

String Correlators: Recursive Expansion, Integration-by-Parts and Scattering Equations

Song He,^{a,b} Fei Teng^c and Yong Zhang^{a,b}

^a*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

^b*School of Physical Sciences, University of Chinese Academy of Sciences, No.19A Yuquan Road, Beijing 100049, China*

^c*Department of Physics and Astronomy, Uppsala University, 75108 Uppsala, Sweden*

E-mail: songhe@itp.ac.cn, fei.teng@physics.uu.se, yongzhang@itp.ac.cn

ABSTRACT: We further elaborate on the general construction proposed in [1], which connects, via tree-level double copy, massless string amplitudes with color-ordered QFT amplitudes that are given by Cachazo-He-Yuan formulas. The current paper serves as a detailed study of the integration-by-parts procedure for any tree-level massless string correlator outlined in the previous letter. We present two new results in the context of heterotic and (compactified) bosonic string theories. First, we find a new recursive expansion of any multitrace mixed correlator in these theories into a logarithmic part corresponding to the CHY integrand for Yang-Mills-scalar amplitudes, plus correlators with the total number of traces and gluons decreased. By iterating the expansion, we systematically reduce string correlators with any number of subcycles to linear combinations of Parke-Taylor factors and similarly for the case with gluons. Based on this, we then derive a CHY formula for the corresponding $(DF)^2 + \text{YM} + \phi^3$ amplitudes. It is the first closed-form result for such multitrace amplitudes and thus greatly extends our result for the single-trace case. As a byproduct, it gives a new CHY formula for all Yang-Mills-scalar amplitudes. We also study consistency checks of the formula such as factorizations on massless poles.

Contents

1	Introduction and review	1
2	The formula for all multitrace amplitudes	5
2.1	Fusion of traces and gluons	6
2.2	Partitions of set and symmetrized fusions	9
2.3	Half-integrands for multitrace amplitudes	10
3	Recursive expansion of string correlator	13
3.1	Pure-scalar examples	14
3.2	General pure-scalar cases	16
3.3	Inserting one gluon	17
3.4	General cases with gluons	18
4	Derivation of the CHY integrand	20
5	Special massless factorizations	22
5.1	An example: factorization of the triple-trace scalar integrand	24
5.2	Generic integrands	25
6	Conclusion and discussion	26
A	Labeled trees and logarithmic Yang-Mills-scalar integrands	27
B	IBP reduction of multibranch graphs	30
C	Towards the generic recursive expansion	33
C.1	Examples: two and three gluons with a single trace	35

1 Introduction and review

Recently, the study of scattering amplitudes has uncovered new structures and symmetries in various quantum field theories (QFT), as well as surprising connections between them (*cf.* [2–4]). The double-copy construction provides a notable example, which describes gravitational scattering amplitudes as “squares” of gauge-theory ones. At tree level, such relations can be derived as the field-theory limit of the celebrated Kawai-Lewellen-Tye (KLT) relations [5] between tree amplitudes in open and closed string theory [6]. Based on a remarkable duality between color and kinematics due to Bern, Carrasco and Johansson (BCJ) [7], double copy has been extended to quantum regime and become the state-of-the-art method for multiloop calculations in supergravity theories [8–12].

The Cachazo-He-Yuan (CHY) formulation [13, 14] has provided a new way to manifest and extend the double copy. Based on the universal scattering equations connecting kinematics of massless particles to moduli space of punctured Riemann spheres [15], the CHY formula expresses tree amplitudes in a large class of theories as integrals over moduli space localized to the solutions of scattering equations. Together with loop-level generalizations [16–20], they have led to new double-copy realization of various theories [21], and one-loop extensions of KLT relations and amplitude relations [22, 23]. What underpins both tree and loop-level CHY formulas are worldsheet models known as ambitwistor string theory [24, 25], where CHY integrands can be obtained as correlators therein. There has been significant progress [26–30] for connecting ambitwistor string theory to the usual string theory, but a complete understanding is still lacking.

String theory has played a crucial role in these developments since the discovery of KLT relations. In particular, amplitude representations that respect color-kinematics duality at tree and loop level have both been realized by string-theory based methods [31–35]. Amplitude relations in gauge theory, *e.g.* BCJ relations [7], and those in Einstein-Yang-Mills (EYM) [36], can also find origin in string theory [37–40]. More interestingly, it has been realized that tree-level superstring amplitudes themselves can be obtained via a double copy [41]. The first example is the discovery that one can decompose disk amplitudes for massless states of type-I theory into field-theory KLT products of universal basis of disk integrals, later called Z integrals, and super-Yang-Mills (SYM) amplitudes: “type-I = $Z \otimes \text{SYM}$ ” [32, 33]. The key point is that all nontrivial α' -dependence of string amplitudes is encoded in the Z integrals, which can also be interpreted as amplitudes in an effective field theory of biadjoint scalars dubbed as Z theory [42–44].

It has been realized in [45, 46] that such a double copy for string amplitudes is general, since it also applies to cases for bosonic and heterotic strings. For (compactified) bosonic open string amplitudes, the same double copy works where the field-theory amplitudes now contain tachyon poles, and they were shown [46] to come from the $(DF)^2 + \text{YM} + \phi^3$ Lagrangian [47], with α' related to its mass parameter:

$$\text{(compactified) bosonic open string} = Z \otimes [(DF)^2 + \text{YM} + \phi^3].$$

Furthermore, by replacing Z integrals by certain sphere integrals, one can generalize the double copy structure to the massless amplitudes of closed and heterotic strings. As conjectured in [48–50] and proven in [51, 52], the latter can be obtained as the single-valued (sv) projection [53, 54] of open-string amplitudes:

$$\begin{aligned} \text{type-II} &= \text{sv}(\text{type-I}) \otimes \text{SYM}, \\ \text{heterotic} &= \text{sv}(\text{type-I}) \otimes [(DF)^2 + \text{YM} + \phi^3]. \end{aligned}$$

In a recent letter [1], we have initiated a systematic study on the double-copy of tree-level massless string amplitude in terms of field-theory amplitudes defined by CHY formulas. The nontrivial part of such CHY formulas can be directly obtained from the original string correlator via an integration-by-parts (IBP) process [1], which we review

here. A generic massless open-string tree amplitude is given by a disk integral:

$$\mathcal{M}_n^{\text{string}}(\rho) = \int_{\rho} \frac{d^n z}{\text{vol SL}(2, \mathbb{R})} \underbrace{\prod_{i < j} |z_{ij}|^{s_{ij}}}_{:= d\mu_n^{\text{string}}} \mathcal{I}_n^{\text{string}}(z), \quad \text{KN} := \prod_{i < j} |z_{ij}|^{s_{ij}}, \quad (1.1)$$

where $z_{ij} := z_i - z_j$ and $s_{ij} := \alpha' k_i \cdot k_j$ are the Mandelstam variables. The color ordering $\rho \in S_n/\mathbb{Z}_n$ is realized by the integration domain $z_{\rho(i)} < z_{\rho(i+1)}$. We denote the Koba-Nielsen factor as KN and the integral measure including it as $d\mu_n^{\text{string}}$. One can fix three punctures, e.g. $(z_1, z_{n-1}, z_n) = (0, 1, \infty)$, using the $\text{SL}(2, \mathbb{R})$ redundancy, and the product in the Koba-Nielsen factor is over $1 \leq i < j \leq n-1$ with this fixing. The (reduced) *string correlator* $\mathcal{I}_n^{\text{string}}$ is a rational function of z 's, and we only require it to have correct $\text{SL}(2)$ weight: $\mathcal{I}_n^{\text{string}} \rightarrow \prod_{a=1}^n (\gamma + \delta z_a)^2 \mathcal{I}_n^{\text{string}}$ under $z_a \rightarrow -\frac{\alpha + \beta z_a}{\gamma + \delta z_a}$ with $\alpha\delta - \beta\gamma = 1$. As shown in [1], using IBP relations, one can write any such integral as a double-copy of field-theory color-ordered amplitudes and the Z integrals

$$\mathcal{M}_n^{\text{string}}(\rho) = \mathcal{M}_n^{\text{FT}} \otimes Z_{\rho} := \sum_{\alpha, \beta \in S_{n-3}} \mathcal{M}_n^{\text{FT}}(\alpha) S[\alpha|\beta] Z_{\rho}(\beta), \quad (1.2)$$

where α, β are color orderings in a minimal basis, and the KLT double copy is defined using $(n-3)!$ -dimensional matrix $S[\alpha|\beta]$ known as the field-theory momentum kernel [6, 55]. The Z integral is a disk integral over a Parke-Taylor (PT) factor of [41]:

$$Z_{\rho}(\pi) := \int_{\rho} d\mu_n^{\text{string}} \text{PT}(\pi), \quad \text{PT}(\pi) := \frac{1}{z_{\pi_1 \pi_2} z_{\pi_2 \pi_3} \cdots z_{\pi_n \pi_1}}. \quad (1.3)$$

Each color-ordered, field-theory amplitude $\mathcal{M}_n^{\text{FT}}$ is defined by a CHY formula, whose color ordering is given by a PT factor $\text{PT}(\rho)$; the nontrivial part is a half-integrand $\mathcal{I}_n^{\text{CHY}}$ that is obtained from the original $\mathcal{I}_n^{\text{string}}$ by IBP, which specifies the theory and external states:

$$\mathcal{M}_n^{\text{FT}}(\rho) := \int \frac{d^n z}{\text{vol SL}(2, \mathbb{C})} \underbrace{\prod_i \delta\left(\sum_{j \neq i} \frac{s_{ij}}{z_{ij}}\right)}_{:= d\mu_n^{\text{CHY}}} \text{PT}(\rho) \mathcal{I}_n^{\text{CHY}}(z). \quad (1.4)$$

Here the integrals are localized by the $n-3$ delta functions imposing scattering equations [13, 15].

Before moving on, let us pause and talk about an equivalent way of expressing string amplitudes as double copy. Note that the KLT double copy of Z integrals with $\mathcal{M}(\rho)$ only concerns the Parke-Taylor factor $\text{PT}(\rho)$ of the latter, and leaves $\mathcal{I}_n^{\text{CHY}}$ intact (which is independent of the ordering ρ). Thus it is natural to put the double-copy inside the CHY integral and write the string amplitude as a CHY formula:

$$\mathcal{M}_n^{\text{string}}(\rho) = \int d\mu_n^{\text{CHY}} \mathcal{Z}_{\rho}(z) \mathcal{I}_n^{\text{CHY}}(z), \quad (1.5)$$

where we have defined a *universal* CHY half-integrand $\mathcal{Z}_\rho(z) := Z_\rho \otimes \text{PT}(z)$ for any open-string amplitude.¹ This part is present regardless of type-I, bosonic or other possible theories, and the difference between these theories is represented by $\mathcal{I}_n^{\text{CHY}}$ only. For the closed string case, we simply replace the Z integral in the definition of \mathcal{Z} by single-valued projection of open-string amplitudes. While this rewriting has been known for a while, we emphasize that here the nontrivial, theory-dependent part in such CHY formulas, $\mathcal{I}_n^{\text{CHY}}$, can be obtained from the string correlator $\mathcal{I}_n^{\text{string}}$ as follows.

Using the technique developed in [1], one in fact obtains an *equivalence class* of CHY half-integrands from the string correlator $\mathcal{I}_n^{\text{string}}$ through the following two steps:

- First we algorithmically reduce the string correlator $\mathcal{I}_n^{\text{string}}$, via IBP relations, to an equivalent class of logarithmic functions \mathcal{I}_n , which can be used as a CHY half-integrand in Eq. (1.4).
- Next we use scattering equations (SE) to obtain equivalent half-integrands, $\mathcal{I}_n^{\text{CHY}}$, which are no longer logarithmic but usually take a more compact form and make some useful properties more manifest.

Logarithmic functions are defined to have only logarithmic singularities, *i.e.* simple poles, on boundaries of the moduli space of n -punctured Riemann spheres. Equivalently, it can be written as a linear combination of PT factors [57–59]. Note that we have an equivalence class of logarithmic functions $\mathcal{I}_n \stackrel{\text{IBP}}{\cong} \mathcal{I}_n^{\text{string}}$: any \mathcal{I}_n gives the same string integral as that of $\mathcal{I}_n^{\text{string}}$. Since they are also equivalence by SE, any \mathcal{I}_n gives the same $\mathcal{M}_n^{\text{FT}}$ as well. One can usually use SE to simplify \mathcal{I}_n greatly and obtain $\mathcal{I}_n^{\text{CHY}}$: while being non-logarithmic, usually it allows an all-multiplicity expression!

In this paper, we obtain two new results, corresponding to the two steps above, for the scalar-gluon correlators of compactified open bosonic strings, or equivalently the holomorphic part of heterotic strings. Recall that a general mixed string correlator for r gluons and $m+1$ scalar traces reads [40]

$$\mathcal{I}_n^{\text{string}}(z) = R(i_1, i_2, \dots, i_r) \prod_{t=1}^{m+1} \text{PT}(W_t), \quad (1.6)$$

where the PT factors follow the definition in Eq. (1.3). The $R(i_1, i_2, \dots, i_r)$ correlator, containing the gluon polarization vectors, is given by a cycle expansion:

$$R(i_1, i_2, \dots, i_r) = \sum_{(I)(J)\cdots(K) \in S_r} \mathcal{R}_{(I)} \mathcal{R}_{(J)} \cdots \mathcal{R}_{(K)}. \quad (1.7)$$

Here, we sum over all the permutations of $\{i_1, \dots, i_r\}$, and write them as products of cycles

¹Such half-integrands have been studied earlier: it was called string-deformed Parke-Taylor factor in [56] and also implicitly studied for the higher-energy limit in [15].

$(I), (J), \dots, (K)$. For length-one and two cycles, we have

$$\mathcal{R}_{(i)} = C_i := \sum_{j \neq i} C_{i,j} = \sum_{j \neq i} \frac{\epsilon_i \cdot k_j}{z_{ij}}, \quad \mathcal{R}_{(ij)} = \frac{\epsilon_i \cdot \epsilon_j}{\alpha' z_{ij}^2}, \quad (1.8)$$

while $\mathcal{R}_{(I)} = 0$ for longer cycles. In C_i , the summation is over all the particle labels that are different from i . We note that Eq. (1.6) also appears in the heterotic string correlator $\mathcal{I}_n^{\text{string}}(z)\mathcal{K}(\bar{z})$ for r gravitons and $m+1$ gluons traces, where $\mathcal{K}(\bar{z})$ is an antiholomorphic type-I superstring correlator.

For step 1, we will propose a systematic method for performing IBP to reduce any multitrace mixed correlator to logarithmic functions. This is based on a new recursive expansion we discover for such string correlators, as we show in (3.14) for pure-scalar case and (3.23) for mixed case. The correlator can be expanded into two parts: the first part is a logarithmic function involving a set of *labeled trees* \mathcal{T} , which has appeared in previous studies of Yang-Mills-scalar CHY integrand [60, 61], and corresponds to \mathcal{I}_n from IBP reducing (compactified) superstring correlators; the second part contains terms with the total number of traces and gluons decreased. By iterating the expansion, any multitrace mixed correlator can be reduced to a logarithmic function. In our arXiv submission, we provide an ancillary Mathematica file which implements the expansion and does IBP reduction for any number of traces and gluons.

For step 2, we further use SE to rewrite the logarithmic function \mathcal{I}_n to a closed-form CHY half-integrand, $\mathcal{I}_n^{\text{CHY}}$ for generic multitrace $(DF)^2 + \text{YM} + \phi^3$ amplitudes. In [1] we have presented a simple formula for all single-trace results, which we checked to high multiplicities but we had not found a proof then. In this paper we extend our construction to any number of traces, and write $\mathcal{I}_n^{\text{CHY}}$ in a relatively simple form once a basic operation called *fusion* is defined. The result is expressed as a sum over *total partitions*, and it turns out that even for the single-trace case, we get an equivalent but distinct formula than that in [1]. The outline of the paper is as follows.

We first present the general CHY-half integrand, $\mathcal{I}_n^{\text{CHY}}$ in section 2. As a byproduct, in the $\alpha' \rightarrow 0$ limit, our formula also gives a new formula for all multitrace Yang-Mills-scalar amplitudes, which are different from the original one in [21]. In section 3, we present the recursive expansion of the string correlator into a manifestly logarithmic part (which will be reviewed in Appendix A), and additional terms for which we can use the expansion again. We derive the expansion for pure-scalar case and outline the derivation for mixed cases, with some details left in the Appendix B and C. In section 4, we will illustrate how to use the result from IBP to obtain the CHY formula summarized in section 2. In section 5, we present an important check of the result which is factorization on massless poles.

2 The formula for all multitrace amplitudes

In this section, we present the complete CHY half-integrands for multitrace amplitudes in $(DF)^2 + \text{YM} + \phi^3$ theory. The allowed external particles are massless gluons A_μ^a and bi-adjoint scalars $\phi^{a\tilde{a}}$, where a is in the adjoint of a gauge group $U(N)$ and \tilde{a} is in the

adjoint of a global symmetry group $U(\tilde{N})$. We consider tree-level amplitudes with a fixed color ordering in a , represented by $\text{PT}(\rho)$ in Eq. (1.4). By “multitrace”, we mean that the global adjoint indices \tilde{a} of the scalars have the structure of $m+1$ traces.

In the limit $\alpha' \rightarrow 0$, the half-integrand reduces to the usual Yang-Mills-scalar one, where the gluons $\{i_1, i_2, \dots, i_r\}$ are packed into a reduced Pfaffian $\text{Pf}'(\Pi)$, and the scalar multitrace structure is described by $m+1$ PT factors $\text{PT}(W_i)$, see section 3 of [21]. Alternatively, we can single out one trace, say $\text{PT}(W_{m+1})$, but treat the rest on the same footing as gluons. As we will see, the result is given by “fusions” of all possible partitions of the set $\{W_1, \dots, W_m, i_1, \dots, i_r\}$. At finite α' , we need to introduce some α' -deformation of the fusion and consider a generalization of the partition: it turns out that we need the so-called *total partitions*.

2.1 Fusion of traces and gluons

The first operation we introduce is the (*weighted*) *fusion* between two traces W_1 and W_2 represented by the PT factors $\text{PT}(W_1)$ and $\text{PT}(W_2)$:

$$\langle W_1, W_2 \rangle := \frac{1}{2} \sum_{\substack{a_1, b_1 \in W_1 \\ a_2, b_2 \in W_2}} s_{b_1 a_2} s_{b_2 a_1} \frac{z_{b_1 a_1} z_{b_2 a_2}}{z_{b_1 a_2} z_{b_2 a_1}} \text{PT}(W_1) \text{PT}(W_2), \quad (2.1)$$

The cross ratio on the right hand side glues $\text{PT}(W_1)$ and $\text{PT}(W_2)$ into a single PT factor:

$$\frac{z_{b_1 a_1} z_{b_2 a_2}}{z_{b_1 a_2} z_{b_2 a_1}} \text{PT}(W_1) \text{PT}(W_2) = (-1)^{|B_1|+|B_2|} \sum_{\substack{\sigma_1 \in A_1 \sqcup B_1^T \\ \sigma_2 \in A_2 \sqcup B_2^T}} \text{PT}(a_1, \sigma_1, b_1, a_2, \sigma_2, b_2). \quad (2.2)$$

The sets A_i and B_i are determined as follows for $i = 1, 2$. For each choice of $a_i, b_i \in W_i$, we write $\text{PT}(W_i) = \text{PT}(a_i, A_i, b_i, B_i)$, using the cyclicity of PT factors. We then break W_i into words (a_i, σ_i, b_i) , sum over all the σ_i 's in $A_i \sqcup B_i^T$ (the shuffle of A_i and the reverse of B_i), and then glue the two words (a_i, σ_i, b_i) into $\text{PT}(a_1, \sigma_1, b_1, a_2, \sigma_2, b_2)$. This process is shown schematically in figure 1. The sum over shuffle can be represented by a wavy line:

$$\overset{a}{\text{wavy line}} \overset{b}{\bullet} := \text{PT}(W) z_{ba} = \text{PT}(a, A, b, B) z_{ba} = \sum_{\sigma \in A \sqcup B^T} \frac{(-1)^{|B|}}{z_{a\sigma_1} z_{\sigma_1 \sigma_2} \cdots z_{\sigma_{|\sigma|} b}}, \quad (2.3)$$

where b is chosen as the end connected to an edge pointing away from W (see figure 1). The fusion $\langle W_1, W_2 \rangle$ merges two color traces into a single trace since it is a linear combination of $\text{PT}(\rho)$ with $\rho \in \text{perm}(W_1 \cup W_2)$. As a concrete example, we consider the fusion of $\text{PT}(12)$ and $\text{PT}(345)$:

$$\begin{aligned} \langle (12), (345) \rangle &= \text{PT}(12) \text{PT}(345) \left[s_{23} s_{51} \frac{z_{21} z_{53}}{z_{23} z_{51}} + s_{23} s_{41} \frac{z_{21} z_{43}}{z_{23} z_{41}} + s_{24} s_{51} \frac{z_{21} z_{54}}{z_{24} z_{51}} + (1 \leftrightarrow 2) \right] \\ &= s_{23} s_{51} \text{PT}(12345) - s_{23} s_{41} \text{PT}(12354) + s_{24} s_{51} \text{PT}(12435) + (1 \leftrightarrow 2). \quad (2.4) \end{aligned}$$

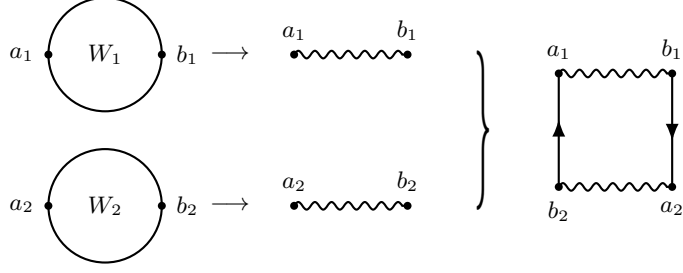


Figure 1: Gluing two cycles W_1 and W_2 into a single one. A directed edge from node i to j represents a factor of $\frac{1}{z_{ij}}$ while a wavy edge is defined by Eq. (2.3).

The generalization to the fusion of r cycles is straightforward (with $a_{r+1} := a_1$):

$$\begin{aligned}
\langle W_1, W_2, \dots, W_r \rangle &:= \frac{1}{2} \left[\prod_{i=1}^r \sum_{a_i, b_i \in W_i} \frac{s_{b_i a_{i+1}} z_{b_i a_i}}{z_{b_i a_{i+1}}} \text{PT}(W_i) \right] \\
&= \frac{1}{2} \left[\prod_{i=1}^r \sum_{a_i, b_i \in W_i} s_{b_i a_{i+1}} \right] \begin{array}{c} a_1 \quad b_1 \quad a_2 \quad b_2 \\ \text{---} \text{---} \text{---} \text{---} \\ \uparrow \quad \quad \downarrow \\ \text{---} \text{---} \text{---} \text{---} \\ b_r \quad a_r \quad \dots \end{array} . \quad (2.5)
\end{aligned}$$

Each wavy line represents a summation over $A_i \sqcup B_i^T$ with i inferred by the ends points. This also defines the fusion for $r = 1$ case: (2.5) reduces to $\langle W \rangle = s_W \text{PT}(W)$, where $s_W := \sum_{i < j \in W} s_{ij}$. The definition (2.5) is clearly cyclic, and the factor $\frac{1}{2}$ cancels the double counting due to the reflection symmetry $\langle W_1, W_2, \dots, W_r \rangle = \langle W_r, \dots, W_2, W_1 \rangle$.

Next, we include gluons into the fusion of traces. We find it convenient to view them as length-one words carrying polarizations, and consider them on the same footing as traces. For a single gluon, we define the fusion to be

$$\langle i \rangle := -C_i = - \sum_{j \neq i} \frac{\epsilon_i \cdot k_j}{z_{ij}} . \quad (2.6)$$

The fusion between two or more gluons $\langle i_1, i_2, \dots, i_r \rangle$ is given as follows. For each length-one word i , we assign a field strength $f_i^{\mu\nu} = k_i^\mu \epsilon_i^\nu - k_i^\nu \epsilon_i^\mu$. We then glue all the i 's in order and Lorentz contract the field strengths in the same way:

$$\langle i_1, i_2, \dots, i_r \rangle := \frac{1}{2} \text{tr}(f_1 f_2 \cdots f_r) \text{PT}(i_1, i_2, \dots, i_r) := \frac{1}{2} \text{tr}(f_1 f_2 \cdots f_r) \begin{array}{c} i_1 \quad i_2 \\ \text{---} \text{---} \\ \uparrow \quad \quad \downarrow \\ \text{---} \text{---} \\ \dots \\ \downarrow \quad \quad \uparrow \\ i_r \end{array} . \quad (2.7)$$

These objects have already appeared in the cycle expansion of gluon CHY integrands [62].

If a trace W is involved in the fusion with a gluon i , we break W as in figure 1, glue the length-one word i to the end points, and then contract the momenta with the field

strength f_i . Namely, we have

$$\begin{aligned} \langle W, i \rangle &:= \frac{\alpha'}{2} \sum_{a,b \in W} (k_b \cdot f_i \cdot k_a) \frac{z_{ba}}{z_{bi} z_{ia}} \text{PT}(W) \\ &= \frac{\alpha'}{2} \sum_{a,b \in W} (-1)^{|B|} (k_b \cdot f_i \cdot k_a) \sum_{\sigma \in A \sqcup B^T} \text{PT}(a, \sigma, b, i) \end{aligned} \quad (2.8)$$

The generalization to arbitrary number of traces and gluons is straightforward:

$$\langle W_1, \overbrace{i_1, \dots, i_s}^{\mathbf{G}_1}, \dots, W_r, \overbrace{j_1, \dots, j_\ell}^{\mathbf{G}_r} \rangle = \frac{\alpha'^r}{2} \left[\prod_{i=1}^r \sum_{a_i, b_i \in W_i} \frac{(k_{b_i} \cdot f_{\mathbf{G}_i} \cdot k_{a_{i+1}}) z_{b_i a_i}}{z_{b_i, \mathbf{G}_i, a_{i+1}}} \text{PT}(W_i) \right], \quad (2.9)$$

where, for example, $z_{b_1, \mathbf{G}_1, a_2} := z_{b_1 i_1} z_{i_1 i_2} \cdots z_{i_s a_2}$ and $(f_{\mathbf{G}_1})_{\mu\nu} := (f_{i_1} f_{i_2} \cdots f_{i_s})_{\mu\nu}$, etc. Similarly, we also have a diagrammatic representation for the fusion. We illustrate with two traces and two sets of gluons in between:

$$\langle W_1, \overbrace{i_1, \dots, i_s}^{\mathbf{G}_1}, W_2, \overbrace{j_1, \dots, j_\ell}^{\mathbf{G}_2} \rangle = \frac{\alpha'^2}{2} \sum_{\substack{a_1, b_1 \in W_1 \\ a_2, b_2 \in W_2}} (k_{b_1} \cdot f_{\mathbf{G}_1} \cdot k_{a_2}) (k_{b_2} \cdot f_{\mathbf{G}_2} \cdot k_{a_1}) \quad \text{Diagram}, \quad (2.10)$$

where the diagram stands for the sum

$$\text{Diagram} = (-1)^{|B_1| + |B_2|} \sum_{\substack{\sigma_1 \in A_1 \sqcup B_1^T \\ \sigma_2 \in A_2 \sqcup B_2^T}} \text{PT}(a_1, \sigma_1, b_1, i_1, \dots, i_r, a_2, \sigma_2, b_2, j_1, \dots, j_\ell). \quad (2.11)$$

Loosely speaking, the gluons participating in fusions are turned into components of a color trace, while the polarization information appears as kinematic coefficients of the color traces. To illustrate our result, let us write explicitly some low-multiplicity examples:

$$\begin{aligned} \langle (12), 3 \rangle &= \alpha' (k_2 \cdot f_3 \cdot k_1) \text{PT}(123), \\ \langle (12), 3, 4 \rangle &= \frac{\alpha'}{2} [(k_2 \cdot f_3 f_4 \cdot k_1) \text{PT}(1234) + (k_1 \cdot f_3 f_4 \cdot k_2) \text{PT}(2134)], \\ \langle (123), 4 \rangle &= \alpha' [(k_3 \cdot f_4 \cdot k_1) \text{PT}(1234) - (k_2 \cdot f_4 \cdot k_1) \text{PT}(1324) - (k_3 \cdot f_4 \cdot k_2) \text{PT}(2134)], \\ \langle (12), (34), 5 \rangle &= \frac{\alpha'}{2} [s_{23} (k_4 \cdot f_5 \cdot k_1) \text{PT}(12345) + s_{24} (k_3 \cdot f_5 \cdot k_1) \text{PT}(12435) + (1 \leftrightarrow 2)], \\ \langle (12), 5, (34), 6 \rangle &= \frac{\alpha'^2}{2} [(k_2 \cdot f_5 \cdot k_3) (k_4 \cdot f_6 \cdot k_1) \text{PT}(125346) + (k_2 \cdot f_5 \cdot k_4) (k_3 \cdot f_6 \cdot k_1) \text{PT}(125436) \\ &\quad + (1 \leftrightarrow 2)]. \end{aligned} \quad (2.12)$$

Finally, as a technical convenience, we require the fusion operation be multilinear on traces, namely, the following relation should hold:

$$\langle \dots, x\text{PT}(W_1) + y\text{PT}(W_2), \dots \rangle := x\langle \dots, \text{PT}(W_1), \dots \rangle + y\langle \dots, \text{PT}(W_2), \dots \rangle, \quad (2.13)$$

where x and y are independent of the worldsheet variables z_i . As a result, nested fusions like $\langle \langle W_1, i, W_2 \rangle, W_3, j \rangle$ are well-defined. We note that fusions are not associative. For example, one can check that $\langle W_1, \langle W_2, W_3 \rangle \rangle \neq \langle \langle W_1, W_2 \rangle, W_3 \rangle \neq \langle W_1, W_2, W_3 \rangle$ by an explicit calculation.

2.2 Partitions of set and symmetrized fusions

A *partition* of set A is a family of nonempty subsets $\{A_1, A_2, \dots, A_m\}$ of A that satisfies $\bigcup_{i=1}^m A_i = A$ and $A_i \cap A_j = \emptyset$ if $i \neq j$, where $1 \leq m \leq |A|$ is the order of the partition. We refer A_i as a *block* in the partition. For $A = \{a_1, a_2\}$ and $\{a_1, a_2, a_3\}$, the collections of all partitions, denoted as $\mathbb{P}[A]$, are

$$\begin{aligned} \mathbb{P}[\{a_1, a_2\}] &= \left\{ \{ \{a_1, a_2\} \}, \{a_1, a_2\} \right\}, \\ \mathbb{P}[\{a_1, a_2, a_3\}] &= \left\{ \{ \{a_1, a_2, a_3\} \}, \{a_1, a_2, a_3\}, \{ \{a_1, a_2\}, a_3 \}, \{ \{a_2, a_3\}, a_1 \}, \{ \{a_1, a_3\}, a_2 \} \right\}. \end{aligned}$$

To avoid cluttered notations, we omit the curly bracket on singleton blocks when confusion is unlikely.² For the partitions with $m \geq 2$, we can further partition each non-singleton A_i into two or more blocks, and continue the process until only singleton blocks remain. On the other hand, if we perform the same operation on the $m = 1$ partition $\{A\}$, we get the same result but with an overall curly bracket [63]. Together they form the family of *total partitions* of A , denoted as $\mathbb{T}[A]$.³ For example, we have

$$\begin{aligned} \mathbb{T}[\{a_1, a_2\}] &= \left\{ \{a_1, a_2\}, \{ \{a_1, a_2\} \} \right\}, \\ \mathbb{T}[\{a_1, a_2, a_3\}] &= \left\{ \{a_1, a_2, a_3\}, \{ \{a_1, a_2\}, a_3 \}, \{ \{a_2, a_3\}, a_1 \}, \{ \{a_1, a_3\}, a_2 \}, \right. \\ &\quad \left. \{ \{a_1, a_2, a_3\} \}, \{ \{ \{a_1, a_2\}, a_3 \} \}, \{ \{ \{a_2, a_3\}, a_1 \} \}, \{ \{ \{a_1, a_3\}, a_2 \} \} \right\}. \end{aligned}$$

For a total partition $\mathbf{A} = \{A_1, A_2, \dots, A_m\} \in \mathbb{T}[A]$, each block A_i may contain nested curly brackets. In contrary, for $\mathbf{A} \in \mathbb{P}[A]$, the A_i 's contain only singleton blocks. It is clear that by construction $\mathbb{P}[A]$ is always a subset of $\mathbb{T}[A]$.

We are interested in the case when the elements of A are a collection of traces and gluons. Now we define the *symmetrized fusion* with α' deformation that acts recursively on the block A_i of a total partition $\mathbf{A} \in \mathbb{T}[A]$:

²The total number of partitions for n elements is known as the Bell number (<https://oeis.org/A000110>), which equals 1, 2, 5, 15, 52 for $n = 1, 2, 3, 4, 5$, etc.

³The only exception is that $\mathbb{T}[\{a_1\}] := \{ \{a_1\} \}$. The number of total partitions under our definition equals the one in <https://oeis.org/A006351>.

- If $A_i = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r\}$ contains only singleton blocks, we define

$$\mathcal{S}_{\alpha'}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r) := \frac{1}{1 - s_{\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_r}} \sum_{\pi \in S_r / \mathbb{Z}_r} \langle \mathbf{a}_{\pi(1)}, \mathbf{a}_{\pi(2)}, \dots, \mathbf{a}_{\pi(r)} \rangle, \quad (2.14)$$

where \mathbf{a}_j can either be a trace W_j or a gluon j . For $r = 1$, we use instead

$$\mathcal{S}_{\alpha'}(W) := \langle W \rangle = s_W \text{PT}(W), \quad \mathcal{S}_{\alpha'}(i) := \langle i \rangle = -C_i. \quad (2.15)$$

- In the $\alpha' \rightarrow 0$ limit, we have

$$\mathcal{S}_0(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r) = \sum_{\pi \in S_r / \mathbb{Z}_r} \langle \mathbf{a}_{\pi(1)}, \mathbf{a}_{\pi(2)}, \dots, \mathbf{a}_{\pi(r)} \rangle, \quad (2.16)$$

but still $\mathcal{S}_0(W) = \langle W \rangle$ and $\mathcal{S}_0(i) = \langle i \rangle$ since they contribute to the leading α' order.

- If A_i contains nested curly brackets, say $A_i = \{A'_1, A'_2, \dots, A'_j, \mathbf{a}_{j+1}, \dots, \mathbf{a}_r\}$, we define

$$\mathcal{S}_{\alpha'}(A_i) := \mathcal{S}_{\alpha'}(\mathcal{S}_{\alpha'}(A'_1), \mathcal{S}_{\alpha'}(A'_2), \dots, \mathcal{S}_{\alpha'}(A'_j), \mathbf{a}_{j+1}, \dots, \mathbf{a}_r). \quad (2.17)$$

The symmetrization is an essential ingredient here since it restores the bosonic exchange symmetry of the gluons and color traces after the fusion.

Using the multilinearity (2.13) of the fusion, we can calculate generic $\mathcal{S}_{\alpha'}(A_i)$ from the inner-most level. We give two examples to demonstrate the construction:

$$\begin{aligned} \mathcal{S}_{\alpha'}(\{(456), 7\}) &= \mathcal{S}_{\alpha'}((456), 7) \\ &= \frac{\alpha' [(k_6 \cdot f_7 \cdot k_4) \text{PT}(4567) - (k_5 \cdot f_7 \cdot k_4) \text{PT}(4657) - (k_6 \cdot f_7 \cdot k_5) \text{PT}(5467)]}{1 - s_{4567}}, \end{aligned} \quad (2.18a)$$

$$\begin{aligned} \mathcal{S}_{\alpha'}(\{\{2, 3\}, 4\}) &= \mathcal{S}_{\alpha'}(\mathcal{S}_{\alpha'}(2, 3), 4) \\ &= \frac{\text{tr}(f_2 f_3)}{1 - s_{23}} \mathcal{S}_{\alpha'}((23), 4) = \frac{\alpha' \text{tr}(f_2 f_3) (k_3 \cdot f_4 \cdot k_2)}{(1 - s_{234})(1 - s_{23})} \text{PT}(234). \end{aligned} \quad (2.18b)$$

The symmetrized fusion is our basic building block for the multitrace CHY integrands, as we will show in the next subsection.

2.3 Half-integrands for multitrace amplitudes

Now we present the half-integrand for $m+1$ traces and r gluons in the $(DF)^2 + \text{YM} + \phi^3$ theory. It is given by an overall factor including $\text{PT}(W_{m+1})$ for the trace W_{m+1} and tachyon poles for traces $\{W_1, \dots, W_m\}$, times a sum over symmetrized fusions of all the

total partitions $\mathbf{A} = \{A_1, A_2, \dots, A_{|\mathbf{A}|}\}$ of $\{W_1, \dots, W_m, i_1, \dots, i_r\}$:

$$\mathcal{I}_n^{\text{CHY}}(W_1, \dots, W_{m+1}, i_1, \dots, i_r) = \frac{\text{PT}(W_{m+1})}{\prod_{i=1}^m (1 - s_{W_i})} \sum_{\mathbf{A} \in \mathbb{T}[W_1 \dots W_m, i_1 \dots i_r]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_{\alpha'}(A_j). \quad (2.19)$$

In particular, the pure-scalar and single-trace integrands are given by

$$\mathcal{I}_n^{\text{CHY}}(W_1, \dots, W_{m+1}) = \frac{\text{PT}(W_{m+1})}{\prod_{i=1}^m (1 - s_{W_i})} \sum_{\mathbf{A} \in \mathbb{T}[W_1 \dots W_m]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_{\alpha'}(A_j), \quad (2.20a)$$

$$\mathcal{I}_n^{\text{CHY}}(W, i_1, \dots, i_r) = \text{PT}(W) \sum_{\mathbf{A} \in \mathbb{T}[i_1 \dots i_r]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_{\alpha'}(A_j). \quad (2.20b)$$

The crossing symmetry among $\{W_1, \dots, W_m, i_1, \dots, i_r\}$ is manifest since $\mathcal{S}_{\alpha'}$ is completely symmetric in its arguments. On the other hand, different choices of W_{m+1} lead to equivalent integrand on the support of scattering equations. Some simple multitrace examples are given as follows:

$$\text{double-trace:} \quad \mathcal{I}_n^{\text{CHY}}(\sigma, \rho) = -\frac{\text{PT}(\rho) \mathcal{S}_{\alpha'}(\sigma)}{1 - s_\sigma} = -\frac{s_\sigma}{1 - s_\sigma} \text{PT}(\rho) \text{PT}(\sigma), \quad (2.21a)$$

$$\text{triple-trace:} \quad \mathcal{I}_n^{\text{CHY}}(\sigma, \tau, \rho) = \frac{\text{PT}(\rho) [\mathcal{S}_{\alpha'}(\sigma) \mathcal{S}_{\alpha'}(\tau) - \mathcal{S}_{\alpha'}(\sigma, \tau)]}{(1 - s_\sigma)(1 - s_\tau)}, \quad (2.21b)$$

$$\text{double-trace one-gluon:} \quad \mathcal{I}_n^{\text{CHY}}(\tau, \rho, q) = \frac{\text{PT}(\rho)}{1 - s_\tau} [\mathcal{S}_{\alpha'}(\tau) \mathcal{S}_{\alpha'}(q) - \mathcal{S}_{\alpha'}(\tau, q)], \quad (2.21c)$$

where σ , ρ and τ are traces and q is a single gluon. The symmetrized fusions involved can be written explicitly as

$$\begin{aligned} \mathcal{S}_{\alpha'}(\sigma) \mathcal{S}_{\alpha'}(\tau) &= s_\sigma s_\tau \text{PT}(\sigma) \text{PT}(\tau), & \mathcal{S}_{\alpha'}(\tau) \mathcal{S}_{\alpha'}(q) &= -s_\tau \text{PT}(\tau) C_q, \\ \mathcal{S}_{\alpha'}(\sigma, \tau) &= \frac{1}{2(1 - s_\rho)} \sum_{\substack{\sigma_1, \sigma_2 \in \sigma \\ \tau_1, \tau_2 \in \tau}} s_{\tau_2 \sigma_1} s_{\sigma_2 \tau_1} \frac{z_{\sigma_2 \sigma_1} z_{\tau_2 \tau_1}}{z_{\tau_2 \sigma_1} z_{\sigma_2 \tau_1}} \text{PT}(\sigma) \text{PT}(\tau), \\ \mathcal{S}_{\alpha'}(\tau, q) &= \frac{\alpha'}{2(1 - s_\rho)} \sum_{\tau_1, \tau_2 \in \tau} (k_{\tau_2} \cdot f_q \cdot k_{\tau_1}) \frac{z_{\tau_2 \tau_1}}{z_{\tau_2 q} z_{q \tau_1}} \text{PT}(\tau). \end{aligned} \quad (2.22)$$

Interestingly, we can rewrite the single-trace integrand (2.20b) into a more familiar cycle expansion form:

$$\sum_{\mathbf{A} \in \mathbb{T}[i_1 \dots i_r]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_{\alpha'}(A_j) = (-1)^r \sum_{(I)(J) \dots (K) \in S_r} \Psi_{(I)} \Psi_{(J)} \dots \Psi_{(K)}. \quad (2.23)$$

This identity holds at the algebraic level, which can be easily checked numerically. However, the inductive proof is lengthy and we omit it here. For length-one and two cycles, the cycle

factor Ψ is given by

$$\Psi_{(i)} = C_i, \quad \Psi_{(ij)} = -T_{ij} \text{PT}(ij) := -\frac{\text{tr}(f_i f_j)}{2(1-s_{ij})} \text{PT}(ij). \quad (2.24)$$

For cycles with $|I| \geq 3$, we have

$$\Psi_{(I)} = -\frac{T_I \text{PT}(I)}{2} := -\frac{1}{2} \left[\frac{1}{(1-s_I)} \sum_{\text{CP}} \text{tr}(F_{I_1} F_{I_2} \dots F_{I_p}) \right] \text{PT}(I). \quad (2.25)$$

The definition of T_{ij} and T_I can be inferred from the above two equations. Here, the summation is over all the cyclic partitions (CP) $\{I_1, I_2, \dots, I_p\}$ of I with $p \geq 2$. Each block I_ℓ of a cyclic partition must conform to the cyclic order determined by I .⁴ For length-one blocks, $F_i^{\mu\nu}$ is just the field strength $f_i^{\mu\nu}$; for longer ones, it is recursively defined as

$$F_{i_1 i_2 \dots i_t}^{\mu\nu} = \alpha' k_{i_1}^\mu T_{i_1 i_2 [i_3 [\dots [i_{t-1} i_t] \dots]]} k_{i_t}^\nu \quad (t \geq 2), \quad (2.26)$$

where the bracket $[ij]$ stands for an antisymmetrization. For example,

$$\begin{aligned} T_{i_1 i_2 [i_3 i_4]} &= T_{i_1 i_2 i_3 i_4} - T_{i_1 i_2 i_4 i_3}, \\ T_{i_1 i_2 [i_3 [i_4 i_5]]} &= T_{i_1 i_2 i_3 i_4 i_5} - T_{i_1 i_2 i_3 i_5 i_4} - T_{i_1 i_2 i_4 i_5 i_3} + T_{i_1 i_2 i_5 i_4 i_3}. \end{aligned} \quad (2.27)$$

The cycle factor (2.25) is of a different form compared with the one defined in [1].⁵ They are of course algebraically equivalent, as one can check explicitly.

Finally, we study the $\alpha' \rightarrow 0$ limit of the half-integrand (2.19). From the definition of $S_{\alpha'}$, one can show that, for example, $\mathcal{S}_{\alpha'}(\{\mathbf{a}_1, \mathbf{a}_2\}, \mathbf{a}_3) = \mathcal{S}_{\alpha'}(S_{\alpha'}(\mathbf{a}_1, \mathbf{a}_2), \mathbf{a}_3)$ gives higher α' order contribution than $\mathcal{S}_{\alpha'}(\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}) = \mathcal{S}_{\alpha'}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. As a result, the leading α' order is contributed solely by $\mathbf{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_{|\mathbf{A}|}\} \in \mathbb{P}[W_1, \dots, W_m, i_1, \dots, i_r]$, where the \mathbf{A}_i 's do not have nested curly brackets:

$$\mathcal{I}_n^{\text{CHY}}(W_1, \dots, W_{m+1}, i_1, \dots, i_r)_{\alpha' \rightarrow 0} = \text{PT}(W_{m+1}) \sum_{\mathbf{A} \in \mathbb{P}[W_1 \dots W_m, i_1 \dots i_r]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_0(\mathbf{A}_j). \quad (2.28)$$

We have also removed all tachyon poles such that $\mathcal{S}_{\alpha'}$ reduces to \mathcal{S}_0 . For the single-trace case, one can easily show that the summation in Eq. (2.28) indeed gives a Pfaffian in the leading α' order. For generic cases, we can establish the equivalence between Eq. (2.28) and the Yang-Mills-scalar integrand given in the squeezed form [21] by using Eq (2.5) and (2.9). After factorizing out the overall product $\prod_{i=1}^{m+1} \text{PT}(W_i)$, one can recognize that Eq. (2.28) agrees exactly with the reduced Pfaffian under the gauge choice of deleting the two rows and columns associated to trace W_{m+1} .

⁴If $|I| = n$, the number of such cyclic partitions is the Eulerian number $A(n, 1) = 2^n - n - 1$.

⁵Comparing with the definition given in [1], we note that the requirement $|I_\ell| \geq 2$ is relaxed in Eq. (2.25).

3 Recursive expansion of string correlator

In this section, we present a recursive expansion for the open-bosonic string integrand (1.6). To achieve this, one needs to perform the IBP reduction in a well-controlled manner. We first show how this can be done for pure-scalar cases,

$$\mathcal{I}_n^{\text{string}} = \text{PT}(W_1) \text{PT}(W_2) \cdots \text{PT}(W_{m+1}). \quad (3.1)$$

Our goal is to reduce this string integrand, via IBP relations, into a combination of logarithmic functions and the string integrands with number of traces decreased by the fusions defined in the previous section. The logarithmic function naturally takes the form of labeled trees which will be reviewed in Appendix A. We can then use recursively the fewer-trace results and land on a logarithmic form integrand eventually.

As the starting point of our derivation, we always take a puncture in W_{m+1} to infinity. Under this gauge, we can break another subcycle W at a chosen puncture $z_a \in W$ by the following IBP relation [40]:

$$\text{PT}(W)(\cdots) \stackrel{\text{IBP}}{\cong} \frac{1}{1-s_W} \sum_{j \notin W} \text{PT}(W) \underbrace{\sum_{b \in W} \frac{s_{bj} z_{ba}}{z_{bj}}}_{C_{a,j}^W}(\cdots), \quad (3.2)$$

where (\cdots) does not involve any punctures in W except for z_a . It is convenient to represent $C_{a,j}^W$ by the following diagram

$$C_{a,j}^W = \sum_{b \in W} \begin{array}{c} a \\ \text{---} \text{wavy line} \text{---} b \text{---} \text{---} j \\ \text{---} \text{---} \end{array} s_{bj}, \quad (3.3)$$

where each term in the summation can be viewed as a chain (dressed with an additional factor s_{bj}). This diagrammatic representation is consistent with the one shown in figure 1. If the (\cdots) contains no subcycles, for example, the double-trace case,

$$\text{PT}(W_1)\text{PT}(W_2) \stackrel{\text{IBP}}{\cong} \frac{1}{1-s_{W_1}} \sum_{j_2 \in W_2} C_{a_1, j_2}^{W_1} \text{PT}(W_2) := \frac{\text{PT}(W_2)}{1-s_{W_1}} \mathcal{T}_{W_2}(W_1), \quad (3.4)$$

the result is already logarithmic and our IBP reduction finishes.

However, for triple-trace and beyond, the above no-subcycle condition no longer holds after breaking one subcycle, say W_1 . Thus further IBP reduction is necessary. We first show by some examples on how to proceed in this situation. We then provide a systematic solution that leads to a recursive expansion for generic pure-scalar integrands. Finally, we write down the expansion for the case with gluons, which is an analog of the pure-scalar case, and leave the derivation to Appendix C.

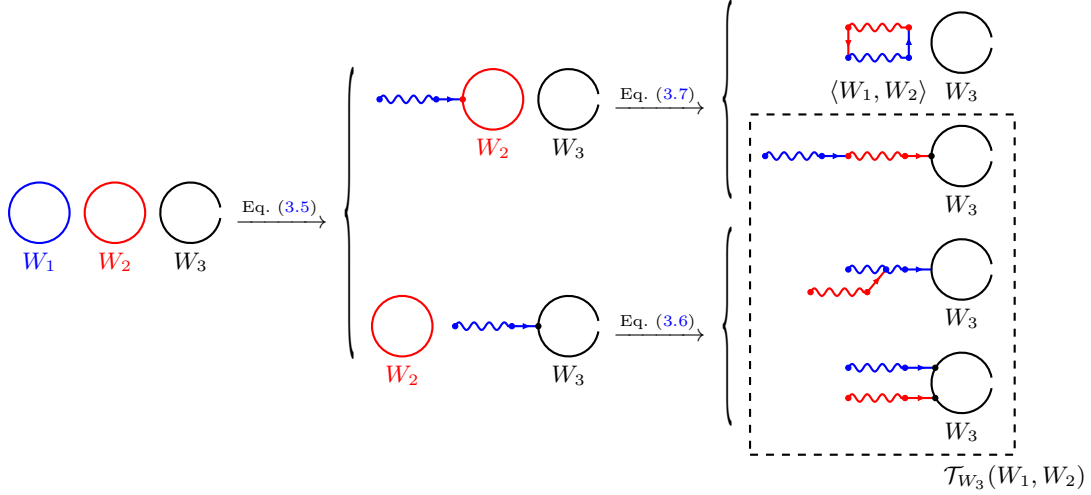


Figure 2: The IBP reduction of the triple-trace string integrand $\text{PT}(W_1)\text{PT}(W_2)\text{PT}(W_3)$. The subcycle $\text{PT}(W_3)$ is broken since we set one of its punctures to infinity.

is given as a linear combination of logarithmic functions and double-trace integrands.

The recursive nature of this derivation is more obvious when we carry on to four traces. After setting a puncture in W_4 to infinity, we break the subcycle W_1 by Eq. (3.2). Now the chain C^{W_1} can either connect to another subcycle (W_2 or W_3), forming a normal tadpole, or to the root W_4 . Next, we break all the normal tadpoles by Eq. (3.2) at the attach point of the tail, including those generated in the process. At the end, the chain C^{W_1} is either connected to the root (maybe through another chain), or appears in an induced tadpole. We then repeat these steps for subcycle W_2 (if exists), followed by W_3 (if exists). This prescription introduces a reference order $W_1 \prec W_2 \prec W_3$ for the subcycles as the priority rank of being broken by the IBP relation (3.2). The result of the above IBP reduction is

$$\prod_{i=1}^4 \text{PT}(W_i) \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W_4)}{\prod_{i=1}^3 (1 - s_{W_i})} \left[\alpha'^3 \mathcal{T}_{W_4}(W_1, W_2, W_3) + \langle W_1, W_2, W_3 \rangle + \langle W_1, W_3, W_2 \rangle + ((1 - s_{W_1}) \text{PT}(W_1) \langle W_2, W_3 \rangle + \text{cyclic}) \right], \quad (3.11)$$

where $\mathcal{T}_{W_4}(W_1, W_2, W_3)$ is a set of labeled trees rooted on W_4 and evaluated under the reference order $W_1 \prec W_2 \prec W_3$. Again we refer the readers to Appendix A for the construction of these labeled trees. Different reference orders will give different \mathcal{T}_{W_4} 's that are equivalent up to an IBP relation. The fusions in the first line of Eq. (3.11) come from induced tadpoles after using the algebraic identity

$$\sum_{j_1 \in W_1} \sum_{j_2 \in W_2} \sum_{j_3 \in W_3} \left(C_{a_1, j_2}^{W_1} C_{j_2, j_3}^{W_2} C_{j_3, j_1}^{W_3} + (2 \leftrightarrow 3) \right) = \langle W_1, W_2, W_3 \rangle + \langle W_1, W_3, W_2 \rangle, \quad (3.12)$$

while the fusions in the second line of Eq. (3.11) are obtained by first using Eq. (3.8) and

then the inversion of the relation (3.2), for example:

$$\sum_{j \notin W_1} C_{a_1, j}^{W_1} \langle W_2, W_3 \rangle \text{PT}(W_4) \stackrel{\text{IBP}}{\cong} (1 - s_{W_1}) \text{PT}(W_1) \langle W_2, W_3 \rangle \text{PT}(W_4). \quad (3.13)$$

Although the above manipulation seems to be a move in the opposite direction, the benefit is that now we can use directly the double- and triple-trace results in Eq. (3.11).

3.2 General pure-scalar cases

Now we provide the IBP reduction algorithm for generic $(m + 1)$ -trace string integrands, where we gauge fix one of the punctures in W_{m+1} to infinity. We first pick an arbitrary reference order, say $\mathbf{R} = W_1 \prec W_2 \prec \dots \prec W_m$, as the priority rank of being broken by the IBP relation (3.2). For each term in the integrand, we carry out the following algorithm:

- (t1) Break the first trace in the reference order (here W_1) by Eq. (3.2), which turns the trace W_1 into a chain C^{W_1} that is attached to another subcycle or the root W_{m+1} .
- (t2) If the chain C^{W_1} appears in the tail of a normal tadpole, break that subcycle using Eq. (3.2) at the attach point of the tail. Repeat this step until C^{W_1} is connected to the root W_{m+1} (maybe through another chain) or appears in an induced tadpole.
- (t3) Repeat step (t1) and (t2) for the next subcycle W_i in the reference order that remains in the original $\text{PT}(W_i)$ configuration.

The algorithm leads to a remarkable recursive expansion for the string integrand:

$$\prod_{i=1}^{m+1} \text{PT}(W_i) \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W_{m+1})}{\prod_{i=1}^m (1 - s_{W_i})} \left[\alpha'^m \mathcal{T}_{W_{m+1}}(\mathbf{R}) - \sum_{\substack{\mathbf{A} \in \mathbb{P}[W_1 \dots W_m] \\ |\mathbf{A}| < m}} (-1)^{|\mathbf{A}|} \mathcal{J}[\mathbf{A}] \right], \quad (3.14)$$

where $\mathcal{T}_{W_{m+1}}(\mathbf{R})$ consists of labeled trees only, and thus logarithmic. It is actually the logarithmic form CHY integrand for the pure-scalar sector of Yang-Mills scalar amplitudes. The explicit form depends on the reference order \mathbf{R} (for example, $W_1 \prec W_2 \prec \dots \prec W_m$), and we defer the details to Appendix A. The second term of Eq. (3.14) is independent of the reference order, in which the summation is over all the partitions \mathbf{A} of $\{W_1, W_2, \dots, W_m\}$ whose number of blocks is less than m . Suppose \mathbf{A} has s singleton blocks and the rest non-singletons, namely, $\mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_s, \mathbf{A}_{s+1}, \dots, \mathbf{A}_{|\mathbf{A}|}\}$, we can write \mathcal{J} as

$$\mathcal{J}[\mathbf{A}] = (-1)^s \left[\prod_{j=1}^s (1 - s_{\mathbf{a}_j}) \text{PT}(\mathbf{a}_j) \right] \left[\prod_{j=s+1}^{|\mathbf{A}|} \mathcal{S}_0(\mathbf{A}_j) \right]. \quad (3.15)$$

They come from the induced tadpoles generated in step (t2) after we use some algebraic identities like

$$\sum_{j_2 \in W_2} C_{a_1, j_2}^{W_1} \sum_{j_3 \in W_3} C_{j_2, j_3}^{W_2} \dots \sum_{j_r \in W_r} C_{j_{r-1}, j_r}^{W_r} + \text{perm}(2, 3, \dots, r) = \mathcal{S}_0(W_1, W_2, \dots, W_r). \quad (3.16)$$

In addition, we need to use the inverse of Eq. (3.2) to obtain the $(1 - s_{a_j}) \text{PT}(\mathbf{a}_j)$ factor.

We note that each singleton block \mathbf{a}_i contributes factor proportional to the original PT factor $\text{PT}(\mathbf{a}_i)$. The condition $|\mathbf{A}| < m$ in Eq. (3.14) guarantees that there exists at least one non-singleton block $\mathbf{A}_j = \{W_{j_1}, W_{j_2}, \dots\}$. Each $\mathcal{S}_0(\mathbf{A}_j)$ then merges the traces in the non-singleton block \mathbf{A}_j into a single trace according to Eq. (2.16). In factor, $\mathcal{S}_0(\mathbf{A}_j)$ is a linear combination of $\text{PT}(\rho)$ with ρ belong to a subset of $\text{perm}(\mathbf{A}_j) = \text{perm}(W_{j_1} \cup W_{j_2} \cup \dots)$. The symmetrization defined in \mathcal{S}_0 takes care of the bosonic exchange symmetry between the original traces.

Therefore, $\mathcal{J}[\mathbf{A}]$ is a product of PT factors taking value in each block of \mathbf{A} respectively, the physical meaning of $\mathcal{J}[\mathbf{A}]$ is then clear: *it is a linear combination of the string integrands with less number of traces, cf. the left hand side of Eq. (3.14)*. It is non-logarithmic because of the existence of subcycles. Nevertheless, we can recursively use Eq. (3.14) to eventually obtain a logarithmic integrand. On the other hand, further IBP reduction on $\mathcal{J}[\mathbf{A}]$ only leads to contributions to higher order of α' . Thus in the limit $\alpha' \rightarrow 0$, we have

$$\prod_{i=1}^{m+1} \text{PT}(W_i) \stackrel{\text{IBP}}{\cong} \alpha'^m \text{PT}(W_{m+1}) \mathcal{T}_{W_{m+1}}(\mathbf{R}) + \mathcal{O}(\alpha'^{m+1}), \quad (3.17)$$

where the first term is the result if we would have started with a type-I superstring correlator and obtained the multitrace structure through a compactification.

Our algorithm can reduce any multitrace correlator to a logarithmic function. In the arXiv submission of this paper, we implement the recursive expansion (3.14) in the ancillary Mathematica notebook `IBP.nb`. Given a reference order, the result can either be exported as a linear combination of labeled trees or further expanded in terms of Parke-Taylor factors in the DDM basis [64]. Run on a laptop, our algorithm can process eleven points with five traces in a few minutes.

3.3 Inserting one gluon

The single-gluon string integrand $C_i \prod_{j=1}^{m+1} \text{PT}(W_j)$ is only slightly more general than the pure-scalar one: each term in the integrand contains exactly one tadpole whose tail is a single gluon. We start with breaking the tadpole at the gluon attach point by Eq. (3.2), and then follow the same prescription as the pure-scalar case. In other words, the subcycle connected with the gluon is always prioritized. This essentially means that we choose the reference order $\mathbf{R} = i \prec W_1 \prec \dots \prec W_m$. The result can again be written as

$$C_i \prod_{i=1}^{m+1} \text{PT}(W_i) \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W_{m+1})}{\prod_{i=1}^m (1 - s_{W_i})} \left[\alpha'^m \mathcal{T}_{W_{m+1}}(i, W_1, \dots, W_m) - \sum_{\substack{\mathbf{A} \in \mathbb{P}[i, W_1 \dots W_m] \\ |\mathbf{A}| < m+1}} (-1)^{|\mathbf{A}|} \mathcal{J}[\mathbf{A}] \right], \quad (3.18)$$

where the definition of $\mathcal{J}[\mathbf{A}]$ is extended to include one single gluon. In particular, if there is a gluon singleton block, namely, $\mathbf{A} = \{i, \mathbf{a}_2, \dots, \mathbf{a}_s, \mathbf{A}_{s+1}, \dots, \mathbf{A}_{|\mathbf{A}|}\}$,

$$\mathcal{J}[\mathbf{A}] = (-1)^s C_i \left[\prod_{j=2}^s (1 - s_{\mathbf{a}_j}) \text{PT}(\mathbf{a}_j) \right] \left[\prod_{j=s+1}^{|\mathbf{A}|} \mathcal{S}_0(\mathbf{A}_j) \right]. \quad (3.19)$$

Otherwise, the definition is the same as Eq. (3.15) but with \mathcal{S}_0 involving fusion between the gluon and traces, which is obtained through algebraic identities like

$$\sum_{j_1 \in W_1} C_{ij_1} \sum_{j_2 \in W_2} C_{j_1 j_2}^{W_1} \cdots \sum_{j_r \in W_r} C_{j_{r-1} j_r}^{W_{r-1}} C_{j_r i}^{W_r} + \text{perm}(1, 2, \dots, r) = \mathcal{S}_0(i, W_1, \dots, W_r). \quad (3.20)$$

As a very simple example of Eq. (3.18), we show the expansion of the double-trace single-gluon integrand under the reference order $i \prec W_1$,

$$C_i \text{PT}(W_1) \text{PT}(W_2) \stackrel{\text{IBP}}{\cong} \left[\mathcal{T}_{W_2}(i, W_1) + \langle i, W_1 \rangle \right] \text{PT}(W_2), \quad (3.21)$$

where $\mathcal{T}_{W_2}(1, W_1)$ is a combination of logarithmic functions:

$$\mathcal{T}_{W_2}(i, W_1) = \sum_{j_2 \in W_2} C_{ij_2} \sum_{j \notin W_1} C_{a_1 j}^{W_1} + \sum_{j_1 \in W_1} \sum_{j_2 \in W_2} C_{ij_1} C_{j_1 j_2}^{W_1}. \quad (3.22)$$

We note that to obtain Eq. (3.18) with a more generic reference ordering in which the gluon i appears between traces, we need to use some new IBP relations that allow us to break a subcycle at a point different from the gluon attach point. We will discuss these IBP relations in Appendix B.

3.4 General cases with gluons

For the most generic string integrand (1.6), we encounter graphs with multiple tails attached to a subcycle that consists of a single trace W_i or a set of gluons. We thus need to perform IBP reductions on these *multibranch graphs* [1], which will be discussed in details in Appendix B. However, using the intuition developed in the single-gluon formula (3.18), one can be convinced that the generic recursive expansion is

$$R(i_1, \dots, i_r) \prod_{j=1}^{m+1} \text{PT}(W_j) \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W_{m+1})}{\prod_{j=1}^m (1 - s_{W_j})} \left[\alpha'^m \mathcal{T}_{W_{m+1}}(\mathbf{R}) - \sum_{\substack{\mathbf{A} \in \mathbb{P}[i_1, \dots, i_r, W_1 \dots W_m] \\ |\mathbf{A}| < r+m}} (-1)^{|\mathbf{A}|} \mathcal{J}[\mathbf{A}] \right], \quad (3.23)$$

where the first term is the logarithmic form CHY integrand for generic Yang-Mills-scalar amplitudes written under the reference order \mathbf{R} . A very convenient choice is to put all

gluons before the traces:

$$\mathbf{R} = i_1 \prec \dots \prec i_r \prec W_1 \prec \dots \prec W_m. \quad (3.24)$$

In the string context, it corresponds to the contribution from a compactified superstring.

The non-logarithmic function \mathcal{J} is generalized from Eq. (3.19) to include more gluons. For a partition \mathbf{A} that contains s singleton blocks, in which t of them are gluons and the rest traces, namely, $\mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_t, \mathbf{b}_{t+1}, \dots, \mathbf{b}_s, \mathbf{A}_{s+1}, \dots, \mathbf{A}_{|\mathbf{A}|}\}$, we have

$$\mathcal{J}[\mathbf{A}] = (-1)^s R(\mathbf{a}_1, \dots, \mathbf{a}_t) \left[\prod_{j=t+1}^s (1 - s_{\mathbf{b}_j}) \text{PT}(\mathbf{b}_j) \right] \left[\prod_{j=s+1}^{|\mathbf{A}|} \mathcal{S}_0(\mathbf{A}_j) \right]. \quad (3.25)$$

Similar to the discussion in section 3.2, the $|\mathbf{A}| < r + m$ condition in Eq. (3.23) guarantees that there must be at least one non-singleton block \mathbf{A}_j in \mathbf{A} and thus at least one nontrivial fusion $\mathcal{S}_0(\mathbf{A}_j)$. It is a linear combination of PT factors taking value in a subset of $\text{perm}(\mathbf{A}_j)$, where the coefficients contain the polarization vectors if \mathbf{A}_j contain gluons.

Now combining the contributions from every block in \mathbf{A} , we can see that $\mathcal{J}[\mathbf{A}]$ is a string integrand with total number of traces and gluons decreased due to the of nontrivial fusions. Consider the triple-trace three-gluon integrand $R(i_1, i_2, i_3) \text{PT}(W_1) \text{PT}(W_2) \text{PT}(W_3)$, with total number of gluons and traces being six. After pulling out the overall factor $\text{PT}(W_3)$, we need to consider the partitions of the set $\{i_1, i_2, i_3, W_1, W_2\}$. Two such examples are

$$\begin{aligned} \mathbf{A} = \{i_2, i_3, W_1, \{i_1, W_2\}\} &\Rightarrow \mathcal{J}[\mathbf{A}] = -R(i_2, i_3)(1 - s_{W_1}) \text{PT}(W_1) \mathcal{S}_0(i_1, W_2), \\ \mathbf{A} = \{i_1, i_2, i_3, \{W_1, W_2\}\} &\Rightarrow \mathcal{J}[\mathbf{A}] = -R(i_1, i_2, i_3) \mathcal{S}_0(W_1, W_2), \end{aligned} \quad (3.26)$$

where the first line gives a linear combination of triple-trace two-gluon integrands and the second line double-trace three-gluon integrands. For both cases, the total number of gluons and traces is five. When there are no gluons, this $\mathcal{J}[\mathbf{A}]$ reduces trivially to the one in Eq. (3.15).

We give several examples for this recursive expansion under the reference order (3.24). First, The recursive expansion for single-trace integrands with two and three gluons will be worked out in detail in Appendix C.1. As a more involving case, the single-trace four-gluon integrand can be expanded as

$$\begin{aligned} R(1, 2, 3, 4) \text{PT}(W_1) &\stackrel{\text{IBP}}{\cong} \text{PT}(W_1) \left[\mathcal{T}_{W_1}(1, 2, 3, 4) + \mathcal{S}_0(1, 2, 3, 4) + (\mathcal{S}_0(1, 2, 3)C_4 + \text{cyclic}) \right. \\ &\quad - \mathcal{S}_0(1, 2)\mathcal{S}_0(3, 4) - \mathcal{S}_0(1, 3)\mathcal{S}_0(2, 4) - \mathcal{S}_0(1, 4)\mathcal{S}_0(2, 3) \\ &\quad + R(1, 2)\mathcal{S}_0(3, 4) + R(3, 4)\mathcal{S}_0(1, 2) + R(1, 3)\mathcal{S}_0(2, 4) \\ &\quad \left. + R(2, 4)\mathcal{S}_0(1, 3) + R(1, 4)\mathcal{S}_0(2, 3) + R(2, 3)\mathcal{S}_0(1, 4) \right]. \end{aligned} \quad (3.27)$$

The next example is the expansion of the double-trace two-gluon integrand,

$$R(1, 2) \text{PT}(W_1) \text{PT}(W_2) \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W_2)}{1 - s_{W_1}} \left[\alpha' \mathcal{T}_{W_2}(1, 2, W_1) + \mathcal{S}_0(1, 2, W_1) + C_1 \mathcal{S}_0(2, W_1) \right. \\ \left. + C_2 \mathcal{S}_0(1, W_1) + (1 - s_{W_1}) \text{PT}(W_1) \mathcal{S}_0(1, 2) \right]. \quad (3.28)$$

In Appendix C, we will provide more details on the derivation of the generic formula (3.23).

4 Derivation of the CHY integrand

The recursive expansion (3.23) of string integrands might have a very wide application. In this section, we show how to use it to derive inductively the CHY integrand of the pure-scalar sector of $(DF)^2 + \text{YM} + \phi^3$, Eq. (2.20a). The derivation of the most generic integrand (2.19) is very similar and we will comment on it at the end.

The induction starts at double trace. The logarithmic function (3.4) can be further simplified by SE as

$$\text{PT}(W_1) \text{PT}(W_2) \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W_2)}{1 - s_{W_1}} \sum_{j_2 \in W_2} C_{a_1, j_2}^{W_1} \stackrel{\text{SE}}{\cong} -\frac{s_{W_1}}{1 - s_{W_1}} \text{PT}(W_1) \text{PT}(W_2). \quad (4.1)$$

We can combine the two-step process and write

$$\text{PT}(W_1) \text{PT}(W_2) \stackrel{\text{IBP+SE}}{\cong} -\frac{s_{W_1}}{1 - s_{W_1}} \text{PT}(W_1) \text{PT}(W_2). \quad (4.2)$$

This CHY integrand was first identified in [46]. We use “IBP+SE” to stand for the process of IBP reduction to logarithmic functions followed by a SE simplification.

To derive the triple-trace CHY integrand, we can directly apply the double-trace result (4.2) to the second term of Eq. (3.9):

$$\langle W_1, W_2 \rangle \text{PT}(W_3) \stackrel{\text{IBP+SE}}{\cong} \frac{-s_{W_1 W_2} \langle W_1, W_2 \rangle \text{PT}(W_3)}{1 - s_{W_1, W_2}} \\ = \left[\langle W_1, W_2 \rangle - \mathcal{S}_{\alpha'}(W_1, W_2) \right] \text{PT}(W_3). \quad (4.3)$$

Finally, the $\langle W_1, W_2 \rangle$ in the above equation, when combined with $\mathcal{T}_{W_3}(W_1, W_2)$, produces the last piece of the triple-trace CHY integrand:

$$\left[\alpha'^2 \mathcal{T}_{W_3}(W_1, W_2) + \langle W_1, W_2 \rangle \right] \text{PT}(W_3) \stackrel{\text{SE}}{\cong} s_{W_1} s_{W_2} \text{PT}(W_1) \text{PT}(W_2) \text{PT}(W_3). \quad (4.4)$$

This completes the derivation of the triple-trace CHY integrand from the string integrand

$$\prod_{i=1}^3 \text{PT}(W_i) \stackrel{\text{IBP+SE}}{\cong} \frac{\text{PT}(W_3)}{(1 - s_{W_1})(1 - s_{W_2})} \left[\mathcal{S}_{\alpha'}(W_1) \mathcal{S}_{\alpha'}(W_2) - \mathcal{S}_{\alpha'}(W_1, W_2) \right], \quad (4.5)$$

which agrees with Eq. (2.20a).

The calculation at four traces is also similar. The last two terms in the first line of Eq. (3.11) are double-trace string integrands, such that Eq. (4.2) leads to

$$\begin{aligned} \text{PT}(W_4) \left(\langle W_1, W_2, W_3 \rangle + \langle W_1, W_3, W_2 \rangle \right) \\ \stackrel{\text{IBP+SE}}{\cong} \text{PT}(W_4) \left(\langle W_1, W_2, W_3 \rangle + \langle W_1, W_3, W_2 \rangle - \mathcal{S}_{\alpha'}(W_1, W_2, W_3) \right). \end{aligned} \quad (4.6)$$

On the other hand, the terms in the second line of Eq. (3.11) are all triple-trace integrands, from which we can generate nested fusions using Eq. (4.5). For example,

$$\begin{aligned} (1 - s_{W_1}) \text{PT}(W_1) \langle W_2, W_3 \rangle \text{PT}(W_4) \\ \stackrel{\text{IBP+SE}}{\cong} \text{PT}(W_4) \left[-s_{W_1} \text{PT}(W_1) \langle W_2, W_3 \rangle + \mathcal{S}_{\alpha'}(W_1) \mathcal{S}_{\alpha'}(W_2, W_3) \right. \\ \left. - \mathcal{S}_{\alpha'}(W_1, \mathcal{S}_{\alpha'}(W_2, W_3)) \right], \end{aligned} \quad (4.7)$$

and the rest are obtained by cyclic permutations. If we plug the above two equations back to Eq. (3.11), add and subtract $\prod_{i=1}^3 [s_{W_i} \text{PT}(W_i)]$, the labeled trees are exactly canceled due to the relation

$$\begin{aligned} \alpha'^3 \mathcal{T}_{W_4}(W_1, W_2, W_3) \stackrel{\text{SE}}{\cong} - \prod_{i=1}^3 [s_{W_i} \text{PT}(W_i)] - \langle W_1, W_2, W_3 \rangle - \langle W_1, W_3, W_2 \rangle \\ + \left(s_{W_1} \text{PT}(W_1) \langle W_2, W_3 \rangle + \text{cyclic} \right), \end{aligned} \quad (4.8)$$

where we use the definition (2.28) for the Yang-Mills-scalar integrand on the right hand side. Collecting all the relevant terms, we get the four-trace CHY integrand

$$\begin{aligned} \prod_{i=1}^4 \text{PT}(W_i) \stackrel{\text{IBP+SE}}{\cong} \frac{\text{PT}(W_4)}{\prod_{i=1}^3 (1 - s_{W_i})} \left[- \prod_{i=1}^3 \mathcal{S}_{\alpha'}(W_i) - \mathcal{S}_{\alpha'}(W_1, W_2, W_3) \right. \\ \left. + \left(\mathcal{S}_{\alpha'}(W_1) \mathcal{S}_{\alpha'}(W_2, W_3) - \mathcal{S}_{\alpha'}(W_1, \mathcal{S}_{\alpha'}(W_2, W_3)) + \text{cyclic} \right) \right], \end{aligned} \quad (4.9)$$

which again agrees with Eq. (2.20a).

In fact, starting from Eq. (3.14), we can derive the $(m+1)$ -trace CHY integrand (2.20a) inductively. Since the second term of Eq. (3.14) is a linear combination of string integrands with fewer traces, we can simplify it using Eq. (2.20a) as our inductive assumption. After some algebras, one can show that the result is

$$\begin{aligned} - \sum_{\substack{\mathbf{A} \in \mathbb{P}[W_1 \dots W_m] \\ |\mathbf{A}| < m}} (-1)^{|\mathbf{A}|} \mathcal{J}[\mathbf{A}] \stackrel{\text{IBP+SE}}{\cong} \sum_{\mathbf{A} \in \mathbb{T}[W_1 \dots W_m]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_{\alpha'}(\mathbf{A}_j) - \sum_{\mathbf{A} \in \mathbb{P}[W_1 \dots W_m]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_0(\mathbf{A}_j). \end{aligned} \quad (4.10)$$

While the first term is precisely the desired $(m+1)$ -trace integrand (2.20a), the second

term exactly cancels the labeled trees:

$$\alpha'^m \mathcal{T}_{W_{m+1}}(\mathbf{R}) \stackrel{\text{SE}}{\cong} \sum_{\mathbf{A} \in \mathbb{P}[W_1 \dots W_m]} (-1)^{|\mathbf{A}|} \prod_{j=1}^{|\mathbf{A}|} \mathcal{S}_0(\mathbf{A}_j), \quad (4.11)$$

since both of them are valid CHY integrands for the same amplitude and thus must equal on the support of the scattering equations. Finally, we note that the most generic CHY integrand (2.19) for multitrace $(DF)^2 + \text{YM} + \phi^3$ can be inductively derived following a procedure similar to Eq. (4.10) and (4.11).

5 Special massless factorizations

As an important consistency check, our integrand (2.19) should demonstrate the correct factorization behavior. In particular, we consider two special massless factorization channels as shown in figure 3: (i) we cut out exactly a single trace σ . The on-shell internal propagator is thus a gluon; (ii) we also cut out part of a second trace $\rho_L \subset \rho$. The on-shell internal propagator is thus a scalar.

We start with introducing essential tools for studying factorization in the CHY framework. We consider a generic physical factorization limit $q_L^2 \rightarrow 0$, where

$$\sum_{i=1}^{n_L} k_i = -q_L, \quad \sum_{i=n_L+1}^n k_i = -q_R = q_L. \quad (5.1)$$

We follow the prescription of [14] and change the variables to

$$\begin{aligned} z_a &= \frac{\zeta}{u_a} & a \in L &= \{1, 2 \dots n_L\}, \\ z_a &= \frac{v_a}{\zeta} & a \in R &= \{n_L + 1, n_L + 2 \dots n\}, \end{aligned} \quad (5.2)$$

where we have fixed one of the v 's, say, $v_{n-1} = v_{n-1}^*$. In terms of the new variables, the scattering equations for L and R are independent of each other at the zeroth order of ζ^2 :

$$a \in L : \quad 0 = \sum_{b \in L \cup \{q_L\} \setminus \{a\}} \frac{k_a \cdot k_b}{u_{ab}} + \mathcal{O}(\zeta^2), \quad (5.3a)$$

$$a \in R : \quad 0 = \sum_{b \in R \cup \{q_R\} \setminus \{a\}} \frac{k_a \cdot k_b}{v_{ab}} + \mathcal{O}(\zeta^2), \quad (5.3b)$$

where we have used the gauge choice $u_{q_L} = v_{q_R} = 0$. On the other hand, ζ^2 satisfy the following equation,

$$0 = -\frac{q_L^2}{2} + \zeta^2 \sum_{a \in R} \sum_{b \in L} \frac{k_a \cdot k_b}{v_a u_b - \zeta^2} \equiv -\frac{q_L^2}{2} + \frac{\zeta^2}{2} F(u, v, k_i) + \mathcal{O}(\zeta^4), \quad (5.4)$$

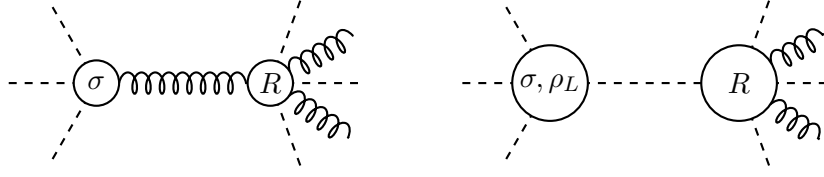


Figure 3: Two special massless factorization channels of our integrand. The dashed lines represent scalars and the curly lines represent gluons. Note that there are no original external gluons on the left.

where $F := \sum_{a \in R} \sum_{b \in L} \frac{2k_a \cdot k_b}{v_a u_b}$ is independent of ζ . In the limit $q_L^2 \rightarrow 0$, there always exists a singular solution

$$\zeta^2 = \frac{q_L^2}{F(u, v, k_i)} + \mathcal{O}(q_L^4). \quad (5.5)$$

We can ignore other solutions of ζ^2 since they are only relevant to subleading orders in the factorization limit.

The benefit of these new worldsheet variables is that, at the leading order of ζ , the CHY integration measure factorizes nicely as [14]

$$d\mu_{\text{CHY}} \sim d\mu_L d\mu_R \frac{d\zeta^2}{\zeta^2} \delta(\zeta^2 F - q_L^2) \frac{\zeta^{2n_L - 2n_R - 4}}{\mathcal{U}^4}, \quad (5.6)$$

where $\mathcal{U} := \prod_{a=1}^{n_L} u_a$, and

$$d\mu_L = (u_1 u_2 u_{12})^2 \left[\prod_{a=3}^{n_L} du_a \delta(E_a^L) \right], \quad d\mu_R = (v_{n-1} v_n u_{n-1, n})^2 \left[\prod_{a=n_L+1}^{n-2} dv_a \delta(E_a^R) \right],$$

$$E_a^L = \sum_{b \in L \cup \{q_L\} \setminus \{a\}} \frac{k_a \cdot k_b}{u_{ab}}, \quad E_a^R = \sum_{b \in R \cup \{q_R\} \setminus \{a\}} \frac{k_a \cdot k_b}{v_{ab}}. \quad (5.7)$$

Namely, $d\mu_L$ and $d\mu_R$ are nothing but the integration measure for $L \cup \{q_L\}$ and $R \cup \{q_R\}$ respectively, where $q_L^2 = q_R^2 = 0$. As a universal building block for gauge amplitudes, the Parke-Taylor factor becomes

$$\text{PT}(1, 2, \dots, n) \sim (-1)^{n_L} \zeta^{n_R - n_L + 2} \mathcal{U}^2 \text{PT}(1, 2, \dots, n_L, q_L) \text{PT}(q_R, n_L + 1, \dots, n), \quad (5.8)$$

where the two PT's on the right hand side are given by u and v variables respectively.⁶ We expect a valid CHY half-integrand $\mathcal{I}_n^{\text{CHY}}$ to behave as

$$\mathcal{I}_n^{\text{CHY}} \sim (-1)^{n_L} \zeta^{n_R - n_L + 2} \mathcal{U}^2 \sum_{\text{states}} \mathcal{I}_L^{\text{CHY}}(1, 2, \dots, n_L, q_L) \mathcal{I}_R^{\text{CHY}}(q_R, n_L + 1, \dots, n), \quad (5.9)$$

where the summation is over the on-shell states on the factorization channel $q_L^2 = 0$. If

⁶In the following, it is understood that in a factorization analysis, particles in L are always associated with the u 's while particles in R with the v 's.

this is true, then the $\zeta^{2n_R-2n_L+4}/\mathcal{U}^4$ factor in the measure (5.6) will be cancel, such that the ζ^2 integration will provide the desired massless pole:

$$\int \frac{d\zeta^2}{\zeta^2} \delta(\zeta^2 F - q_L^2) = \frac{1}{q_L^2}, \quad (5.10)$$

We thus obtain the correct factorization behavior

$$A(1, 2, \dots, n) \sim \sum_{\text{states}} A_L(1, 2, \dots, n_L, q_L) \frac{1}{q_L^2} A_R(q_R, n_{L+1}, \dots, n), \quad (5.11)$$

where A_L and A_R are the amplitude given by the half-integrand \mathcal{I}_L and \mathcal{I}_R .

5.1 An example: factorization of the triple-trace scalar integrand

We apply the above construction to study the factorization behavior of the triple-trace scalar integrand (2.21b). We first consider the scalar factorization channel. In particular, we cut through the trace ρ such that $L = \sigma \cup \rho_L$ and $R = \tau \cup \rho_R$, where $\rho_L \cup \rho_R = \rho$ and $\rho_L \cap \rho_R = \emptyset$. The leading order is contributed only by the first term of (2.21b). Using Eq. (5.2), we find that

$$\begin{aligned} \mathcal{I}_n^{\text{CHY}}(\sigma, \tau, \rho) &\sim \zeta^{n_R-n_L+2} \mathcal{U}^2 \left[\frac{\text{PT}(\rho_1) s_\sigma \text{PT}(\sigma)}{1-s_\sigma} \right] \left[\frac{\text{PT}(\rho_2) s_\tau \text{PT}(\tau)}{1-s_\tau} \right] \\ &= \zeta^{n_R-n_L+2} \mathcal{U}^2 \mathcal{I}_L^{\text{CHY}}(\sigma, \rho_1) \mathcal{I}_R^{\text{CHY}}(\tau, \rho_2), \end{aligned} \quad (5.12)$$

which is the correct factorization behavior according to Eq. (5.9). We note that in our factorization analyses, we always omit a possible overall sign, but keep track carefully the relative signs in our integrands.

Next, we consider the gluon factorization channel $L = \sigma$ and $R = \tau \cup \rho$. The first term of Eq. (2.21b) behaves as

$$\begin{aligned} \frac{\text{PT}(\rho) \mathcal{S}_{\alpha'}(\sigma) \mathcal{S}_{\alpha'}(\tau)}{(1-s_\sigma)(1-s_\tau)} &\sim \zeta^{n_R-n_L+2} \mathcal{U}^2 \text{PT}(\sigma) \sum_{a \in R, b \in L} \frac{s_{ab}}{v_a u_b} \frac{s_\tau \text{PT}(\tau) \text{PT}(\rho)}{1-s_\tau} \\ &\sim \alpha' \zeta^{n_R-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \left[\mathcal{S}_{\alpha'}(q_L) \text{PT}(\sigma) \right] \left[\frac{\text{PT}(\rho) \mathcal{S}_{\alpha'}(\tau) \mathcal{S}_{\alpha'}(q_R)}{1-s_\tau} \right]. \end{aligned} \quad (5.13)$$

Besides changing the variables to (5.2), we also replace s_σ by Eq. (5.5), which leads to the first line. To achieve the second line, we first insert the on-shell completeness relation⁷

$$\sum_{\text{states}} \epsilon_\mu(q_L) \epsilon_\nu(q_R) \rightarrow \eta_{\mu\nu} \quad (5.14)$$

⁷The arrow means that we have excluded terms that vanish on-shell.

into $s_{ab} = \alpha' k_a^\mu \eta_{\mu\nu} k_b^\nu$. Similarly, the second term of Eq. (2.21b) behaves as

$$\begin{aligned}
\frac{\text{PT}(\rho) \mathcal{S}_{\alpha'}(\sigma, \tau)}{(1-s_\sigma)(1-s_\tau)} &\sim \frac{\zeta^{n_R-n_L+2} \mathcal{U}^2 \text{PT}(\sigma) \text{PT}(\tau) \text{PT}(\rho)}{2(1-s_\tau)(1-s_\rho)} \sum_{\substack{\sigma_1, \sigma_2 \in \sigma \\ \tau_1, \tau_2 \in \tau}} \left(\frac{s_{\tau_2 \sigma_1}}{u_{\sigma_1}} s_{\sigma_2 \tau_1} - \frac{s_{\sigma_2 \tau_1}}{u_{\sigma_2}} s_{\tau_2 \sigma_1} \right) \frac{v_{\tau_1 \tau_2}}{v_{\tau_1} v_{\tau_2}} \\
&\sim \alpha'^2 \zeta^{n_R-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \left[\frac{C_{q_L} \text{PT}(\sigma) \text{PT}(\tau) \text{PT}(\rho)}{2(1-s_\tau)(1-s_\rho)} \sum_{\tau_1, \tau_2 \in \tau} \frac{(k_{\tau_2} \cdot f_{q_R} \cdot k_{\tau_1}) v_{\tau_1 \tau_2}}{v_{\tau_2} v_{\tau_1}} \right] \\
&\sim \alpha' \zeta^{n_R-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \left[\mathcal{S}_{\alpha'}(q_L) \text{PT}(\sigma) \right] \left[\frac{\text{PT}(\rho) \mathcal{S}_{\alpha'}(\tau, q_R)}{1-s_\tau} \right]. \quad (5.15)
\end{aligned}$$

We have used the completeness relation (5.14), the momentum conservation (5.1) and the gauge choice $v_{q_R} = 0$ to get the final result.

Combining Eq. (5.13) and (5.15), we find that the triple-trace integrand (2.21b) indeed factorizes into a single-trace one-gluon and a double-trace one-gluon integrand, namely,

$$\mathcal{I}_n^{\text{CHY}}(\sigma, \tau, \rho) \sim \alpha' \zeta^{n_R-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \mathcal{I}_L^{\text{CHY}}(\sigma, q_L) \mathcal{I}_R^{\text{CHY}}(\tau, \rho, q_R), \quad (5.16)$$

where $\mathcal{I}_L^{\text{CHY}}(\sigma, q_L) = C_{q_L} \text{PT}(\sigma) = -\mathcal{S}_{\alpha'}(q_L) \text{PT}(\sigma)$ and $\mathcal{I}_R^{\text{CHY}}$ is given by Eq. (2.21c). Our result thus agrees with the general requirement (5.9).

5.2 Generic integrands

We briefly talk about how the above two special factorization channels work for generic multitrace integrand (2.19). We first consider the scalar channel $L = W_1 \cup W_{m+1}^L$, where W_{m+1}^L is part of the trace W_{m+1} treated specially in Eq. (2.19). One can show that the more mixed L and R are in a PT factor, the higher order of ζ it will give rise to:

$$\text{PT}(L) \text{PT}(R) \sim \zeta^{n_R-n_L}, \quad \text{PT}(LR) \sim \zeta^{n_R-n_L+2}, \quad \text{PT}(LRLR) \sim \zeta^{n_R-n_L+4}, \text{ etc.}$$

Since the overall factor $\text{PT}(W_{m+1})$ already mixes L and R once, the leading order $\zeta^{n_R-n_L+2}$ must be contributed by the terms containing also the stand-alone factor $\text{PT}(W_1)$. Therefore, only those total partitions that have a singleton block W_1 is relevant at the leading order. This immediately leads to the correct factorization behavior

$$\begin{aligned}
\mathcal{I}_n^{\text{CHY}}(W_1, \dots, W_{m+1}, i_1, \dots, i_r) &\sim \zeta^{n_R-n_L+2} \mathcal{U}^2 \mathcal{I}_L^{\text{CHY}}(W_1, W_{m+1}^L) \\
&\quad \times \mathcal{I}_R^{\text{CHY}}(W_2, \dots, W_m, W_{m+1}^R, i_1, \dots, i_r), \quad (5.17)
\end{aligned}$$

where $W_{m+1}^L \cup W_{m+1}^R = W_{m+1}$ and $W_{m+1}^L \cap W_{m+1}^R = \emptyset$.

The analysis of the gluon channel $L = W_1$ is only slightly more difficult. The relevant terms are either of the form $\text{PT}(LR)$ or $s_L \text{PT}(L)$. These terms are generated by $\mathcal{S}_{\alpha'}(W_1, \dots)$, where the “...” may contain traces, gluons, or their symmetrized fusion $\mathcal{S}_{\alpha'}$.

A calculation similar to Eq. (5.13) and (5.15) shows that

$$\mathcal{S}_{\alpha'}(W_1, \dots) \sim \zeta^{\#-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \mathcal{I}_L^{\text{CHY}}(W_1, q_L) \mathcal{S}_{\alpha'}(q_R, \dots), \quad (5.18)$$

where $\#$ is the number of particles contained in “...”. In other words, we can simply replace W_1 by q_R . We note that the prescription still holds when $\mathcal{S}_{\alpha'}(W_1, \dots)$ is nested in some other symmetrized fusions:

$$\mathcal{S}_{\alpha'}(\dots, \mathcal{S}_{\alpha'}(W_1, \dots)) \sim \zeta^{\#-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \mathcal{I}_L^{\text{CHY}}(W_1, q_L) \mathcal{S}_{\alpha'}(\dots, \mathcal{S}_{\alpha'}(q_R, \dots)). \quad (5.19)$$

Although the outer level $\mathcal{S}_{\alpha'}$ may further mix W_1 with the others, the contribution is subleading and thus can be ignored. On the other hand, the terms still of the form $\text{PT}(LR)$ are exactly captured by (5.19). The final result of this channel is

$$\begin{aligned} \mathcal{I}_n^{\text{CHY}}(W_1, \dots, W_{m+1}, i_1, \dots, i_r) &\sim \zeta^{n_R-n_L+2} \mathcal{U}^2 \sum_{\text{states}} \mathcal{I}_L^{\text{CHY}}(W_1, q_L) \\ &\times \mathcal{I}_R^{\text{CHY}}(W_2, \dots, W_{m+1}, i_1, \dots, i_r, q_R), \end{aligned} \quad (5.20)$$

agreeing with Eq. (5.9). Schematically, we can obtain $\mathcal{I}_R^{\text{CHY}}$ by replacing the trace W_1 by q_R in the original $\mathcal{I}_n^{\text{CHY}}$.

By iterating the two cuts discussed above, we eventually land on the single-trace integrand (2.20b). This factorization analysis provides a simple but nontrivial consistency check to our integrands.

6 Conclusion and discussion

In this work we have continued our study of the two-step method proposed in [1]: namely (1) IBP reduction of correlators of string amplitudes to a logarithmic function, and (2) rewriting the logarithmic function into a closed-form CHY half-integrand for field-theory amplitudes using scattering equations. We present two main results regarding heterotic and compactified bosonic strings for arbitrary multiplicities and number of traces. The final outcome of our calculation is remarkably simple CHY formulas for general $(DF)^2 + \text{YM} + \phi^3$ amplitudes, which extend our previous formula for the single-trace case greatly. As a paraphrase, our result gives the half-integrand needed in a rewriting of open-string amplitude as a CHY formula. In order to derive the formula, we find that the key new result is a recursive expansion for multi-trace string correlators. It provides an efficient algorithm for reducing multi-trace correlator to logarithmic functions, which is useful for other purposes as well. Note that we have left out one special case, which is the formula for pure-graviton case: while it can be obtained from factorization already from our single-trace formula, it would be highly desirable to obtain a closed-form result for it as well.

Our results opens various interesting avenues for further investigations. First of all, they may bring new insight into the $(DF)^2 + \text{YM} + \phi^3$ theory especially in the multi-trace sector, as well as conformal (super-)gravity which can be obtained from a double copy with

(super)-YM [47, 65]. Moreover, it would be interesting to compute matrix elements with higher-dimensional operators from bosonic/heterotic string corrections, and our formulas can serve as a starting point for extracting such corrections, in a way similar to [66, 67]. Of course having such a general formula for a large class of amplitudes, including those in Einstein-Yang-Mills theory, provides more applications. For example, one could use it for extracting BCJ numerators and discovering new amplitude relations, which have been recently studied further in [61, 68–70].

The recursive expansion certainly has more applications. Most directly it gives the BCJ numerators for the $(DF)^2 + \text{YM} + \phi^3$ theory. From a more mathematical point of view, it allows us to reduce non-logarithmic functions with multiple cycles to logarithmic ones, both for IBP reduction of string correlator and, in the $\alpha' \rightarrow \infty$ limit, for manipulating CHY integrand using SE. As shown in the paper and in the ancillary file, the recursive expansion allows us to do such calculations in a very efficient way. It is also interesting to relate our general procedure to various ideas in the literature, such as intersection theory [29, 71], studies of disk/sphere integrals from a mathematical point of view [51, 52, 72], and positive geometries related to string worldsheet [59, 73, 74].

As we have pointed out in [1], our method applies to any string correlator for massless external states with the correct $\text{SL}(2)$ weight, and it would be interesting to study more examples beyond type I, bosonic and heterotic cases, such as the dual model proposed in ref. [75]. More importantly, it would be highly desirable to apply our method to string correlators with massive states, as well as to cases at genus one [76–82]. On the other hand, the CHY half-integrands here contain explicit α' dependence, and it would be interesting to understand if there are worldsheet models, such as ambitwistor string [20, 24, 25] theory, underpin it (see [28] for some progress which gives at least correct three-point amplitudes). Investigations along these lines may shed new light into the universality and origin of CHY/ambitwistor string constructions.

Acknowledgments

It is our pleasure to thank Thales Azevedo, Yi-Jian Du, Bo Feng, Henrik Johansson, Oliver Schlotterer and Ellis Yuan for inspiring discussions. S.H.’s research is supported in part by the Thousand Young Talents program, the Key Research Program of Frontier Sciences of CAS under Grant No. QYZDB-SSW-SYS014 and Peng Huanwu center under Grant No. 11747601. F.T. is supported by the Knut and Alice Wallenberg Foundation under grant KAW 2013.0235, and the Ragnar Söderberg Foundation (Swedish Foundations’ Starting Grant).

A Labeled trees and logarithmic Yang-Mills-scalar integrands

As shown in ref. [83], labeled trees form a basis for logarithmic functions on the worldsheet. In this section, we give the rules to write down the logarithmic Yang-Mills-scalar CHY

integrand $\mathcal{T}_{W_{m+1}}(\mathbf{R})$ in terms of a labeled-tree expansion:

$$\mathcal{T}_{W_{m+1}}(\mathbf{R}) = \sum_{T \in \mathbf{T}(W_{m+1})} N_{\mathbf{R}}(T) \mathcal{C}(T). \quad (\text{A.1})$$

To carry out this expansion, we need to first construct the relevant labeled trees, and then define the map $N_{\mathbf{R}}$ and \mathcal{C} for each labeled tree. The summation in Eq. (A.1) is over $\mathbf{T}(W_{m+1})$, the labeled trees with roots in W_{m+1} , the nodes of which are labels of all the external particles. The function \mathcal{C} maps a tree T into a rational function of worldsheet variables:⁸ each edge is mapped to a z_{ij} factor in the denominator, where i and j are the labels of the nodes connected by the edge. Each tree also carries a dual kinematic factor $N_{\mathbf{R}}$, the evaluation of which depends on the choice of reference order \mathbf{R} . Very interestingly, these dual kinematic factors form a basis for the DDM basis BCJ numerators [60, 61], while the reference order \mathbf{R} characterizes certain generalized gauge degrees of freedom.

We start with constructing the relevant labeled trees $\mathbf{T}(W_{m+1})$. We first treat the gluons and traces on the same footing, and draw all the rooted trees on W_{m+1} with nodes $\{i_1, \dots, i_r, W_1, \dots, W_m\}$. In all there are $(r+m+1)^{r+m-1}$ such trees. For example, the spanning trees for the double-trace single-gluon case are

$$\begin{array}{ccc} \begin{array}{c} W_1 \quad i \\ | \quad | \\ \bullet \quad \bullet \\ | \\ W_2 \end{array} & \begin{array}{c} i \quad W_1 \\ | \quad | \\ \bullet \quad \bullet \\ | \\ W_2 \end{array} & \begin{array}{c} W_1 \quad i \\ | \quad / \\ \bullet \quad \bullet \\ | \\ W_2 \end{array} . \end{array} \quad (\text{A.2})$$

Next, given a reference order \mathbf{R} , we decompose each tree into a collection of paths and blow up the traces according to the following procedures:

- (1) draw a path from the first element of \mathbf{R} to the root. Then draw another path towards the root from the first element of \mathbf{R} that has not been traversed. This path will end on a previous path. Repeat the process until all nodes are traversed. This decomposes each tree into a set of paths, denoted as $\mathcal{P}[T]$.
- (2) replace the root by a chain evaluated to $\text{PT}(W_{m+1})$ after restoring the gauge:

$$W_{m+1} \xrightarrow{\bullet} \bigcirc, \quad \mathcal{C}\left(\bigcirc\right) \xrightarrow{\text{restore gauge}} \text{PT}(W_{m+1}). \quad (\text{A.3})$$

- (3) if a trace W_i appears in the mid of a path, blow it up according to

$$W_i \xrightarrow{\bullet} \begin{array}{c} a_i \quad b_i \\ \text{---} \text{wavy} \text{---} \\ \bullet \end{array}, \quad (\text{A.4})$$

where by our convention b_i is the end closer to the root. We will sum over all pairs of a_i and b_i in W_i .

⁸In [83], it is called the Cayley function. For a trivial tree with a single node, we define $\mathcal{C}[\bullet] = 1$.

(4) if a trace W_i appears at the start of a path, still blow it up as (A.4). However, only b_i will be summed in W_i , while $a_i \in W_i$ is arbitrary but fixed. Across our construction, we keep the same choice of a_i if this situation happens.⁹

(5) if a path ends on a trace W_i , then the end point can take any value in W_i .

Accordingly, the three spanning trees in Eq. (A.2) generate the following labeled trees relevant to the logarithmic CHY integrand:

$$\mathbf{T}(W_2) : \quad \begin{array}{ccc}
 \begin{array}{c} b_1 \quad a_1 \quad i \\ \text{---} \text{---} \text{---} \\ \downarrow \\ (j_2) \end{array} & \begin{array}{c} i \quad b_1 \quad a_1 \\ \text{---} \text{---} \text{---} \\ \downarrow \\ (j_2) \end{array} & \begin{array}{c} a_1 \quad b_1 \quad i \\ \text{---} \text{---} \text{---} \\ \downarrow \\ (l_2) \quad j_2 \end{array}, \quad (\text{A.5}) \\
 a_1, b_1 \in W_1 & b_1 \in W_1 & b_1 \in W_1 \\
 j_2 \in W_2 & j_2 \in W_2 & j_2, l_2 \in W_2
 \end{array}$$

in which we have used the reference order $\mathbf{R} = i \prec W_1$. All the paths are directed towards the root, and different ones are illustrated by different colors. For each $T \in \mathbf{T}(W_{m+1})$, the map \mathcal{C} is defined as

$$\begin{array}{c} i \\ \text{---} \text{---} \text{---} \\ \rightarrow \\ j \end{array} \rightarrow \frac{1}{z_{ij}}, \quad \begin{array}{c} a_i \\ \text{---} \text{---} \text{---} \\ b_i \end{array} \rightarrow \text{PT}(W_i) z_{b_i a_i}. \quad (\text{A.6})$$

This definition is compatible with the one introduced in section 2.1. On the other hand, for each path p in the path set $\mathcal{P}[T]$, we can define a path factor $\varphi(p)$ obtained from the following rule:

node	position in the path		
	start	middle	end
gluon i	ϵ_i^μ	$f_i^{\mu\nu}$	k_i^ν
trace W_i	$k_{b_i}^\mu$	$k_{a_i}^\mu, k_{b_i}^\nu$	$k_{j_i}^\nu$

(A.7)

where the Lorentz indices are contracted with their neighbors on the path. The map $N_{\mathbf{R}}(T)$ is given by the product of all these path factors:

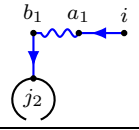
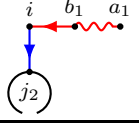
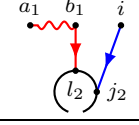
$$N_{\mathbf{R}}(T) = \prod_{p \in \mathcal{P}[T]} \varphi(p). \quad (\text{A.8})$$

The outcome of $N_{\mathbf{R}}$ depends on the reference order \mathbf{R} , since different \mathbf{R} 's lead to different path sets for a given tree.

According to Eq. (A.6) and (A.7), the labeled trees in Eq. (A.5) are evaluated as

⁹This choice eliminates some redundancy in the construction. In [61], a_i is called a fiducial particle.

follows under the reference order $\mathbf{R} = i \prec W_1$:

	\sum_T	$N_{\mathbf{R}}(T)$	$\mathcal{C}(T)$
	$\sum_{a_1, b_1 \in W_1} \sum_{j_2 \in W_2}$	$(\epsilon_i \cdot k_{a_1})(k_{b_1} \cdot k_{j_2})$	$\frac{\text{PT}(W_1) z_{b_1 a_1}}{z_{i a_1} z_{b_1 j_2}}$
	$\sum_{b_1 \in W_1} \sum_{j_2 \in W_2}$	$(\epsilon_i \cdot k_{j_2})(k_{b_1} \cdot k_i)$	$\frac{\text{PT}(W_1) z_{b_1 a_1}}{z_{b_1 i} z_{i j_2}}$
	$\sum_{b_1 \in W_1} \sum_{j_2, l_2 \in W_2}$	$(\epsilon_i \cdot k_{j_2})(k_{b_1} \cdot k_{l_2})$	$\frac{\text{PT}(W_1) z_{b_1 a_1}}{z_{b_1 j_2} z_{i l_2}}$

such that the logarithmic integrand $\mathcal{T}_{W_2}(i, W_1)$ is obtained simply by adding the three rows together. In particular, the $a_1 \in W_1$ in the second and third row is the same, and not summed over. Different choice of a_1 leads to equivalent \mathcal{T}_{W_2} , and thus it exposes certain redundancy in both the string and CHY integrand. We note that if we choose $\mathbf{R} = W_1 \prec i$ instead, the first two classes of labeled trees in the above table are modified into

$$\begin{array}{c}
 \begin{array}{c}
 \text{Diagram: } j_2 \text{ (root) } \xrightarrow{\text{blue}} b_1 \xrightarrow{\text{blue}} j_1 \xrightarrow{\text{blue}} a_1 \xrightarrow{\text{red}} i \\
 \text{Diagram: } i \xrightarrow{\text{red}} b_1 \xrightarrow{\text{red}} a_1
 \end{array}
 \end{array}
 \longrightarrow
 \sum_{j_1, b_1 \in W_1} \sum_{j_2 \in W_2} (k_{b_1} \cdot k_{j_2})(\epsilon_i \cdot k_{j_1}) \frac{\text{PT}(W_1) z_{b_1 a_1}}{z_{i j_1} z_{b_1 j_2}}, \quad (\text{A.10a})$$

$$\begin{array}{c}
 \begin{array}{c}
 \text{Diagram: } j_2 \text{ (root) } \xrightarrow{\text{blue}} i \\
 \text{Diagram: } i \xrightarrow{\text{red}} b_1 \xrightarrow{\text{red}} a_1
 \end{array}
 \end{array}
 \longrightarrow
 \sum_{b_1 \in W_1} \sum_{j_2 \in W_2} (k_{b_1} \cdot f_i \cdot k_{j_2}) \frac{\text{PT}(W_1) z_{b_1 a_1}}{z_{b_1 i} z_{i j_2}}, \quad (\text{A.10b})$$

while the third class remains the same. The resultant integrand $\mathcal{T}_{W_2}(W_1, i)$ is of course equivalent to $\mathcal{T}_{W_2}(i, W_1)$ both as string and CHY integrand.

B IBP reduction of multibranch graphs

In our previous letter [1], we have shown that generic multibranch graphs can be algebraically rearranged into tadpoles and then processed by using Eq. (3.2). In this section, we introduce a new IBP reduction for multibranch graphs that naturally leads to our recursive expansion (3.23).

We may view a multibranch graph as a collection of subtrees planted on a subcycle consisting of a color trace and/or gluons. If we denote the subcycle as W , each node $i \in W$ is the root of a tree B_i . Moreover, we use \mathbf{s}_i to denote the immediate successors of i in the tree B_i . By definition, the set \mathbf{s}_i can be empty while B_i at least contains one node, the root

i . For generic multibranch graphs, we have

$$\mathcal{C} \left[\begin{array}{c} \text{---} \\ \text{---} \end{array} \right] = \text{PT}(W) \prod_{i \in W} \mathcal{C}[\mathbf{B}_i], \quad (\text{B.1})$$

where \mathcal{C} is defined in Appendix A. As special cases, single subcycles correspond to all $\mathbf{B}_i = \{i\}$ while tadpoles have exactly one nontrivial \mathbf{B}_i that at the same time is a chain.

For any multibranch graph, we can absorb all the $\frac{1}{z_{bj}}$ factors with $b \in W$ and $j \in s_b$ into the Koba-Nielsen factor. Then using directly Eq. (3.2), we get

$$\text{PT}(W)(\dots) \prod_{i \in W} \mathcal{C}[\mathbf{B}_i] \stackrel{\text{IBP}}{\cong} \frac{(\dots)}{1 - s_W} \sum_{b \in W} \text{PT}(W) z_{ba} \left[\sum_{j \notin W \cup s_b} \frac{s_{bj}}{z_{bj}} + \sum_{j \in s_b} \frac{s_{bj} - 1}{z_{bj}} \right] \prod_{i \in W} \mathcal{C}[\mathbf{B}_i], \quad (\text{B.2})$$

where as before $a \in W$ is arbitrary and (\dots) does not involve any punctures in W . Both terms on the right hand side of Eq. (B.2) contain *induced subcycles*, which are not present in the original integrand but appear as a result of IBP. Here, they consists of nodes originally in the branches but only part of the nodes in W . To manifest the recursive pattern, we need further operations to make all the nodes in W to appear in the induced subcycles. We first demonstrate this process by an example.

We can treat the simplest tadpole $\text{PT}(W) \frac{1}{z_{pq}}$ as multibranch and apply Eq. (B.2). The generalization is that we can now break the subcycle at any point, not just the tail attach point p . The result is

$$\frac{\text{PT}(W)(\dots)}{z_{pq}} \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W)(\dots)}{1 - s_W} \left[\sum_{\substack{b \in W \\ j \notin W \cup \{q\}}} \frac{z_{ba} s_{bj}}{z_{bj} z_{pq}} - \sum_{b \in W \setminus \{p\}} \frac{z_{ba} s_{bq}}{z_{bq} z_{qp}} + \frac{z_{pa}(1 - s_{pq})}{z_{pq} z_{qp}} \right], \quad (\text{B.3})$$

where $p \in W$ and $\mathbf{B}_p = \overset{p}{\rightleftarrows} q$. The last term is a tadpole, and the numerator cancels the tachyon pole introduced by the IBP relation (3.2),

$$\frac{\text{PT}(W) z_{pa}(1 - s_{pq})(\dots)}{z_{pq} z_{qp}} \stackrel{\text{IBP}}{\cong} \text{PT}(W) \left[\sum_{b \in W \setminus \{p\}} \frac{z_{pa} s_{bq}}{z_{bq} z_{qp}} + \sum_{j \notin W \cup \{q\}} \frac{z_{pa} s_{qj}}{z_{pq} z_{qj}} \right] (\dots). \quad (\text{B.4})$$

Noticing that $z_{ba} - z_{pa} = z_{bp}$, we can collapse the first term of (B.4) and the second term of (B.3) into a single subcycle. The final result is

$$\frac{\text{PT}(W)(\dots)}{z_{pq}} \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W)(\dots)}{1 - s_W} \left[\sum_{\substack{b \in W \\ j \notin W \cup \{q\}}} \frac{z_{ba} s_{bj}}{z_{bj} z_{pq}} + \sum_{j \notin W \cup \{q\}} \frac{z_{pa} z_{qj}}{z_{pq} z_{qj}} + \sum_{b \in W} \frac{z_{pb} s_{bq}}{z_{bq} z_{qp}} \right]. \quad (\text{B.5})$$

The first two terms are trees planted on the remaining integrand. The third term features

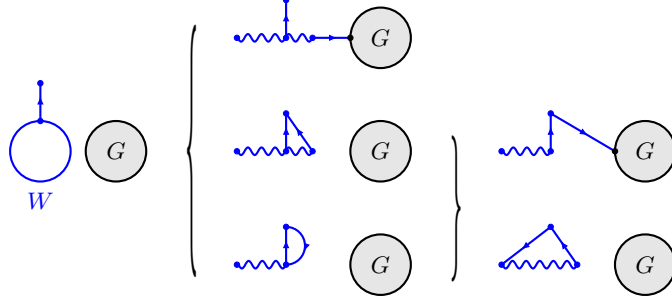


Figure 4: The IBP reduction of $\text{PT}(W)\frac{1}{z_{pq}}$, where G denotes other parts of the graph disconnected to the tadpole.

a “fusion” between the subcycle W and the branch, and it can be further broken by using Eq. (3.2). Schematically, we can represent the above reduction process as in figure 4.

The example gives some important ideas on the reduction of generic multibranch graphs. As discussed before, the goal is to include all nodes in the original subcycle W (represented by the wavy line) into the induced subcycles. In Eq. (B.2), we have generated a family of length-two subcycles $\text{PT}(b, j)$ featuring a numerator $(s_{bj} - 1)$, where $b \in W$ and $j \in \mathfrak{s}_j \subset \mathcal{B}_p$. Further IBP reduction on them will not lead to new tachyon poles. Next, we can absorb all the edges connecting j and its immediate successors \mathfrak{s}_j into the Koba-Nielsen and break the subcycle at b . In this way, we can push the induced length-two subcycles with numerators $(s_{bj} - 1)$ towards the end of the branches. This process ends at the leaves of each branch \mathcal{B}_p , where length-two tadpoles are generated and processed using relations like Eq. (B.4). Then after using some algebraic identities, we arrive at

$$\text{PT}(W)(\dots) \prod_{i \in W} \mathcal{C}[\mathcal{B}_i] \stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W)(\dots)}{1 - s_W} \times \left[\sum_{\substack{p \in W \\ j \notin \cup_i \mathcal{B}_i}} \sum_{\tilde{p} \in \mathcal{B}_p} \frac{z_{pa} s_{\tilde{p}j}}{z_{\tilde{p}j}} + \frac{1}{2} \sum_{\substack{p \in W \\ r \in W \\ (\tilde{p}, \tilde{r}) \neq (p, r)}} \sum_{\tilde{p} \in \mathcal{B}_p, \tilde{r} \in \mathcal{B}_r} \frac{z_{pr} s_{\tilde{p}\tilde{r}}}{z_{\tilde{p}\tilde{r}}} \right] \prod_{i \in W} \mathcal{C}[\mathcal{B}_i], \quad (\text{B.6})$$

where the $\frac{1}{2}$ in the second term cancels a double counting in the summation. The two terms in Eq. (B.6) can be represented by two kinds of graphs

$$\text{Eq. (B.6)} \rightarrow \left\{ \begin{array}{l} \text{Tree topology: } \text{blob } W \text{ with branches } a, \dots, \tilde{p} \text{ and } j \\ \text{Multibranch topology: } \text{blob } W \text{ with branches } a, \dots, \tilde{p}, \tilde{r} \text{ and } \tilde{p} \end{array} \right\}, \quad (\text{B.7})$$

where we represent each branch \mathcal{B}_i by a blob. Very nicely, the first term has a tree topology and there exists an arbitrary choice $a \in W$, while the second term has a multibranch topology and no arbitrary choice is involved.¹⁰ This resembles the structure of our recursive

¹⁰In practice, we often choose a to be the attach point of a certain branch, see Appendix C.

expansion (3.23). The induced subcycles all contain the original subcycle W as a whole (the wavy line) and at least one node from the branches, such that the total number of nodes in the branches is reduced. We end our discussion with an example,

$$\begin{aligned} \frac{\text{PT}(W)(\dots)}{z_{12}z_{23}z_{45}} &\stackrel{\text{IBP}}{\cong} \frac{\text{PT}(W)(\dots)}{(1-s_W)z_{12}z_{23}z_{45}} \left[\sum_{j \notin W \cup \{2,3,5\}} \left(\sum_{b \in W} z_{ba} \frac{s_{bj}}{z_{bj}} + \frac{z_{1a}s_{2j}}{z_{2j}} + \frac{z_{1a}s_{3j}}{z_{3j}} + \frac{z_{4a}s_{5j}}{z_{5j}} \right) \right. \\ &\quad \left. + \frac{z_{14}s_{25}}{z_{25}} + \frac{z_{14}s_{35}}{z_{35}} + \sum_{b \in W} \left(\frac{z_{b1}s_{b2}}{z_{b2}} + \frac{z_{b1}s_{b3}}{z_{b3}} + \frac{z_{b4}s_{b5}}{z_{b5}} \right) \right], \end{aligned} \quad (\text{B.8})$$

where $1, 4 \in W$ and $2, 3, 5 \notin W$. The branches are $B_1 = \underline{1} \rightarrow \underline{2} \rightarrow \underline{3}$ and $B_4 = \underline{4} \rightarrow \underline{5}$. As before $a \in W$ is arbitrary and (\dots) does not involve any punctures in $W \cup \{2, 3, 5\}$.

C Towards the generic recursive expansion

In this section, we try to derive the generic recursive expansion (3.23) from the string integrand (1.6) with r gluons and $m+1$ traces. The reduced gluon integrand $R(i_1, \dots, i_r)$ gives rise to several new features compared with the pure-scalar case. Besides the traces $\text{PT}(W_i)$, there are new length-two gluon subcycles introduced by the $\mathcal{R}_{(ij)} = -\frac{\epsilon_i \cdot \epsilon_j}{\alpha'}$ factors. There are additional gluon subcycles contributed by the product of C_i 's, which are of the form $C_{ij}C_{jk}C_{ki}$. To derive the generic recursive expansion, a very useful start point is to rewrite the gluon part *algebraically* into the following form:

$$\begin{aligned} R(i_1, \dots, i_r) &= - \sum_{\substack{\mathbf{A} \in \mathbb{P}[i_1, \dots, i_r] \\ |\mathbf{A}| < r}} (-1)^{|\mathbf{A}| + |\text{sg}(\mathbf{A})|} R(\text{sg}(\mathbf{A})) \prod_{|\mathbf{A}_j|=2} \mathcal{S}_0(\mathbf{A}_j) \prod_{|\mathbf{A}_j| \geq 3} \left[\mathcal{S}_0(\mathbf{A}_j)_{\epsilon \cdot \epsilon \rightarrow 0} \right] \\ &\quad + \sum_{\substack{\mathbf{A} \in \mathbb{P}[i_1, \dots, i_r] \\ \text{with all } |\mathbf{A}_j| \leq 2}} \det(-\mathbf{C}_{\text{sg}(\mathbf{A})}) \prod_{|\mathbf{A}_j|=2} (1 - s_{\mathbf{A}_j}) \mathcal{R}_{(\mathbf{A}_j)}, \end{aligned} \quad (\text{C.1})$$

where the summation in the first line is over all the partitions $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{|\mathbf{A}|}\}$ of the gluon set $\{i_1, \dots, i_r\}$ except for the all-singleton partition $\mathbf{A} = \{i_1, \dots, i_r\}$. The set $\text{sg}(\mathbf{A}) := \{\mathbf{A}_j \in \mathbf{A} \mid |\mathbf{A}_j| = 1\}$ is the collection of all the singleton blocks in \mathbf{A} . Suppose there are exactly s gluons $\{i'_1, i'_2, \dots, i'_s\} \subset \{i_1, i_2, \dots, i_r\}$ which are singleton blocks in a certain partition \mathbf{A} , then we have $R(\text{sg}(\mathbf{A})) = R(i'_1, \dots, i'_s)$. The second factor $\prod_{|\mathbf{A}_j|=2} \mathcal{S}_0(\mathbf{A}_j)$ is a product of all length-two blocks in the certain partition \mathbf{A} . The third factor comes from the gluon subcycles of the form $C_{ij}C_{jk}C_{ki}$ mentioned before. For example,

$$\mathcal{S}_0(i, j, k)_{\epsilon \cdot \epsilon \rightarrow 0} = \left(\langle i, j, k \rangle + \langle i, k, j \rangle \right)_{\epsilon \cdot \epsilon \rightarrow 0} = C_{ij}C_{jk}C_{ki} + C_{ik}C_{kj}C_{ji}. \quad (\text{C.2})$$

The summation in the second line is over all the partitions with only singleton and length-two blocks. The matrix $\mathbf{C}_{\text{sg}(\mathbf{A})} = \mathbf{C}_{i'_1, \dots, i'_s}$ is an $s \times s$ matrix whose off-diagonal entries are C_{ij} and the diagonal ones are $-C_i$. According to the matrix-tree theorem [63], $\det(-\mathbf{C}_{i'_1, \dots, i'_s})$ is a combination of labeled trees rooted on the complement set of $\{i'_1, \dots, i'_s\}$ in $\{i_1, \dots, i_r, W_1, \dots, W_{m+1}\}$.

The benefit of this rewriting is the following. Comparing with the recursive expansion (3.23), one can easily see that the first line of Eq. (C.1) is already part of the \mathcal{J} in Eq. (3.23). More precisely, it is the part that only gluons are involved in the fusions. However, the $\epsilon \cdot \epsilon$ contributions in the fusions with length-three and beyond are missing. Therefore, to reach the recursive expansion, this line does not need any further manipulation and we only need to perform IBP reduction on the second line of Eq. (C.1). Interestingly, all the gluon subcycles there are length-two and dressed by a numerator $(1 - s_{ij})$, which cancels the tachyon pole generated by the IBP. For convenience, we later refer the length-two gluon subcycles in the second line of Eq. (C.1) as *B-type*, and all the subcycles in the first line of Eq. (C.1) as *C-type*.

We are now ready to give the IBP algorithm that leads to the generic recursive expansion (3.23). We need to first gauge fix a puncture in W_{m+1} to infinity and then fix a reference order \mathbf{R} for gluons and the rest of traces as the priority list of being processed by IBP. It is convenient to put gluons before traces, for example, using the order (3.24), although there are no restrictions in principle. Starting with the first element in \mathbf{R} ,

- *If it is a gluon, say i , for each term in the string integrand, we do the following:*
 - (g1) If i appears in the B-type subcycle $(1 - s_{ij})\mathcal{R}_{(ij)}$, choose $a = i$ and break it using (B.6).
 - (g2) If i appears in a branch of a trace W_j or B-type subcycle $(1 - s_{jk})\mathcal{R}_{(jk)}$, we choose a as the attach point of that branch and break the subcycle using (B.6).
 - (g3) Repeat this process until in every term i is connected to the root W_{m+1} or an induced subcycle.¹¹ Then proceed to the next element in \mathbf{R} .
- *If it is a trace, say W_i , we choose the same $a_i \in W_i$ for every term in the string integrand and do the following:*
 - (t1') If W_i is in its original form and has not been processed, break it at a_i using (B.6). This turns W_i , together with its branches, into a tree planted on the other part of the integrand.
 - (t2') If the tree generated in step (t1') appears in a branch of another trace or B-type subcycle in its original form, choose a as the attach point of the branch and break the subcycle using (B.6).
 - (t3') Repeat this process until in every term W_i is connected to the root W_{m+1} or an induced subcycle. Then proceed to the next element in \mathbf{R} .

We note that the trace rules are the direct generalization of (t1), (t2) and (t3) for pure-scalar cases. The algorithm terminates when all the elements in \mathbf{R} are traversed. The outcome will provide the $\epsilon \cdot \epsilon$ terms in the labeled trees and in the gluon fusions together with the fusions involving traces. We will then arrive at the recursive expansion (3.23) after some straightforward algebra.

¹¹See the definition of induced subcycles below Eq. (B.2). We note that i can never connect to a C-type subcycle since they do not appear in the second line of Eq. (C.1).

C.1 Examples: two and three gluons with a single trace

We will demonstrate by two examples that the above algorithm indeed leads to the correct expansion (3.23).

The single-trace two-gluon string integrand is $R(1, 2)\text{PT}(W_1) = (C_1 C_2 + \mathcal{R}_{(12)})\text{PT}(W_1)$. We can expand the $C_1 C_2 \text{PT}(W_1)$ part as

$$C_1 C_2 \text{PT}(W_1) = \left[\sum_{\ell_1, \ell_2 \in W_1} C_{1\ell_1} C_{2\ell_2} + \sum_{j \in W_1} \underbrace{\frac{\epsilon_1 \cdot k_2 \epsilon_2 \cdot k_j}{z_{12} z_{2,j}}}_{C_{12} C_{2j}} + C_{21} \sum_{j \in W_1} C_{1j} + \underbrace{\frac{\epsilon_1 \cdot k_2 \epsilon_2 \cdot k_1}{z_{12} z_{21}}}_{C_{12} C_{21}} \right] \text{PT}(W_1), \quad (\text{C.3})$$

where the first three terms of the right hand side are combinations in labeled trees and the last term is a C-type subcycle. We rewrite the subcycle $\mathcal{R}_{(12)}$ as $\mathcal{R}_{(12)}(1 - s_{12}) + \mathcal{R}_{(12)} s_{12}$ and perform IBP on the the first part,

$$\mathcal{R}_{(12)} \text{PT}(W_1) \stackrel{\text{IBP}}{\cong} \left(\sum_{j \in W_1} \frac{\epsilon_1 \cdot \epsilon_2 k_2 \cdot k_j}{z_{12} z_{2j}} - \frac{\epsilon_1 \cdot \epsilon_2 k_2 \cdot k_1}{z_{12} z_{21}} \right) \text{PT}(W_1). \quad (\text{C.4})$$

Combining these two equations, we have

$$R(1, 2)\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} \underbrace{\left(\sum_{\ell_1, \ell_2 \in W_1} C_{1\ell_1} C_{2\ell_2} + \sum_{j \in W_1} \frac{\epsilon_1 \cdot f_2 \cdot k_j}{z_{12} z_{2,j}} + C_{21} \sum_{j \in W_1} C_{1j} + \langle 1, 2 \rangle \right)}_{\mathcal{T}_{W_1}(1, 2)} \text{PT}(W_1), \quad (\text{C.5})$$

which agrees with the general formula (3.23). The result corresponds to the reference order $1 \prec 2$ because we choose to break the subcycle at gluon 1 in Eq. (C.4), *cf.* the rule (g1).

We then try to derive the recursive expansion for the single-trace three-gluon string integrand. According to Eq. (3.23), it is

$$R(1, 2, 3)\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} \left[\mathcal{T}_{W_1}(1, 2, 3) + \langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle + \left(\langle 1, 2 \rangle C_3 + \text{cyclic} \right) \right] \text{PT}(W_1). \quad (\text{C.6})$$

As described in Appendix A, there are 16 relevant spanning trees rooted on W_1 . Each such spanning tree can be decomposed into a collection of paths according to the reference ordering $1 \prec 2 \prec 3$. In the following, we will use a set of paths to denote a spanning tree. For examples,

$$\begin{array}{c} 1 \quad 2 \quad 3 \\ \diagdown \quad | \quad / \\ \quad \quad \quad W_1 \end{array} = \{ \overset{1}{\rightarrow} W_1, \overset{2}{\rightarrow} W_1, \overset{3}{\rightarrow} W_1 \}, \quad \begin{array}{c} 1 \quad 2 \quad 3 \\ \rightarrow \quad \rightarrow \quad \rightarrow \\ \quad \quad \quad | \\ \quad \quad \quad W_1 \end{array} = \{ \overset{1}{\rightarrow} \overset{2}{\rightarrow} W_1, \overset{3}{\rightarrow} \overset{2}{\rightarrow} \}, \quad (\text{C.7})$$

where different paths are drawn with different colors.

We rewrite the string integrand this way,

$$R(1, 2, 3)\text{PT}(W_1) = \left[\det(-\mathbf{C}_{123}) + (\langle 1, 2 \rangle C_3 + \text{cyclic}) + C_{12}C_{23}C_{31} + C_{13}C_{32}C_{21} \right. \\ \left. + \left((1 - s_{12})\mathcal{R}_{(12)}C_3 + \text{cyclic} \right) \right] \text{PT}(W_1), \quad (\text{C.8})$$

where \mathbf{C}_{123} is the 3×3 matrix where the off-diagonal elements are C_{ij} and the diagonal ones are $-C_i$. According to matrix tree theorem [63], $\det(-\mathbf{C}_{123})$ is a combination of labeled trees rooted on W_1 . Actually, $\det(-\mathbf{C}_{123})$ is the part of $\mathcal{T}_{W_1}(1, 2, 3)$ with $\epsilon \cdot \epsilon$ absent,

$$\det(-\mathbf{C}_{123}) = \mathcal{T}_{W_1}(1, 2, 3) \Big|_{\epsilon \cdot \epsilon \rightarrow 0}. \quad (\text{C.9})$$

Equivalently, they are the reference order independent part of \mathcal{T}_{W_1} . Therefore, they are already part of the final recursive expansion. Similarly, $\langle 1, 2 \rangle C_3$ and its cyclic are the ingredient of the recursive expansion, while $C_{12}C_{23}C_{31} + C_{13}C_{32}C_{21}$ is the part of $\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle$ with $\epsilon \cdot \epsilon$ absent according to Eq. (C.2).

All that left to be done is to perform IBP on $(1 - s_{12})\mathcal{R}_{(12)}C_3$ and its cyclic to get the missing $\epsilon \cdot \epsilon$ pieces. For this simple example, Eq. (B.5) is adequate since only tadpoles appear after expanding C_i . We proceed with the reference order $1 \prec 2 \prec 3$.

For $(1 - s_{12})\mathcal{R}_{(12)}C_3$, we set $a = 1$ according to (g1) and break the subcycle $\text{PT}(1, 2)$ using Eq. (B.5). Note that $C_3 = C_{31} + C_{32} + \sum_{j \in W_1} C_{3j}$. For C_{31} , we have

$$(1 - s_{12})\mathcal{R}_{(12)}C_{31}\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} - \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}} C_{31} \left(\frac{k_2 \cdot k_3}{z_{23}} + \sum_{j \in W_1} \frac{k_2 \cdot k_j}{z_{2j}} \right) \text{PT}(W_1), \quad (\text{C.10})$$

where the first term contains a subcycle $\text{PT}(1, 2, 3)$, which contributes to $\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle$. The second term contributes to the tree $\{ \underline{1 \rightarrow 2} W_1, \underline{3 \rightarrow 1} \}$ in $\mathcal{T}_{W_1}(1, 2, 3)$. According to the rule (A.7), the path $\underline{1 \rightarrow 2} W_1$ has the kinematic factor $\epsilon_1 \cdot f_2 \cdot k_j = \epsilon_1 \cdot k_2 \epsilon_2 \cdot k_j - \epsilon_1 \cdot \epsilon_2 k_2 \cdot k_j$ with $j \in W_1$. Indeed, the first part $\epsilon_1 \cdot k_2 \epsilon_2 \cdot k_j$ is given by Eq. (C.9), while the second part is supplemented by Eq. (C.10). Similarly, for C_{32} , we have

$$(1 - s_{12})\mathcal{R}_{(12)}C_{32}\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} - \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}} C_{32} \left[\frac{k_3 \cdot k_1}{z_{31}} + \sum_{j \in W_1} \left(\frac{k_2 \cdot k_j}{z_{2j}} + \frac{k_3 \cdot k_j}{z_{3j}} \right) \right] \text{PT}(W_1), \quad (\text{C.11})$$

where the first term contributes to $\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle$, and second term contributes to the labeled tree $\{ \underline{1 \rightarrow 2} W_1, \underline{3 \rightarrow 2} \}$ and $\{ \underline{1 \rightarrow 2 \rightarrow 3} W_1 \}$. Finally, for $\sum_{j \in W_1} C_{3j}$, we have

$$(1 - s_{1,2})\mathcal{R}_{(12)} \sum_{j \in W_1} C_{3,j} \text{PT}(W_1) \stackrel{\text{IBP}}{\cong} - \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}} \sum_{j \in W_1} C_{3,j} \sum_{\ell \in W_1 \cup \{3\}} \frac{k_2 \cdot k_\ell}{z_{2\ell}} \text{PT}(W_1), \quad (\text{C.12})$$

which contributes to the labeled tree $\{ \underline{1 \rightarrow 2} W_1, \underline{3} W_1 \}$ and $\{ \underline{1 \rightarrow 2 \rightarrow 3} W_1 \}$.

Next, $(1 - s_{1,3})\mathcal{R}_{(13)}C_2$ can be processed similarly by Eq. (B.5) with $a = 1$,

$$(1 - s_{1,3})\mathcal{R}_{(13)}C_2\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} -\frac{\epsilon_1 \cdot \epsilon_3}{z_{13}}\text{PT}(W_1) \left[C_{21} \frac{k_3 \cdot k_2}{z_{32}} + C_{23} \frac{k_2 \cdot k_1}{z_{21}} \right. \\ \left. + C_2 \sum_{j \in W_1} \frac{k_3 \cdot k_j}{z_{3j}} + C_{23} \sum_{j \in W_1} \frac{k_2 \cdot k_j}{z_{2j}} + \frac{k_3 \cdot k_2}{z_{32}} \sum_{j \in W_1} C_{2,j} \right], \quad (\text{C.13})$$

where the first line contributes to $\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle$ and the second line contributes to the labeled tree $\{ \underline{1 \rightarrow 3 \rightarrow W_1}, \underline{2 \rightarrow 1} \}$, $\{ \underline{1 \rightarrow 3 \rightarrow W_1}, \underline{2 \rightarrow 3} \}$, $\{ \underline{1 \rightarrow 3 \rightarrow W_1}, \underline{2 \rightarrow W_1} \}$ and $\{ \underline{1 \rightarrow 3 \rightarrow 2 \rightarrow W_1} \}$.

Finally, for $(1 - s_{23})\mathcal{R}_{(23)}C_1 = (1 - s_{23})\mathcal{R}_{(23)}(C_{12} + C_{13} + \sum_{j \in W_1} C_{1j})$, we invoke (g2) to set $a = 2$ and 3 for the first two terms. In the last term, since 1 is connected to the root, we move on to the next particle in the reference order, which is 2, according to (g3). For the first two terms, using (B.5) with the proper a , we get

$$(1 - s_{23})\mathcal{R}_{(23)}C_{13}\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} -C_{13} \frac{\epsilon_2 \cdot \epsilon_3}{z_{23}} \left(\frac{k_2 \cdot k_1}{z_{21}} + \sum_{j \in W_1} \frac{k_2 \cdot k_j}{z_{2j}} \right) \text{PT}(W_1), \\ (1 - s_{23})\mathcal{R}_{(23)}C_{12}\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} -C_{12} \frac{\epsilon_2 \cdot \epsilon_3}{z_{23}} \left(\frac{k_3 \cdot k_1}{z_{31}} + \sum_{j \in W_1} \frac{k_3 \cdot k_j}{z_{3j}} \right) \text{PT}(W_1). \quad (\text{C.14})$$

In both results, the first term contributes to $\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle$, while the second term contributes to the labeled tree $\{ \underline{1 \rightarrow 3 \rightarrow 2 \rightarrow W_1} \}$ and $\{ \underline{1 \rightarrow 2 \rightarrow 3 \rightarrow W_1} \}$ respectively. For the $\sum_j C_{1j}$ part, we set $a = 2$ to break the subcycle $\text{PT}(2, 3)$ because of the rule (g1),

$$(1 - s_{23})\mathcal{R}_{(23)} \sum_{j \in W_1} C_{1j}\text{PT}(W_1) \stackrel{\text{IBP}}{\cong} -\frac{\epsilon_2 \cdot \epsilon_3}{z_{23}} \sum_{j \in W_1} C_{1j} \sum_{\ell \in W_1 \cup \{1\}} \frac{k_3 \cdot k_\ell}{z_{3\ell}} \text{PT}(W_1). \quad (\text{C.15})$$

We see that this choice of a , the result contributes to the labeled tree $\{ \underline{2 \rightarrow 3 \rightarrow W_1}, \underline{1 \rightarrow W_1} \}$ and $\{ \underline{2 \rightarrow 3 \rightarrow 1 \rightarrow W_1} \}$.

The six terms containing subcycle $\text{PT}(1, 2, 3)$ in Eq. (C.10), (C.11), (C.13) and (C.14) are the exact the remaining components to make up $\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle$ together with $C_{12}C_{23}C_{31} + C_{13}C_{32}C_{21}$. According to the definition of fusion (2.7), we have

$$\langle 1, 2, 3 \rangle + \langle 1, 3, 2 \rangle = C_{12}C_{23}C_{31} + C_{13}C_{32}C_{21} - \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}} \left[C_{31} \frac{k_2 \cdot k_3}{z_{23}} + C_{32} \frac{k_3 \cdot k_1}{z_{31}} \right] \\ - \frac{\epsilon_1 \cdot \epsilon_3}{z_{13}} \left[C_{21} \frac{k_3 \cdot k_2}{z_{32}} + C_{23} \frac{k_2 \cdot k_1}{z_{21}} \right] - \frac{\epsilon_2 \cdot \epsilon_3}{z_{23}} \left[C_{13} \frac{k_2 \cdot k_1}{z_{21}} + C_{12} \frac{k_3 \cdot k_1}{z_{31}} \right]. \quad (\text{C.16})$$

Meanwhile, summing over all the aforementioned labeled trees, we exactly reproduce

$\mathcal{T}_{W_1}(1, 2, 3)$ constructed from the rules in Appendix A. The explicit expression is

$$\begin{aligned} \mathcal{T}_{W_1}(1, 2, 3) = & \det(-\mathbf{C}_{123}) - \frac{\epsilon_1 \cdot \epsilon_2}{z_{12}} \left[C_3 \sum_{j \in W_1} \frac{k_2 \cdot k_j}{z_{2j}} + C_{32} \sum_{j \in W_1} \frac{k_3 \cdot k_j}{z_{3j}} + \frac{k_2 \cdot k_3}{z_{23}} \sum_{j \in W_1} C_{3,j} \right] \\ & - \frac{\epsilon_1 \cdot \epsilon_3}{z_{13}} \left[C_2 \sum_{j \in W_1} \frac{k_3 \cdot k_j}{z_{3j}} + C_{23} \sum_{j \in W_1} \frac{k_2 \cdot k_j}{z_{2j}} + \frac{k_3 \cdot k_2}{z_{32}} \sum_{j \in W_1} C_{2,j} \right] \\ & - \frac{\epsilon_2 \cdot \epsilon_3}{z_{23}} \left[\sum_{\ell_1 \in W_1} C_{1\ell_1} \sum_{j \in W_1 \cup \{1\}} \frac{k_3 \cdot k_j}{z_{3j}} + C_{13} \sum_{j \in W_1} \frac{k_2 \cdot k_j}{z_{2j}} + C_{12} \sum_{j \in W_1} \frac{k_3 \cdot k_j}{z_{3j}} \right]. \quad (\text{C.17}) \end{aligned}$$

This finishes the derivation of the recursive expansion (C.6) for three gluons and one trace from the the string integrand (1.6).

References

- [1] S. He, F. Teng and Y. Zhang, *String amplitudes from field-theory amplitudes and vice versa*, *Phys. Rev. Lett.* **122** (2019) 211603, [[1812.03369](#)].
- [2] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, A. B. Goncharov, A. Postnikov and J. Trnka, *Grassmannian Geometry of Scattering Amplitudes*. Cambridge University Press, 2016, [10.1017/CBO9781316091548](#).
- [3] J. M. Henn and J. C. Plefka, *Scattering Amplitudes in Gauge Theories*, *Lect. Notes Phys.* **883** (2014) pp.1–195.
- [4] H. Elvang and Y.-t. Huang, *Scattering Amplitudes in Gauge Theory and Gravity*. Cambridge University Press, 2015.
- [5] H. Kawai, D. C. Lewellen and S. H. H. Tye, *A Relation Between Tree Amplitudes of Closed and Open Strings*, *Nucl. Phys.* **B269** (1986) 1–23.
- [6] Z. Bern, L. J. Dixon, M. Perelstein and J. S. Rozowsky, *Multileg one loop gravity amplitudes from gauge theory*, *Nucl. Phys.* **B546** (1999) 423–479, [[hep-th/9811140](#)].
- [7] Z. Bern, J. J. M. Carrasco and H. Johansson, *New Relations for Gauge-Theory Amplitudes*, *Phys. Rev.* **D78** (2008) 085011, [[0805.3993](#)].
- [8] Z. Bern, J. J. M. Carrasco and H. Johansson, *Perturbative Quantum Gravity as a Double Copy of Gauge Theory*, *Phys. Rev. Lett.* **105** (2010) 061602, [[1004.0476](#)].
- [9] Z. Bern, S. Davies, T. Dennen, A. V. Smirnov and V. A. Smirnov, *Ultraviolet Properties of $N=4$ Supergravity at Four Loops*, *Phys. Rev. Lett.* **111** (2013) 231302, [[1309.2498](#)].
- [10] Z. Bern, S. Davies and T. Dennen, *Enhanced ultraviolet cancellations in $\mathcal{N} = 5$ supergravity at four loops*, *Phys. Rev.* **D90** (2014) 105011, [[1409.3089](#)].
- [11] Johansson, Henrik and Kälin, Gregor and Mogull, Gustav, *Two-loop supersymmetric QCD and half-maximal supergravity amplitudes*, *JHEP* **09** (2017) 019, [[1706.09381](#)].
- [12] Z. Bern, J. J. Carrasco, W.-M. Chen, A. Edison, H. Johansson, J. Parra-Martinez et al., *Ultraviolet Properties of $\mathcal{N} = 8$ Supergravity at Five Loops*, *Phys. Rev.* **D98** (2018) 086021, [[1804.09311](#)].

- [13] F. Cachazo, S. He and E. Y. Yuan, *Scattering of Massless Particles in Arbitrary Dimensions*, *Phys. Rev. Lett.* **113** (2014) 171601, [[1307.2199](#)].
- [14] F. Cachazo, S. He and E. Y. Yuan, *Scattering of Massless Particles: Scalars, Gluons and Gravitons*, *JHEP* **07** (2014) 033, [[1309.0885](#)].
- [15] F. Cachazo, S. He and E. Y. Yuan, *Scattering equations and Kawai-Lewellen-Tye orthogonality*, *Phys. Rev.* **D90** (2014) 065001, [[1306.6575](#)].
- [16] T. Adamo, E. Casali and D. Skinner, *Ambitwistor strings and the scattering equations at one loop*, *JHEP* **04** (2014) 104, [[1312.3828](#)].
- [17] Y. Geyer, L. Mason, R. Monteiro and P. Tourkine, *Loop Integrands for Scattering Amplitudes from the Riemann Sphere*, *Phys. Rev. Lett.* **115** (2015) 121603, [[1507.00321](#)].
- [18] F. Cachazo, S. He and E. Y. Yuan, *One-Loop Corrections from Higher Dimensional Tree Amplitudes*, *JHEP* **08** (2016) 008, [[1512.05001](#)].
- [19] Y. Geyer, L. Mason, R. Monteiro and P. Tourkine, *Two-Loop Scattering Amplitudes from the Riemann Sphere*, *Phys. Rev.* **D94** (2016) 125029, [[1607.08887](#)].
- [20] Y. Geyer and R. Monteiro, *Two-Loop Scattering Amplitudes from Ambitwistor Strings: from Genus Two to the Nodal Riemann Sphere*, *JHEP* **11** (2018) 008, [[1805.05344](#)].
- [21] F. Cachazo, S. He and E. Y. Yuan, *Scattering Equations and Matrices: From Einstein To Yang-Mills, DBI and NLSM*, *JHEP* **07** (2015) 149, [[1412.3479](#)].
- [22] S. He and O. Schlotterer, *New Relations for Gauge-Theory and Gravity Amplitudes at Loop Level*, *Phys. Rev. Lett.* **118** (2017) 161601, [[1612.00417](#)].
- [23] S. He, O. Schlotterer and Y. Zhang, *New BCJ representations for one-loop amplitudes in gauge theories and gravity*, *Nucl. Phys.* **B930** (2018) 328–383, [[1706.00640](#)].
- [24] L. Mason and D. Skinner, *Ambitwistor strings and the scattering equations*, *JHEP* **07** (2014) 048, [[1311.2564](#)].
- [25] E. Casali, Y. Geyer, L. Mason, R. Monteiro and K. A. Roehrig, *New Ambitwistor String Theories*, *JHEP* **11** (2015) 038, [[1506.08771](#)].
- [26] W. Siegel, *Amplitudes for left-handed strings*, [[1512.02569](#)].
- [27] E. Casali and P. Tourkine, *On the null origin of the ambitwistor string*, *JHEP* **11** (2016) 036, [[1606.05636](#)].
- [28] T. Azevedo and R. L. Jusinskis, *Connecting the ambitwistor and the sectorized heterotic strings*, *JHEP* **10** (2017) 216, [[1707.08840](#)].
- [29] S. Mizera, *Scattering Amplitudes from Intersection Theory*, *Phys. Rev. Lett.* **120** (2018) 141602, [[1711.00469](#)].
- [30] N. E. J. Bjerrum-Bohr, P. H. Damgaard, P. Tourkine and P. Vanhove, *Scattering Equations and String Theory Amplitudes*, *Phys. Rev.* **D90** (2014) 106002, [[1403.4553](#)].
- [31] C. R. Mafra, O. Schlotterer and S. Stieberger, *Explicit BCJ Numerators from Pure Spinors*, *JHEP* **07** (2011) 092, [[1104.5224](#)].
- [32] C. R. Mafra, O. Schlotterer and S. Stieberger, *Complete N-Point Superstring Disk Amplitude I. Pure Spinor Computation*, *Nucl. Phys.* **B873** (2013) 419–460, [[1106.2645](#)].

- [33] C. R. Mafra, O. Schlotterer and S. Stieberger, *Complete N -Point Superstring Disk Amplitude II. Amplitude and Hypergeometric Function Structure*, *Nucl. Phys.* **B873** (2013) 461–513, [[1106.2646](#)].
- [34] C. R. Mafra and O. Schlotterer, *Towards one-loop SYM amplitudes from the pure spinor BRST cohomology*, *Fortsch. Phys.* **63** (2015) 105–131, [[1410.0668](#)].
- [35] S. He, R. Monteiro and O. Schlotterer, *String-inspired BCJ numerators for one-loop MHV amplitudes*, *JHEP* **01** (2016) 171, [[1507.06288](#)].
- [36] M. Chiodaroli, M. Günaydin, H. Johansson and R. Roiban, *Scattering amplitudes in $\mathcal{N} = 2$ Maxwell-Einstein and Yang-Mills/Einstein supergravity*, *JHEP* **01** (2015) 081, [[1408.0764](#)].
- [37] N. E. J. Bjerrum-Bohr, P. H. Damgaard and P. Vanhove, *Minimal Basis for Gauge Theory Amplitudes*, *Phys. Rev. Lett.* **103** (2009) 161602, [[0907.1425](#)].
- [38] S. Stieberger, *Open & Closed vs. Pure Open String Disk Amplitudes*, [[0907.2211](#)].
- [39] S. Stieberger and T. R. Taylor, *New relations for Einstein–Yang–Mills amplitudes*, *Nucl. Phys.* **B913** (2016) 151–162, [[1606.09616](#)].
- [40] O. Schlotterer, *Amplitude relations in heterotic string theory and Einstein–Yang–Mills*, *JHEP* **11** (2016) 074, [[1608.00130](#)].
- [41] J. Broedel, O. Schlotterer and S. Stieberger, *Polylogarithms, Multiple Zeta Values and Superstring Amplitudes*, *Fortsch. Phys.* **61** (2013) 812–870, [[1304.7267](#)].
- [42] J. J. M. Carrasco, C. R. Mafra and O. Schlotterer, *Abelian Z -theory: NLSM amplitudes and α' -corrections from the open string*, *JHEP* **06** (2017) 093, [[1608.02569](#)].
- [43] C. R. Mafra and O. Schlotterer, *Non-abelian Z -theory: Berends–Giele recursion for the α' -expansion of disk integrals*, *JHEP* **01** (2017) 031, [[1609.07078](#)].
- [44] J. J. M. Carrasco, C. R. Mafra and O. Schlotterer, *Semi-abelian Z -theory: NLSM+ ϕ^3 from the open string*, *JHEP* **08** (2017) 135, [[1612.06446](#)].
- [45] Y.-t. Huang, O. Schlotterer and C. Wen, *Universality in string interactions*, *JHEP* **09** (2016) 155, [[1602.01674](#)].
- [46] T. Azevedo, M. Chiodaroli, H. Johansson and O. Schlotterer, *Heterotic and bosonic string amplitudes via field theory*, *JHEP* **10** (2018) 012, [[1803.05452](#)].
- [47] H. Johansson and J. Nohle, *Conformal Gravity from Gauge Theory*, [[1707.02965](#)].
- [48] O. Schlotterer and S. Stieberger, *Motivic Multiple Zeta Values and Superstring Amplitudes*, *J. Phys.* **A46** (2013) 475401, [[1205.1516](#)].
- [49] S. Stieberger, *Closed superstring amplitudes, single-valued multiple zeta values and the Deligne associator*, *J. Phys.* **A47** (2014) 155401, [[1310.3259](#)].
- [50] S. Stieberger and T. R. Taylor, *Closed String Amplitudes as Single-Valued Open String Amplitudes*, *Nucl. Phys.* **B881** (2014) 269–287, [[1401.1218](#)].
- [51] O. Schlotterer and O. Schnetz, *Closed strings as single-valued open strings: A genus-zero derivation*, *J. Phys.* **A52** (2019) 045401, [[1808.00713](#)].
- [52] F. Brown and C. Dupont, *Single-valued integration and superstring amplitudes in genus zero*, [[1810.07682](#)].
- [53] O. Schnetz, *Graphical functions and single-valued multiple polylogarithms*, *Commun. Num. Theor. Phys.* **08** (2014) 589–675, [[1302.6445](#)].

- [54] F. Brown, *Single-valued Motivic Periods and Multiple Zeta Values*, [SIGMA 2 \(2014\) e25](#), [[1309.5309](#)].
- [55] N. E. J. Bjerrum-Bohr, P. H. Damgaard, B. Feng and T. Sondergaard, *Proof of Gravity and Yang-Mills Amplitude Relations*, [JHEP 09 \(2010\) 067](#), [[1007.3111](#)].
- [56] S. Mizera and G. Zhang, *A String Deformation of the Parke-Taylor Factor*, [Phys. Rev. D96 \(2017\) 066016](#), [[1705.10323](#)].
- [57] S. Mizera, *Combinatorics and Topology of Kawai-Lewellen-Tye Relations*, [JHEP 08 \(2017\) 097](#), [[1706.08527](#)].
- [58] N. Arkani-Hamed, Y. Bai and T. Lam, *Positive Geometries and Canonical Forms*, [JHEP 11 \(2017\) 039](#), [[1703.04541](#)].
- [59] N. Arkani-Hamed, Y. Bai, S. He and G. Yan, *Scattering Forms and the Positive Geometry of Kinematics, Color and the Worldsheet*, [JHEP 05 \(2018\) 096](#), [[1711.09102](#)].
- [60] F. Teng and B. Feng, *Expanding Einstein-Yang-Mills by Yang-Mills in CHY frame*, [JHEP 05 \(2017\) 075](#), [[1703.01269](#)].
- [61] Y.-J. Du, B. Feng and F. Teng, *Expansion of All Multitrace Tree Level EYM Amplitudes*, [JHEP 12 \(2017\) 038](#), [[1708.04514](#)].
- [62] C. S. Lam and Y.-P. Yao, *Evaluation of the Cachazo-He-Yuan gauge amplitude*, [Phys. Rev. D93 \(2016\) 105008](#), [[1602.06419](#)].
- [63] R. P. Stanley, *Enumerative Combinatorics: Volume 2*. Cambridge University Press, New York, NY, USA, 1st ed., 2001.
- [64] V. Del Duca, L. J. Dixon and F. Maltoni, *New color decompositions for gauge amplitudes at tree and loop level*, [Nucl. Phys. B571 \(2000\) 51–70](#), [[hep-ph/9910563](#)].
- [65] H. Johansson, G. Mogull and F. Teng, *Unraveling conformal gravity amplitudes*, [JHEP 09 \(2018\) 080](#), [[1806.05124](#)].
- [66] S. He and Y. Zhang, *New Formulas for Amplitudes from Higher-Dimensional Operators*, [JHEP 02 \(2017\) 019](#), [[1608.08448](#)].
- [67] L. M. Garozzo, L. Queimada and O. Schlotterer, *Berends-Giele currents in Bern-Carrasco-Johansson gauge for F^3 - and F^4 -deformed Yang-Mills amplitudes*, [JHEP 02 \(2019\) 078](#), [[1809.08103](#)].
- [68] Y.-J. Du and F. Teng, *BCJ numerators from reduced Pfaffian*, [JHEP 04 \(2017\) 033](#), [[1703.05717](#)].
- [69] Y.-J. Du and Y. Zhang, *Gauge invariance induced relations and the equivalence between distinct approaches to NLSM amplitudes*, [JHEP 07 \(2018\) 177](#), [[1803.01701](#)].
- [70] L. Hou and Y.-J. Du, *A graphic approach to gauge invariance induced identity*, [JHEP 05 \(2019\) 012](#), [[1811.12653](#)].
- [71] S. Mizera, *Aspects of Scattering Amplitudes and Moduli Space Localization*, [[1906.02099](#)].
- [72] P. Vanhove and F. Zerbini, *Closed string amplitudes from single-valued correlation functions*, [[1812.03018](#)].
- [73] S. He, G. Yan, C. Zhang and Y. Zhang, *Scattering Forms, Worldsheet Forms and Amplitudes from Subspaces*, [JHEP 08 \(2018\) 040](#), [[1803.11302](#)].
- [74] N. Arkani-Hamed, *Talk at Amplitudes 2019*, July 2019 at Trinity College, Dublin.

- [75] C. Baadsgaard, N. E. J. Bjerrum-Bohr, J. L. Bourjaily and P. H. Damgaard, *String-Like Dual Models for Scalar Theories*, *JHEP* **12** (2016) 019, [[1610.04228](#)].
- [76] A. Tsuchiya, *More on One Loop Massless Amplitudes of Superstring Theories*, *Phys. Rev.* **D39** (1989) 1626.
- [77] L. Dolan and P. Goddard, *Current Algebra on the Torus*, *Commun. Math. Phys.* **285** (2009) 219–264, [[0710.3743](#)].
- [78] C. R. Mafra and O. Schlotterer, *Double-Copy Structure of One-Loop Open-String Amplitudes*, *Phys. Rev. Lett.* **121** (2018) 011601, [[1711.09104](#)].
- [79] C. R. Mafra and O. Schlotterer, *Towards the n -point one-loop superstring amplitude I: Pure spinors and superfield kinematics*, [1812.10969](#).
- [80] C. R. Mafra and O. Schlotterer, *Towards the n -point one-loop superstring amplitude II: Worldsheet functions and their duality to kinematics*, [1812.10970](#).
- [81] C. R. Mafra and O. Schlotterer, *Towards the n -point one-loop superstring amplitude III: One-loop correlators and their double-copy structure*, [1812.10971](#).
- [82] J. E. Gerken, A. Kleinschmidt and O. Schlotterer, *Heterotic-string amplitudes at one loop: modular graph forms and relations to open strings*, *JHEP* **01** (2019) 052, [[1811.02548](#)].
- [83] X. Gao, S. He and Y. Zhang, *Labelled tree graphs, Feynman diagrams and disk integrals*, *JHEP* **11** (2017) 144, [[1708.08701](#)].