated Eisenstein integrals, Forum of Mathematics, Sigma 8 (2020) 1. arXiv:1708.03354, doi:10.1017/fms.2020.24.

[427] E. D'Hoker, M. B. Green, Zhang-Kawazumi Invariants and Superstring Amplitudes, J. Number Theor. 144 (2014) 111–150. arXiv:1308.4597.

[428] B. Pioline, A Theta lift representation for the Kawazumi-Zhang and Faltings invariants of genus-two Riemann surfaces, J. Number Theor. 163 (2016) 520–541. arXiv:1504.04182, doi:10.1016/j.jnt.2015.12.021.

[429] E. D'Hoker, M. B. Green, B. Pioline, Higher genus modular graph functions, string invariants, and their exact asymptotics, Commun. Math. Phys. 366 (3) (2019) 927–979. arXiv:1712.06135, doi:10.1007/s00220-018-3244-3.

[430] E. D'Hoker, M. B. Green, B. Pioline, Asymptotics of the D^8R^4 genus-two string invariant, Commun. Num. Theor. Phys. 13 (2019) 351–462. arXiv:1806.02691, doi:10.4310/CNTP.2019.v13.n2.a3.

[431] A. Basu, Eigenvalue equation for genus two modular graphs, JHEP 02 (2019) 046. arXiv:1812.00389, doi:10.1007/JHEP02(2019)046.

[432] E. D'Hoker, O. Schlotterer, Identities among higher genus modular graph tensors, Commun. Num. Theor. Phys. 16 (1) (2022) 35–74. arXiv:2010.00924, doi:10.4310/CNTP.2022.v16.n1.a2.

[433] J. Broedel, N. Matthes, G. Richter, O. Schlotterer, Twisted elliptic multiple zeta values and non-planar one-loop open-string amplitudes, J. Phys. A 51 (28) (2018) 285401. arXiv:1704.03449, doi:10.1088/1751-8121/aac601.

[434] B. Enriquez, Analogues elliptiques des nombres multizétas, Bull. Soc. Math. France 144 (3) (2016) 395–427. arXiv:1301.3042, doi:10.24033/bsmf.2718.

[435] F. Brown, A. Levin, Multiple elliptic polylogarithms (2011). arXiv:1110.6917.

[436] F. Zerbini, Single-valued multiple zeta values in genus 1 superstring amplitudes, Commun. Num. Theor. Phys. 10 (2016) 703–737. arXiv:1512.05689, doi:10.4310/CNTP.2016.v10.n4.a2.

[437] J. Broedel, O. Schlotterer, F. Zerbini, From elliptic multiple zeta values to modular graph functions: open and closed strings at one loop, JHEP 01 (2019) 155. arXiv:1803.00527, doi:10.1007/JHEP01(2019)155.

[438] D. Zagier, F. Zerbini, Genus-zero and genus-one string amplitudes and special multiple zeta values, Commun. Num.

Theor. Phys. 14 (2) (2020) 413–452. arXiv:1906.12339, doi:10.4310/CNTP.2020.v14.n2.a4.
[439] J. E. Gerken, A. Kleinschmidt, O. Schlotterer, All-order differential equations for one-loop closed-string integrals and modular graph forms, JHEP 01 (2020) 064. arXiv:1911.03476, doi:10.1007/JHEP01(2020)064.
[440] J. E. Gerken, A. Kleinschmidt, O. Schlotterer, Generating series of all modular graph forms from iterated Eisenstein integrals, JHEP 07 (07) (2020) 190. arXiv:2004.05156, doi:10.1007/JHEP07(2020)190.
[441] J. E. Gerken, A. Kleinschmidt, C. R. Mafra, O. Schlotterer, B. Verbeek, Towards closed strings as single-valued open strings at genus one, J. Phys. A 55 (2) (2022) 025401. arXiv:2010.10558, doi:10.1088/1751-8121/abe58b.
[442] U. Gran, GAMMA: A Mathematica package for performing gamma matrix algebra and Fierz transformations in arbitrary dimensions (5 2001). arXiv:hep-th/0105086.
[443] S. Guttenberg, Superstrings in General Backgrounds, Ph.D. thesis, Vienna, Tech. U. (2007). arXiv:0807.4968.
[444] D. Z. Freedman, A. Van Proeyen, Supergravity, Cambridge Univ. Press, Cambridge, UK, 2012.
[445] P. Mukhopadhyay, On D-brane boundary state analysis in pure-spinor formalism, JHEP 03 (2006) 066. arXiv:hep-th/0505157, doi:10.1088/1126-6708/2006/03/066.
[446] J. W. van Holten, A. van Proeyen, N=1 supersymmetry algebras in d=2,3,4 mod 8, Journal of Physics A: Mathematical and General 15 (12) (1982) 3763–3783. doi:10.1088/0305-4470/15/12/028.
URL https://doi.org/10.1088/0305-4470/15/12/028
[447] P. Kuusela, "GammaMaP" - A Mathematica Package for Clifford Algebras, Gamma Matrices and Spinors (5 2019). arXiv:1905.00429.
[448] W. Mueck, General (anti-)commutators of gamma matrices (11 2007). arXiv:0711.1436.
[449] V. Alexandrov, D. Krotov, A. Losev, V. Lysov, On Pure Spinor Superfield Formalism, JHEP 10 (2007) 074. arXiv:0705.2191, doi:10.1088/1126-6708/2007/10/074.
[450] H. Georgi, Lie Algebras In Particle Physics: from Isospin To Unified Theories, Frontiers in Physics, Avalon Publishing, 1999.
URL https://books.google.co.uk/books?id=g4yEuH5rBMUC
[451] P. A. Grassi, G. Policastro, P. van Nieuwenhuizen, An Introduction to the covariant quantization of superstrings, Class. Quant. Grav. 20 (2003) S395–S410. arXiv:hep-th/0302147, doi:10.1088/0264-9381/20/12/304.
[452] R. Schiappa, N. Wyllard, D-brane boundary states in the pure spinor superstring, JHEP 07 (2005) 070. arXiv:hep-th/0503123, doi:10.1088/1126-6708/2005/07/070.
[453] Y. Aisaka, Y. Kazama, A New first class algebra, homological perturbation and extension of pure spinor formalism for superstring, JHEP 02 (2003) 017. arXiv:hep-th/0212316, doi:10.1088/1126-6708/2003/02/017.
[454] R. Brauer, H. Weyl, Spinors in n dimensions, American Journal of Mathematics 57 (2) (1935) 425–449.
[455] D. B. Kaplan, M. Unsal, A Euclidean lattice construction of supersymmetric Yang-Mills theories with sixteen supercharges, JHEP 09 (2005) 042. arXiv:hep-lat/0503039, doi:10.1088/1126-6708/2005/09/042.
[456] A. Pais, On spinors in n dimensions, Journal of Mathematical Physics 3 (6) (1962) 1135–1139.
[457] M. Lothaire, Combinatorics on words, Vol. 17, Cambridge university press, 1997.
[458] G. Melançon, C. Reutenauer, Lyndon words, free algebras and shuffles, Canadian Journal of Mathematics 41 (4) (1989) 577–591.
[459] G. Griffing, Dual lie elements and a derivation for the cofree coassociative coalgebra, Proceedings of the American Mathematical Society 123 (11) (1995) 3269–3277.
[460] J. E. Humphreys, Introduction to Lie algebras and representation theory, Vol. 9, Springer Science & Business Media, 2012.
[461] W. Fulton, J. Harris, Representation theory: a first course, Vol. 129, Springer Science & Business Media, 2013.
[462] R. Feger, T. W. Kephart, R. J. Saskowski, LieART 2.0 – A Mathematica application for Lie Algebras and Representation Theory, Comput. Phys. Commun. 257 (2020) 107490. arXiv:1912.10969, doi:10.1016/j.cpc.2020.107490.
[463] C. Stahn, Fermionic superstring loop amplitudes in the pure spinor formalism, JHEP 05 (2007) 034. arXiv:0704.0015, doi:10.1088/1126-6708/2007/05/034.
[464] G. Trivedi, Correlation functions in Berkovits' pure spinor formulation, Mod. Phys. Lett. A 17 (2002) 2239–2248. arXiv:hep-th/0205217, doi:10.1142/S0217732302008939.
[465] J. A. de Azcarraga, J. M. Izquierdo, J. C. Perez Bueno, On the generalizations of Poisson structures, J. Phys. A 30 (1997) L607–L616. arXiv:hep-th/9703019, doi:10.1088/0305-4470/30/18/001.
[466] J. A. de Azcarraga, J. M. Izquierdo, n-ary algebras: A Review with applications, J. Phys. A 43 (2010) 293001. arXiv:1005.1028, doi:10.1088/1751-8113/43/29/293001.