String amplitudes from field-theory amplitudes and \textit{vice versa}

Song He,\textsuperscript{1,2} Fei Teng,\textsuperscript{3} and Yong Zhang\textsuperscript{1,2,}\textsuperscript{*}

\textsuperscript{1}CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China
\textsuperscript{2}School of Physical Sciences, University of Chinese Academy of Sciences, No.19A Yuquan Road, Beijing 100049, China
\textsuperscript{3,}\textsuperscript{3}Department of Physics and Astronomy, Uppsala University, 75108 Uppsala, Sweden

We present an integration-by-parts reduction of any massless tree-level string correlator to an equivalence class of logarithmic functions, which can be used to define a field-theory amplitude via a Cachazo-He-Yuan (CHY) formula. The string amplitude is then shown to be the double copy of the field-theory one and a special disk/sphere integral. The construction is generic as it applies to any correlator that is a rational function of correct SL(2) weight. By applying the reduction to open bosonic/heterotic strings, we get a closed-form CHY integrand for the \((DF)^2 + YM + \phi^3\) theory.

I. INTRODUCTION

Recent years have seen enormous advances in understanding structures of scattering amplitudes in quantum field theories (\textit{c.f.} \textsuperscript{1–3}), and many crucial insights have originated from string theory. The field-theory limit of the celebrated Kawai-Lewellen-Tye (KLT) relations \textsuperscript{4} between tree amplitudes in open and closed string theory gives double copy relations linking gauge-theory amplitudes to gravitational ones \textsuperscript{5}. Based on a remarkable duality between color and kinematics due to Bern, Carrasco and Johansson (BCJ) \textsuperscript{6}, double copy has been extended to quantum level and become the state-of-the-art method for multi-loop calculations in supergravities \textsuperscript{7}. Apart from the original KLT relations, string theory has provided constructions of amplitude representation that respects color-kinematics duality at tree and loop level \textsuperscript{8–12}, and amplitude relations in gauge theory, \textit{i.e.} BCJ relations \textsuperscript{13,14}, and beyond \textsuperscript{15,16}.

More surprisingly, it has been realized that tree-level open- and closed-superstring amplitudes themselves can be obtained via a double copy \textsuperscript{17}. By rearranging the worldsheet correlator with massless states in type-I theory \textsuperscript{9–10}, we can decompose disk amplitudes into field-theory KLT products of super-Yang-Mills (SYM) amplitudes and universal basis of disk integrals called \textit{Z} integrals. The former contain the full polarization information but no nontrivial \(\alpha'\)-dependence, while the latter encode \(\alpha'\) expansion and are interpreted as amplitudes in an effective field theory dubbed \(Z\) theory \textsuperscript{18–20}. Very recently, this has been generalized to bosonic and heterotic strings \textsuperscript{21,22}: now the field-theory amplitudes also contain tachyon poles, and they were conjectured to come from the \((DF)^2 + YM + \phi^3\) Lagrangian, with \(\alpha'\) related to its mass parameter \textsuperscript{23}. With \(Z\) integrals replaced by certain sphere integrals, we can get closed-string amplitudes via a double copy from these field theories \textsuperscript{24}.

An alternative framework that manifests and extends the double copy in field theories is the Cachazo-He-Yuan (CHY) formulation \textsuperscript{25,26}, which expresses tree amplitudes in a large class of massless theories based on scattering equations \textsuperscript{27}. Together with loop-level generalizations \textsuperscript{28–32}, these have led to new double-copy realizations of various theories \textsuperscript{33}, and one-loop extensions of KLT and amplitude relations \textsuperscript{34,35}. It is now well established that tree and loop CHY formulas can be derived from worldsheet models known as ambitwistor string theory \textsuperscript{36,37}, where CHY integrand is given by correlators therein. Despite significant progress \textsuperscript{38,42}, it is fair to say that the origin of ambitwistor strings, and the relations to strings at finite \(\alpha'\) remain mysterious. A related outstanding question is how such double copies for strings can be understood via ambitwistor strings, especially for bosonic and heterotic strings which would require CHY formulas with nontrivial \(\alpha'\) dependence.

In this letter, we present a double-copy construction for any massless string amplitude in terms of field-theory amplitudes defined by a CHY formula, where the integrand can be directly obtained from the original string correlator. We show algorithmically how to reduce a completely general rational function with correct SL(2) weight (which we refer to as a string correlator), via integration by parts (IBP) to an equivalent class of logarithmic functions, and any of them serves as a CHY half-integrand that gives desired field-theory amplitudes. Once the logarithmic class is reached, the next step is to use scattering equations (SE) to obtain equivalent half-integrands, which are no longer logarithmic but usually take a more compact form and make some properties more manifest. The same holds for closed string correlators where we reduce the holomorphic part by IBP.

As the main application illustrating the power of this method, we perform IBP for all scalar-gluon correlators of compactified open bosonic strings, or equivalently the holomorphic part in heterotic strings, to produce logarithmic functions as CHY half-integrands. By using SE, we further present a remarkably simple formula for the half-integrand to any multiplicities which gives amplitudes for the single-trace sector of the \((DF)^2 + YM + \phi^3\) theory. The formula for multitrace amplitudes and pure gluon ones are contained in our result as well, thus it gives CHY representations for all tree amplitudes in this theory.
II. CHY INTEGRAND FROM STRING CORRELATOR

The double copy construction at tree-level can be most conveniently expressed by KLT product of color-ordered amplitudes: $M_C = M_A \otimes M_B$ is defined as
\begin{equation}
M_C = \sum_{\alpha, \beta \in S_{n-3}} M_A(\alpha) S[\alpha | \beta] M_B(\beta),
\end{equation}
where the $(n-3)!$ dimensional matrix $S[\alpha | \beta]$ is the so-called momentum kernel [43] for orderings $\alpha, \beta$ in a minimal basis. Instead of writing $S$ explicitly, it suffices to say that it is the inverse of the matrix formed by double-partial amplitudes in bi-adjoint $\phi^3$ theory: $S = m^{-1}$ [26], for any choice of basis such that $m$ is invertible.

A generic massless open-string tree amplitude is given by a disk integral with $\rho \in S_n/Z_n$ denoting the ordering for integration domain $z_{\rho(i)} < z_{\rho(i+1)}$:
\begin{equation}
M_{\text{string}}^{\text{string}}(\rho) = \int_{\rho} \frac{d^n z}{\text{vol SL}(2, \mathbb{R})} \prod_{i < j} |z_{ij}|^{s_{ij}} \mathcal{I}_n(z),
\end{equation}
where $s_{ij} := \alpha' k_i \cdot k_j$, $z_{ij} := z_i - z_j$; one can fix three punctures, e.g. $(z_1, z_{n-1}, z_n) = (0, 1, \infty)$, using SL(2, R) redundancy, and the product in the Koba-Nielsen (KN) factor is over $1 \leq i < j \leq n-1$ with this fixing. After stripping it off, the (reduced) string correlator $\mathcal{I}_n$ is a rational function of $z$’s which depends on details of vertex operators. The only requirement here is that it has the correct SL(2) weight: with $z_n \to \frac{\alpha' \beta z_n}{\gamma + \delta z_n}$ and $\alpha \delta - \beta \gamma = 1$, it must transform as $\mathcal{I}_n \to \prod_{i=1}^n (\gamma + \delta z_i)^2 \mathcal{I}_n$.

A. General claim

Our general claim, as summarized in Fig. 1, is that $M_{\text{string}}^{\text{string}}(\rho)$ is the double copy of $Z$ integrals (disk integral with a Parke-Taylor (PT) factor [17])
\begin{equation}
Z_\rho(\pi) = \int_{\rho} \frac{d\mu_{\text{string}}^{\text{string}}}{z_{\pi_1 \pi_2} z_{\pi_2 \pi_3} \cdots z_{\pi_n \pi_1}} := \int_{\rho} d\mu_{\text{string}}^{\text{PT}}(\pi)
\end{equation}
with color-ordered amplitudes in a field theory. The latter is defined by a CHY formula, with integrands given by a PT factor as defined above, and a logarithmic function $\mathcal{I}_n'$ that can be obtained from $\mathcal{I}_n$:
\begin{equation}
M_{\text{PT}}^{\text{n}}(\rho) := \int_{\rho} \frac{d^n z}{\text{vol SL}(2, \mathbb{C})} \prod_i \delta(i) \sum_{j \neq i} s_{ij} \mathcal{I}_n'(z).\end{equation}

Here the integrals are localized by $n-3$ delta functions imposing scattering equations [25] [27]. Regardless of the form of $\mathcal{I}_n$, by repeated use of IBP, we will present an algorithm to reduce it to a logarithmic function, which is defined to have only logarithmic singularities on boundaries of the moduli space of $n$-punctured Riemann spheres; equivalently, such a function can be written as a linear combination of PT factors [44] [46].

The idea that string amplitude can be written as a double copy is not new, which has been realized in various theories [9] [10] [21] [22] and expected to hold for generic $\mathcal{I}_n$. This can be shown using intersection theory, where $\mathcal{I}_n'$ is obtained from $\mathcal{I}_n$ by computing intersection numbers [47]. In this work we will give a concrete, streamlined IBP algorithm which allows us to reduce a generic $\mathcal{I}_n$ explicitly to a class of logarithmic functions; the algorithm gives $\mathcal{I}_n'$ that is not in a minimal basis of PT factors which makes further simplifications possible.

We actually have an equivalence class of logarithmic functions denoted as $\mathcal{I}_n^{\text{IBP}} \cong \mathcal{I}_n$: any $\mathcal{I}_n'$ gives the same string integral as that of $\mathcal{I}_n$. As we will show shortly, $\mathcal{I}_n'$ gives exactly $M_{\text{PT}}^{\text{n}}$ needed for the double copy, via CHY formula with a Parke-Taylor factor; any $\mathcal{I}_n'$ in the class gives the same result since they are also in an equivalent class on the support of SE. While the explicit form of $\mathcal{I}_n'$ gets more complicated as $n$ grows, we emphasize that it can usually be simplified greatly by use of SE (see Fig. 1). The manipulation from $\mathcal{I}_n'$ to $\mathcal{I}_n^{\text{SYM}}$ is in a sense the reversed one as going from $\mathcal{I}_n$ to $\mathcal{I}_n'$: while being non-logarithmic, usually $\mathcal{I}_n'$ allows an all-multiplicity expression that makes other properties, such as gauge invariance, manifest. The primary example of this double copy is “type-I = $Z \otimes \text{SYM}$” [9] [10], where the class of logarithmic functions from type-I correlator can be used to produce SYM amplitudes. It is now well known that the CHY half-integrand for external gluons, which takes a remarkably simple form $\mathcal{I}_n'' = \text{Pf}^2 \Psi_n$ [48], is equivalent to $\mathcal{I}_n'$ on the support of SE.

Last but not least, one can generalize the double copy to closed-string amplitudes by replacing $Z$ integrals by certain sphere integrals. As conjectured in [24] [49] [50] and proven in [51] [52], the latter can be obtained as the single-valued (sv) projection [53] [54] of open-string amplitudes, and e.g., we have “type-II = sv(type-I) \otimes \text{SYM}”. All of our discussions hold for such double copies where IBP is applied to the holomorphic part of
closed-string correlator $\mathcal{I}_n(z)$ (with the antiholomorphic part untouched): schematically we write \( \mathcal{M}^\text{closed}_n = \mathcal{M}^\text{PT}_n \otimes \text{sv}(\mathcal{M}^\text{open}_n) \), with the CHY half-integrand for $\mathcal{M}^\text{PT}_n$ obtained from the holomorphic correlator, $\mathcal{I}^\text{IBP}_n \cong \mathcal{I}_n$.

B. Derivation: IBP reduction and double copy

We can conveniently represent any rational functions involving only $z_{ij}$’s by a graph: we denote each $z_{ij}$ in the denominator (numerator) by a solid (dashed) line directed from node $i$ to $j$. With the gauge fixing, e.g. $z_n \to \infty$, all the labeled trees are logarithmic functions since we can write them as a sum of PT factors \([55]\).

The derivation of our general claim consists of two steps. We first give a method for reducing any string correlator to logarithmic functions via IBP relations. Next, we show that such logarithmic functions are precisely the CHY integrands for those field theories that enter the string-amplitude double copy mentioned above.

a. IBP reduction

We start with a simple example $\mathcal{I}_4 = \frac{1}{z_{12}^7 z_{34}^{z_2}}$, which, after we gauge fix $(z_1, z_3, z_4) = (0, 1, \infty)$ and subtract the total derivative $\partial z_2 z_2^{-1}(1-z_2)\gamma$, reduces to a logarithmic function, in this case a PT factor after restoring the SL(2) covariance:

$$\frac{1}{z_{12}^7 z_{34}^{z_2}} \overset{\text{IBP}}{\cong} \frac{1}{1 - s z_{12} z_{23} z_{34} z_{41}} = \frac{t}{1 - s} \text{PT}(1234), \quad (5)$$

where $s = s_{12}$ and $t = s_{23}$. This example represents the simplest scenario where only a single subcycle and a labeled line are left after gauge fixing $z_n \to \infty$. The next simplest case is represented by tadpole graphs, as depicted in Fig. 2a, in which a single labeled line connects to the subcycle. One can reduce them to a sum of labeled trees by recursively using IBP relations \([10]\). For example, the IBP reduction of Fig. 2a will be given in Appendix A.

![Diagram](image)

FIG. 2. Examples of rational functions of $z_{ij}$ with no numerators or connected subcycles after gauge fixing. For this case, gauge fixing always results in a single labeled tree.

By connecting more labeled lines to a tadpole, we get multibranch graphs, as shown in Fig. 2b. If we remove a subgraph including part of the subcycle, the rest of the graph turns into a labeled tree, which can be written as a sum of labeled lines through algebraic operations \([55]\). For example, we can view the chain \{4, 1, 2, 3, 5\} in Fig. 2b as a labeled tree rooted on 1. Then partial fraction identities lead to

$$\begin{align*}
\mathcal{I}_n + \mathcal{I}_n + \mathcal{I}_n + \mathcal{I}_n + \mathcal{I}_n & = (4) 1 2 3 4 5 \\
\mathcal{I}_n + \mathcal{I}_n + \mathcal{I}_n + \mathcal{I}_n + \mathcal{I}_n & = (5) 1 2 3 4 5 . \quad (6)
\end{align*}$$

In this way, we can turn all the multibranch graphs into tadpoles, and consequently labeled trees via IBP.

The most general graph with no $z_{ij}$ in the numerator or connected subcycles has the form of Fig. 2c consisting of a product of tadpole and multibranch graphs, and a single labeled tree \([55]\). Following the above prescription, we can turn a tadpole or multibranch subgraph into labeled trees planted on the other connected components, reducing the number of subcycles by one. This algorithm allows us to recursively eliminate all the subcycles, and thus the function $\mathcal{I}_n$ we obtain is logarithmic.

As we will see shortly, the above discussion applies to the $\mathcal{I}_n$ of open bosonic and heterotic strings. We defer the treatment of generic rational functions involving numerators and connected subcycles to Appendix A.

b. String double copy

Now we prove that $\mathcal{M}^\text{PT}_n$ given by the $\mathcal{I}_n$ obtained above is indeed the one in the double copy $\mathcal{M}^\text{string}_n = \mathcal{M}^\text{PT}_n \otimes Z_n$. As a logarithmic function, we can write $\mathcal{I}_n = \sum_n N_n \text{PT}(\pi)$, but the coefficients $N_n$ are not unique since the PT factors are not linearly independent. As shown for open strings \([9, 17]\) and CHY integrals \([20]\), they satisfy BCJ relations by IBP and SE equivalence moves respectively. Consequently, any of them can be expressed in terms of the $(n-3)!$ dimensional minimal basis:

$$\text{PT}(\pi) \overset{\text{IBP/SE}}{\cong} \sum_{\beta \in S_{n-3}} \left[ \sum_{\alpha \in S_{n-3}} m(\pi|\alpha) S[\alpha|\beta] \right] \text{PT}(\beta).$$

By using Eq. (2) and (3), we have

$$\mathcal{M}^\text{string}(\rho) = \sum_{\alpha, \beta \in S_{n-3}} \left[ \sum_\pi N_\pi m(\pi|\alpha) S[\alpha|\beta] Z_\rho(\beta) \right]$$

for any $\rho \in S_{n-3}/Z_n$. The sum in the bracket is a field-theory amplitude defined by CHY formula: recalling that $m(\pi|\alpha) = \int d\mu_\alpha^{\text{CHY}} \text{PT}(\pi) \text{PT}(\alpha)$, thus we have

$$\sum_\pi N_\pi m(\pi|\alpha) = \int d\mu_\alpha^{\text{CHY}} \sum_\pi N_\pi \text{PT}(\pi) \text{PT}(\alpha), \quad (7)$$

which gives exactly $\mathcal{M}^\text{PT}_n(\alpha)$ as in Eq. (1). Note that Eq. (7) makes it manifest that $\mathcal{M}^\text{PT}_n(\rho)$ satisfies BCJ relations, which is necessary for a consistent KLT double copy. One can also define $\mathcal{M}^{\text{KLT}}_n$ by the LHS, which can be rewritten in a form that respects color/kinematics duality with BCJ numerators given by linear combinations.
of \( N_n \)’s. However, we find it extremely useful to turn to CHY formula on the RHS: from our explicit result \( I_n \), we can use SE to simplify it to some \( T_n' \), for which it is possible to obtain a closed form.

One can apply the IBP reduction to type-I correlators \( I_n = K_n(z) \) and obtain logarithmic functions, where the coefficients \( N_n \) give the BCJ numerators for SYM \([10]\). For gluons, the standard CHY half-integrand \( T''_n = PT \Psi_n \) can also be reduced via SE to \( T'_n \), giving equivalent BCJ numerators \([51]\). Remarkably, these BCJ numerators are free of any tachyon poles (though they appear in intermediate steps as in above examples), and unlike \( I_n \), both \( T_n \) and \( T_n' \) are homogenous in \( \alpha' \), all of which are of course guaranteed by the worldsheet supersymmetry. Next we will turn to bosonic/heterotic strings, where BCJ numerators and \( T_n' \) contains tachyon poles and nontrivial \( \alpha' \) dependence. We will see that our two-step method again gives a remarkably simple formula for \( T_n' \).

### III. NEW FORMULAS FOR BOSONIC AND HETEROISTIC STRINGS

Let’s first review correlators for (compactified) open bosonic strings and heterotic strings. Since the former is identical to the holomorphic part of the latter, we will focus on the heterotic-string case and everything carries over to the open-string case. By a slight abuse of notation, we denote closed-string measure as \( d\mu = e^{2\pi i m z} \), and the single-trace tree amplitude of \( m \) gravitons and \( r = n - m \) gluons is given by \([10]\)

\[
M^\text{het}_{m,r} = \int |d\mu|^2 PT(\rho) R_m(z) K_n(z),
\]

where \( \rho \in S_r/Z_r \) denotes the gluon color ordering. The holomorphic function \( R_m \) contains half of the graviton vertex operator correlation function, taking the form

\[
R_m(z) = \sum_{(I)\cdots(K)\in S_m} R_{(I)} R_{(J)} \cdots R_{(K)},
\]

where we sum over permutations in \( S_m \), each of which is written as a product of disjoint cycles. The cycle factor \( R_{(I)} \) is given by

\[
R_{(I)} = \sum_{b \neq a} \frac{\epsilon_{ab} \cdot k_b}{z_{ab}} := C_a,
\]

and \( R_{(I)} = 0 \) for \( |I| > 2 \). The antiholomorphic function \( K_n(z) \) is the type-I correlator that encodes vertex operators of gluons and the other half of the gravitons. The generalization to \( t \) color traces is straightforward:

\[
M^\text{het}_{m,r}(\{\rho_t\}) = \int |d\mu|^2 \left[ \prod_{i=1}^t PT(\rho_i) \right] R_m(z) K_n(z),
\]

where \( \rho_t \in S_r/Z_r \) and \( \sum_{i=1}^t r_i = r \).

For the single-trace integrand, we can always gauge fix the gluon \( z_n \to \infty \) such that we only need to break the \( z \)-cycles in \( R_m(z) \) by IBP. The one-graviton integrand is automatically in the logarithmic form since \( R_1 = C_1 \) contains no \( z \)-cycles, i.e., \( I_n(1; \rho) = I'_n(1; \rho) = PT(\rho) C_1 \). For two gravitons, we have \( R_2(z) = C_1 C_2 + \frac{\alpha' z_{12}^2}{2} \), which has only one \( z \)-cycle \( \frac{1}{z_{12}} \). We can break it by the IBP

\[
\frac{1}{z_{12} z_{21}} (\cdots) \equiv \sum_{i=3}^{n-1} \frac{\epsilon_i k_i z_{in}}{1 - s_{12} z_{12} z_{2i}} (\cdots),
\]

which holds when \( (\cdots) \) does not involve \( z_2 \). This leads to the logarithmic function

\[
T'_n(12; \rho) = \left[ \sum_{i=3}^{n-1} \frac{\epsilon_1 k_1 z_{in}}{s_{12} z_{12} z_{2i}} \frac{\epsilon_2 k_2 z_{jn}}{z_{2j} z_{2n}} + \frac{\alpha' \epsilon_1 k_2 \epsilon_2 k_1 - \epsilon_1 \epsilon_2}{\alpha'(1 - s_{12}) z_{12}} \sum_{i=3}^{n-1} \frac{\epsilon_i z_{in}}{s_{12} z_{12} z_{2i}} \right] PT(\rho)
\]

with the \( SL(2) \) covariance restored. As another example, we present the IBP of the three-graviton integrand in Appendix E. Similar calculation can also be found in \([10]\) for a slightly different purpose. While the complexity of \( I_n \) grows quickly with more gravitons, it follows from our algorithm above. Moreover, in the second step, by using SE we can greatly simplify it and obtain \( T'_n \), that is a closed formula for the most generic single-trace integrand.

#### A. The single-trace integrand

Now we present the single-trace CHY half-integrand \( T''_n \), resulted from our two-step reduction of string integrand \([5]\):

\[
T''_n(12 \cdots m; \rho) = PT(\rho) P_m
\]

where \( P_m \) is given by the cycle expansion

\[
P_m = \sum_{(I)\cdots(K)\in S_m} \Psi_{(I)} \Psi_{(J)} \cdots \Psi_{(K)}.
\]

For length one and two cycles, we have \( \Psi_{(a)} = C_a \) and \( \Psi_{(ab)} = -T_{ab} PT(ab) \), where \( T_{ab} \) is given by

\[
T_{ab} := \frac{\text{tr}(ab)}{1 - s_{ab}} := \frac{\text{tr}(f_a f_b)}{2(1 - s_{ab})},
\]

and \( f^\mu_a = k^\mu_a - k^\nu_a k^\mu_a \) is the linearized field strength. For longer cycles \( (|I| \geq 3) \), we have \( \Psi_{(I)} = -\frac{T_I PT(I)}{2} \), where \( T_I \) is recursively generated from \( T_{ab} \) through

\[
T_I = \frac{1}{1 - s_I} \left[ \text{tr}(I) + \sum_{CP} G_{i1}^l G_{i2}^l \cdots G_{iI}^l \right],
\]

where \( \rho_t \in S_r/Z_r \) and \( \sum_{i=1}^t r_i = r \).
with \( \text{tr}(I) := \text{tr}(f_{a_1} f_{a_2} \cdots f_{a_{|I|}}) \) for \( |I| \geq 3 \). In this equation, the summation is over all cyclic partitions (CP) of a cycle \( I \) into words, \( \{I_1, I_2, \ldots, I_p\} \) where the length of each word \( I_k \) is at least two; \( i \) denotes the first element of \( I_i \), which is the label in \( I \) that succeeds the word \( I_{i-1} \). For example, for \( I = \{1, 2, 3, 4\} \), we have four CPs with only one word, 1234, 2341, 3412, 4123, and two more with two words, \{12, 34\} and \{23, 41\}; the sum on the RHS of Eq. (17) thus include terms like \( G_1^{234}, G_3^1G_4^{31} \), etc. Given a word \( A = a_1a_2 \cdots a_s \) and the next label \( b \), the function \( G \) is defined as

\[
G_A^B = \sum_{q=2}^{s-\delta_{a,b}} T_{a_1 a_2 [a_3] \cdots [a_{q-1}, a_q]} V_{a_{q+1} \cdots a_s}^{a_{q+b}}, \tag{18}
\]

where \( V_{a_{q+1} \cdots a_s}^{a_{q+b}} = \alpha' k_{a_1} f_{a_{q+1}} \cdots f_{a_s}, k_b \), and in particular \( V^{a_{q+b}}_{a_{q+1} \cdots a_s} = s_{ab} \). The bracket \( [i \ j] \) stands for the index antisymmetrization, for example, \( T_{ab|c \ d} := T_{abcd} - T_{adbc} \). The summation range \( s-\delta_{a,b} \) ensures that the right hand side of Eq. (17) only contains those \( T \)'s with shorter cycles, since \( b \) agrees with \( a_1 \) only when \( A = I \).

Note that the closed-formula for \( T_n^m \) is still a conjecture, since it is generalized from explicit calculations at low points. However, as we will discuss in great detail in \cite{58}, the formula has passed various nontrivial consistency checks. On the support of SE, we have checked numerically up to eleven points that Eq. \( (14) \) agrees with the logarithmic form \( T_n^m \) resulted from the IBP of the string integrand \cite{8}. We have also checked up to eight points that the gauge amplitude produced by Eq. \( (14) \) agrees with that calculated from the Feynman diagrams of \( (DF)^2 + YM + \phi^3 \) with massless external particles.

Here we list some all-multiplicity checks as well. From Eq. \( (17) \) and \( (18) \), one can see that \( T_I \) satisfies the same cyclic and reflection property as \( \text{tr}(I) \). Gauge invariance is manifest since \( T_I \) is built from \( f \)'s and \( C_n \) is gauge invariant on the support of SE. Moreover, in the limit \( k_m \to 0 \), we have \( T_I \to 0 \) if \( m \in I \), such that \( \mathcal{P}_m \) factorizes as \( \mathcal{P}_m \to C_m \mathcal{P}_{m-1} \), recovering the universal single soft behavior \cite{59} after the CHY integration. Furthermore, as \( \alpha' \to 0 \), all the tachyon poles \( (1-s_f)^{-1} \) drop out, and by induction it is easy to see that all the \( G \)-factors vanish as \( \mathcal{O}(\alpha') \). Consequently, \( \mathcal{P}_m \) reduces to the cycle expansion of a Pfaffian \cite{60}, reproducing the single-trace YM-scalar integrand PT(\( \rho \)) PT(\( \Psi \)) \cite{61}. On the other hand, the \( T_I \) 's vanish as \( \mathcal{O}(\alpha'^{-1}) \) when \( \alpha' \to \infty \), leaving only \( C_n \)'s in \( \mathcal{P}_m \), which lands exactly on the \( (DF)^2 + \phi^3 \) integrand PT(\( \rho \))(\( \prod_{i=1}^n C_{i} \)) \cite{62}.

The most important check is to show that \( (14) \) gives the correct massless factorization properties for the \( (DF)^2 + YM + \phi^3 \) amplitudes \cite{58}. Following \( 26 \), \( dp_{\text{CHY}} \) and PT's split correspondingly in the factorization limit \( y^2 \to 0 \), thus we only need to consider the behavior of \( \mathcal{P}_m \). The scalar channel, which separates the gluons into two sets, is simple: the leading order gives a simple pole \( 1/y^2 \), and \( \mathcal{P}_m \) factorizes nicely since all \( \Psi(I) \)'s that contain gluons in both sets only contribute to higher orders. In other words, the cycle expansion form \( (15) \) trivializes the scalar factorization channel.

The gluon channel is more interesting. For simplicity, we consider the special one that separates the gluons and the scalars. The leading order is now a double pole \( 1/y^2 \) whose coefficient is proportional to \( \mathcal{P}_m \) evaluated on the \( (m+1) \)-point kinematics, while the subleading order provides the simple pole \( 1/y \). The leading order must vanish, and indeed we have checked up to eight points that

\[
\mathcal{P}_m = 0 \quad \text{for } m \text{ and } (m+1) \text{-point kinematics}. \tag{19}
\]

When \( \alpha' = 0 \), this statement is simply that the matrix \( \Psi_m \) has corank two for \( m \)-point kinematics. Remarkably, \( \mathcal{P}_m \) inherits this corank-two property for finite \( \alpha' \) ! One can verify this statement by showing that the coefficient of each tachyon pole structure of \( \mathcal{P}_m \) vanishes on the support of the SEs, as we will present in \cite{58}.

### B. Examples

Now we provide a few examples for our generic single-trace formula. The integrand involving two gravitons is given by PT(\( \rho \)) PT(\( \Psi \)), where

\[
\mathcal{P}_2 = C_1 C_2 + \Psi_{(12)} = C_1 C_2 - \frac{\text{tr}(12) \text{PT}(12)}{1 - s_{12}}, \tag{20}
\]

which agrees with the logarithmic function \( (13) \) after some use of SE. For three gravitons, Eq. \( (15) \) gives

\[
\mathcal{P}_3 = C_1 C_2 C_3 + \left( C_1 \Psi_{(23)} + C_2 \Psi_{(13)} + C_3 \Psi_{(12)} \right) + \Psi_{(123)} + \Psi_{(132)}
\]

\[
= C_1 C_2 C_3 - \left[ C_1 \text{tr}(23) \text{PT}(23) \frac{1}{1 - s_{23}} + \text{cyc}(123) \right]
\]

\[
= - \frac{1}{2} \left[ T_{123} \text{PT}(123) + T_{132} \text{PT}(132) \right], \tag{21}
\]

where, according to Eq. \( (17) \), \( T_{123} \) is given by

\[
T_{123} = \frac{1}{1 - s_{123}} \left[ \text{tr}(123) + G_1^{123} + G_2^{31} + G_3^{132} \right]
\]

\[
= \frac{\text{tr}(123)}{1 - s_{123}} + \alpha' \left[ \frac{\text{tr}(12)(k_2 \cdot f_3 k_1)}{(1 - s_{12})(1 - s_{13})} + \text{cyc}(123) \right]
\]

The explicit derivation of \( \mathcal{P}_3 \) can be found in Appendix \( 15 \). In the four-graviton integrand \( \mathcal{P}_4 \), the only piece that has not appeared in \( \mathcal{P}_2 \) or \( \mathcal{P}_3 \) is \( T_{abcd} \) in the length-four cycle factor \( \Psi_{(abcd)} \), which is given by

\[
T_{1234} = \frac{\text{tr}(1234) + T_{123} V_{34}^{21} + T_{123} V_{41}^{31} + \text{cyc}(1234)}{1 - s_{1234}}
\]

\[
+ T_{123} s_{23} T_{34} s_{41} + T_{234} s_{24} T_{41} s_{12} \left[ \frac{1}{1 - s_{1234}} \right]. \tag{22}
\]

The first example of index antisymmetrization for \( G \) function as shown in Eq. \( (18) \) is

\[
G_1^{12345} = T_{12} V_{345}^{21} + T_{123} V_{45}^{31} + (T_{1234} - T_{1243}) V_{5}^{41}.
\]
It is contained in

\[ T_{12345} = \frac{\text{tr}(12345)}{1 - s_{12345}} + \frac{G_1^{12345} + \text{cyc}(12345)}{1 - s_{12345}} \]

\[ + T_{12825}(T_{34}V^3 + T_{45}V_{51}) + \text{cyc}(12345) \]

that appears in the length-five cycle factor \( \Psi_{(12345)} \).

C. Towards multi-trace integrands

Interestingly, we can get multi-trace integrands by taking certain coefficients of the single-trace integrand \( \mathcal{I}''(h; \rho) \). As the most general formula will be given in \[58\], here we simply present it for some cases. To start with, the double-trace integrand can be obtained by the prescription

\[ \mathcal{I}''_{\rho}(h'; \sigma, \rho) = \left( \text{the coefficient of } \epsilon_{\sigma_1} k_{\sigma_2} \epsilon_{\sigma_3} k_{\sigma_4} \epsilon_{\sigma_5} k_{\sigma_6} \text{ in } \mathcal{I}''_{\rho}(h; \rho) \right), \]

where \( \sigma \subseteq h \) denotes another gluon trace (in addition to \( \rho \) and \( h' = h \backslash \sigma \) denotes the new set of gravitons. The simplest one is the pure-gluon double-trace case already given in \[16, 22\]: \( \mathcal{I}''_{\rho}(\sigma, \rho) = -\text{PT}(\sigma)\text{PT}(\rho)\frac{1}{1-s_{\rho}} \). We also give a closed-form expression for double trace with one additional graviton, denoted by label \( q \):

\[ \mathcal{I}''_{\rho}'(q; \sigma, \rho) = \frac{\text{PT}(\rho)}{1-s_{\rho}} \left[ -s_{\sigma} \text{PT}(\sigma)C_q - \frac{\alpha'}{2(1-s_{\rho})} \right] \]

\[ \times \sum_{\gamma \in \{\sigma, XwY^{T}, \sigma\}} (-1)^{|\gamma|} k_{\sigma}, f_q \cdot k_{\rho} \text{PT}(\gamma, q) \],

where we first sum over all pairs of labels \( \sigma, \sigma' \), and for each choice we read out words \( X \) and \( Y \) by rewriting the cycle \( \sigma \) in the form \( (\sigma_1, X, \sigma_2, Y) \); then the sum is over words \( \gamma \) with endpoints \( \gamma_1 = \sigma \), and \( |\gamma| = \sigma' \), with shuffle of \( X \) and \( Y^{T} \) in between. On the support of SEs, Eq. \([25]\) is invariant under the relabeling of the traces, although this property is not explicit.

Furthermore, we can extract a third gluon trace \( \tau \) from the double-trace integrand \( \mathcal{I}''_{\rho}(h'; \sigma, \rho) \) by the same process of Eq. \([24]\). Remarkably, the pure-gluon triple-trace result also takes a compact form:

\[ \mathcal{I}''_{\rho}(\tau, \sigma, \rho) = \frac{\text{PT}(\rho)}{(1-s_{\tau})(1-s_{\sigma})} \left[ s_{\bar{\tau}} s_{\sigma} \text{PT}(\tau)\text{PT}(\sigma) \right] \]

\[ - \frac{1}{2(1-s_{\rho})} \sum_{\tau, \pi, \tau' \in \tau} \sum_{\sigma, \sigma' \in \sigma} \sum_{\beta \in \{\tau, XwY^{T}, \tau\}} (-1)^{|\tau|+|\gamma|} s_{\tau}, s_{\sigma}, s_{\bar{\tau}} \cdot \text{PT}(\beta, \gamma) \]. \( \tag{26} \)

At the leading order of \( \alpha' \), both Eq. \([25]\) and \([26]\) reproduce the YM-scalar theory in the corresponding multi-trace sectors \([33, 63, 64]\).

We have checked that Eq. \([25]\) and \([26]\) agree with IBP result starting from Eq. \([11]\), on the support of SE. We can get generic multitrace integrands by repeating Eq. \([24]\) which has been numerically checked up to eleven points and five traces with IBP results.

We provide two examples for Eq. \([25]\) and \([26]\), both of which are at six points:

\[ \mathcal{I}''_{6}(1; 23, 456) = -s_{23}\frac{C_{1}\text{PT}(23)\text{PT}(456)}{1-s_{23}} \]

\[ - \alpha'(k_3, f_1, k_2)\text{PT}(123)\text{PT}(456) \]

\[ \frac{1}{(1-s_{23})(1-s_{456})}, \tag{27} \]

\[ \mathcal{I}''_{6}(12, 34, 56) = \text{PT}(56) \left[ \frac{s_{12}s_{34}\text{PT}(12)\text{PT}(34)}{(1-s_{12})(1-s_{34})} \right] \]

\[ - \frac{s_{23}s_{41}\text{PT}(1234) + s_{24}s_{31}\text{PT}(1243)}{(1-s_{12})(1-s_{34})(1-s_{56})} \]. \( \tag{28} \)

IV. CONCLUSION AND DISCUSSION

In this work we have presented an IBP reduction method for correlators of any massless string amplitude, \( \mathcal{I} \), to a class of logarithmic functions, \( \mathcal{I}'' \). The latter can be used as CHY half-integrand (with the other half given by PT) to define field-theory amplitudes, and their double copy with disk/sphere integrals gives the original string amplitude. Although \( \mathcal{I}'' \) generally takes a complicated form, one can use SE to simplify it and even get a closed form to all multiplicities. We have demonstrated this two-step method for open (compactified) bosonic/heterotic strings, and obtained remarkably simple formulas for single-trace and certain multi-trace sectors of \((DF)^2 + YM + \phi^3\) theory. It can bring new insight into this theory, as well as conformal gravity which is obtained from a double copy with YM \([23, 63]\).

Our discussion is completely generic since it only requires \( \mathcal{I} \) to be a rational function with correct SL(2) weight. It can thus be applied beyond those well-studied conventional string theories, for example, to the dual model proposed in Ref. \([66]\). An important open question concerns the application of IBP reduction to string correlators with massive states, and it would be fascinating to see if similar structures exist there.

On the other hand, the CHY half-integrands obtained in this work contain explicit \( \alpha' \) dependence. It will be interesting to understand how they can emerge from certain model of ambitwistor string \([32, 36, 37]\). Some attempts have been made along this direction, and give the correct three-point results \([40]\). Last but not least, it is highly desirable to extend certain aspects of our construction to loop-level string amplitudes/correlators, and see how they are related to field-theory amplitudes/ ambitwistor string correlators.
It is our pleasure to thank Thales Azevedo, Henrik Johansson, Sebastian Mizera and Oliver Schlotterer for inspiring discussions and helpful comments on our draft. 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-SYS014 and Peng Huanwu center under Grant No. 11747601. F.T. is supported by the Knut and Alice Wallenberg Foundation under grant S1/16.

ACKNOWLEDGMENTS

We first provide more details on the IBP reduction of tadpole graphs. Suppose the labeled tree is planted on $z_1$, and the subcycle $PT(12 \cdots m)$ is multiplied by a function independent of $z_2$ to $z_m$, we can break it as

$$PT(12 \cdots m) \stackrel{IBP}{=} \left( \sum_{\ell=2}^{m} \sum_{j=m+1}^{n-1} \sum_{\rho \in X\omega Y \tau} \frac{(-1)^{|Y|}}{1 - s_{12} \cdots m} \right) \cdot \left( \sum_{\rho \in X\omega Y \tau} s_{\ell j} \right).$$

(App. 1)

where $X$ and $Y$ are obtained by matching $(1, 2 \cdots m) = (1, X, \ell, Y)$. The total derivative of $z_1$ is not used to derive this IBP relation (A1), which enable us to recursively include nodes on the labeled tree into the subcycle. For example, Fig. 2a can be reduced as:

\[
\begin{align*}
\text{IBP} & \equiv \frac{s_{24}}{1 - s_{123}} \frac{1}{z_{12}} \frac{1}{z_{23}} \frac{1}{z_{34}} + \frac{s_{25}}{1 - s_{123}} \frac{1}{z_{24}} \frac{1}{z_{13}} \frac{1}{z_{23}} - \frac{s_{14}}{1 - s_{123}} \frac{1}{z_{12}} \frac{1}{z_{23}} \frac{1}{z_{13}} - \frac{s_{15}}{1 - s_{123}} \frac{1}{z_{12}} \frac{1}{z_{13}} \frac{1}{z_{24}} + \frac{s_{26}}{1 - s_{123}} \frac{1}{z_{24}} \frac{1}{z_{13}} \frac{1}{z_{23}}.
\end{align*}
\]

(App. 2)

Therefore, we can turn tadpole graphs into a sum of labeled trees by repeating the IBP relation (A1).

In Sec. II.B, we have proved that those gauge-fixed string correlators without numerators or connected subcycles can be reduced into logarithmic functions via IBP. Now we complete the proof for generic rational function $\mathcal{I}_n$ with correct SL(2) weight by showing that we can always reduce them to the former case. We start with an example $\mathcal{I}_3 = \frac{z_{12} z_{24}}{z_{12} z_{23} z_{34}}$. We can absorb the numerator into the Koba-Nielsen factor as

$$\frac{z_{13} z_{24}}{z_{12} z_{23}} \text{KN} = \frac{1}{z_{12} z_{23}} \frac{z_{13} z_{24}}{z_{12} z_{23}} \text{KN} = \frac{1}{z_{12} z_{23}} \text{KN}' .$$

(App. 3)

Now we can gauge fix $z_4 \rightarrow \infty$ and perform the IBP as in Eq. (5), while the Mandelstam variables in $\text{KN}'$ are shifted as $u \rightarrow u + 1$ and $s \rightarrow s - 1$. We then restore the gauge covariance and release the cross ratio from $\text{KN}'$:

$$\frac{1}{z_{12} z_{23}} \text{KN}' \stackrel{IBP}{=} \frac{t}{1 - (s - 1)} \frac{1}{z_{12} z_{23}} \text{PN}(1234) \frac{t}{s - (s - 1)} \text{KN}.'$$

(App. 4)

The product in the bracket can be rearranged into a sum of rational functions without $z_{ij}$ in the numerators:

$$\text{PN}(1234) \frac{z_{13} z_{24}}{z_{12} z_{23}} = \frac{z_{12} + z_{23}}{z_{12} z_{23}} \frac{z_{24}}{z_{12} z_{23}} = \text{PN}(1234) - \text{PT}(12) \text{PT}(34) ,$$

(App. 5)

Although the second term is not a logarithmic function, we have reduced the problem to the one already solved in Eq. (5): by using the IBP relation (5) again, we obtain the final logarithmic function:

$$\frac{z_{13} z_{24}}{z_{12} z_{23}} \stackrel{IBP}{=} \frac{(1 - s - t) t}{(1 - s)(2 - s)} \text{PT}(1234) ,$$

(App. 6)

which involves exotic tachyon poles.

Our example captures the essence of the general idea. In any SL(2) weight two integrand $\mathcal{I}_n$, if $z_{ab}$ appear in the numerator, $a$ and $b$ must also appear in the denominator in, say, $z_{ac}$ and $z_{bd}$. We can remove the numerator $z_{ab}$ by absorbing the cross ratio $\frac{z_{ab} z_{cd}}{z_{ac} z_{bd}}$ into the Koba-Nielsen factor, schematically shown as

$$\frac{a}{b} \frac{a}{c} \frac{a}{d} \frac{b}{c} \frac{c}{d} \frac{d}{b} \text{KN} = \frac{a}{b} \frac{a}{c} \frac{a}{d} \frac{b}{c} \frac{c}{d} \frac{d}{b} \text{KN}.'$$

(App. 7)

By repeating this operation, we obtain an integrand that contains only disjoint subcycles. After gauge fixing, we

\[
\begin{align*}
\text{IBP} & \equiv \frac{s_{24}}{1 - s_{123}} \frac{1}{z_{12}} \frac{1}{z_{23}} \frac{1}{z_{34}} + \frac{s_{25}}{1 - s_{123}} \frac{1}{z_{24}} \frac{1}{z_{13}} \frac{1}{z_{23}} - \frac{s_{14}}{1 - s_{123}} \frac{1}{z_{12}} \frac{1}{z_{23}} \frac{1}{z_{13}} - \frac{s_{15}}{1 - s_{123}} \frac{1}{z_{12}} \frac{1}{z_{13}} \frac{1}{z_{24}} + \frac{s_{26}}{1 - s_{123}} \frac{1}{z_{24}} \frac{1}{z_{13}} \frac{1}{z_{23}}.
\end{align*}
\]
can then follow Sec. [13] and reduce it to a sum of PT factors via IBP relations, like the first row of Eq. (A4).

Next, we release a cross ratio, say, \( z_{ac}z_{bd} \) from the deformed Koba-Nielsen factor \( K_N' \). For each PT factor, we first use KK to move \( b \) next to \( a \), and then gauge fix \( z_c \rightarrow \infty \). The resultant integrand only contains a tadpole and a labeled line:

\[
\begin{align*}
\text{c} & \quad \text{a} \quad \text{b} \\
\text{a} & \quad \text{b} \\
\text{c} & \quad \text{d} \\
\end{align*}
\]

which can be turned into a sum of PT factors by IBP according to Sec. [13]. This process is captured by Eq. (A5) and (A6) in our example. By recursively releasing cross ratios from the deformed Koba-Nielsen factor, we will eventually land on a logarithmic function together with the original string measure, which completes our constructive proof.

**Appendix B: Single trace integrand with three gravitons**

In this appendix, we derive the three-graviton CHY integrand [21] from IBP. This calculation is similar to the one in Ref. [10], but we present the result in a different form. We start with the \( R_3(z) \) given by Eq. (9) and (10):

\[
R_3(z) = C_1 C_2 C_3 + \frac{\epsilon_{	ext{E}} \cdot \epsilon_2}{\alpha^{z_{12}}_z} C_3 + \frac{\epsilon_{	ext{E}} \cdot \epsilon_3}{\alpha^{z_{13}}_z} C_2 + \frac{\epsilon_{	ext{E}} \cdot \epsilon_3}{\alpha^{z_{23}}_z} C_1,
\]

and we work with the gauge fix \( z_n \rightarrow \infty \).

Our first step is to separate out the length-two and three cycles contained in \( C_1 C_2 C_3 \) and combine them with the second row of Eq. (B1):

\[
R_3(z) = C_1 C_2 C_3 - C_{12} C_{21} C_3 - C_{13} C_{31} C_2 - C_{23} C_{32} C_1 - C_{12} C_{23} C_{31} - C_{13} C_{32} C_{21} \quad \text{(B2)}
\]

where we have defined \( C_{ij} := \frac{\epsilon_{k_i}}{z_{ij}} \) and

\[
E_{ij} = \frac{\epsilon_{i} \cdot \epsilon_{j} - \alpha' k_i \epsilon_j - k_j}{1 - s_{ij}} = \epsilon_{i} \cdot \epsilon_{j} - \alpha' T_{ij}.
\]

In Eq. (B2), the two rows above the dashed line do not contain any \( z \)-cycles, as one can check straightforwardly. The first three terms below the dashed line of Eq. (B2) contain only length-three cycles, while the last two terms contain only length-three cycles. All the length-two cycles can be broken by Eq. (12) and the following relation:

\[
\frac{1}{z_{12} z_{21}} \quad \frac{1}{z_{31}} \quad \text{IBP} \quad \frac{1}{1 - s_{ij}} \quad \sum_{i=4}^{n-1} \frac{s_{2i}}{z_{2i} z_{31} z_{12}} + \frac{s_{23}}{z_{12} z_{23} z_{31}} \quad \text{(B4)}
\]

which holds when its coefficient does not contain \( z_2 \). We note that this IBP generates new length-three cycles.

Next, we can break all the length-three cycles by the IBP relation

\[
\frac{1}{z_{12} z_{23} z_{31}} \quad \text{IBP} \quad \frac{1}{1 - s_{123}} \quad \sum_{i=4}^{n-1} \left[ \frac{s_{3i}}{z_{12} z_{23} z_{3i}} - \frac{s_{2i}}{z_{13} z_{32} z_{2i}} \right] \quad \text{(B5)}
\]

which holds when its coefficient does not contain either \( z_2 \) or \( z_3 \). The final result is

\[
R_3(z) \quad \text{IBP} \quad C_1 C_2 C_3 - C_{12} C_{21} C_3 - C_{13} C_{31} C_2 - C_{23} C_{32} C_1 - C_{12} C_{23} C_{31} - C_{13} C_{32} C_{21} \quad \text{(B6)}
\]

After being multiplied by the Parke-Taylor factor, Eq. (B6) is just the logarithmic form integrand \( \mathcal{I}'_n \) for the three-graviton case.

Since \( \mathcal{I}'_n \) is also a CHY integrand, we can use SE to rewrite it into a more compact form. For example, the summation in the last row of Eq. (B6) can be reduced to

\[
\sum_{i=4}^{n-1} \left[ \frac{s_{3i}}{z_{12} z_{23} z_{3i}} - \frac{s_{2i}}{z_{13} z_{32} z_{2i}} \right] = - \frac{s_{123}}{z_{12} z_{23} z_{31}}.
\]

(B7)
One can easily show that on the support of SE, Eq. (B6) agrees with the $P_3$ given in Eq. (21). For generic cases, the operation resembles the inverse one as shown in [67–69].

**Appendix C: Pure gluon integrand**

It turns out to be convenient to write pure $(DF)^2 + YM$ integrand $I_n' = PT(12 \cdots n) P_n'$ as

$$P_n' = (-1)^{n(n-1)/2} Pf'(\Psi_n) + \alpha' Q_n,$$  \hspace{1cm} (C1)

where the second term vanishes at $\alpha' = 0$. Moreover, by taking $\alpha' = 0$ in $Q_n$, we should reproduce the half-integrands for a single insertion of $F^3$ vertex [70]. This instructs us to use a similar cycle expansion for $Q_n$:

$$Q_n = \sum_{(J)/(J)-(K)\in S_n} (N_{1}>1-1)\langle \Psi(J) \Psi(K) \cdots \Psi(K) \rangle,$$  \hspace{1cm} (C2)

where $N_{1}>1$ counts the number of cycles with length at least two. The bracket $\langle \cdots \rangle$ stands for certain “averaging” over cycles with length at least two, which will be spelled out shortly, while length-one cycles factor out as:

$$\langle \Psi(a) \Psi(J) \cdots \Psi(K) \rangle = C_a \langle \Psi(J) \cdots \Psi(K) \rangle,$$  \hspace{1cm} (C3)

The longest cycles involved in $Q_n$ have length $n - 2$.

At four points, the $(DF)^2 + YM$ amplitude is explicitly given in [22], which can be produced by

$$Q_4 = -C_1 C_2 C_3 C_4 + \langle \Psi(12) \Psi(34) \rangle + \langle \Psi(23) \Psi(41) \rangle + \langle \Psi(13) \Psi(24) \rangle,$$  \hspace{1cm} (C4)

where the last three terms are given by the relabeling of

$$\langle \Psi(12) \Psi(34) \rangle = \frac{1}{2} T_{12} \langle \Psi(34) + \langle \Psi(12) T_{34} \rangle.$$  \hspace{1cm} (C5)

When going to higher points, the single soft limit fixes all the terms involving length-one cycles, while the requirement of no double pole at massless factorization channels puts strong constraints on the terms without length-one cycles. Actually, we find that up to seven points these constraints lead to a unique answer to the ansatz [22], which involves a modified version of $T_I$:

$$\hat{T_I} = \frac{1}{1 - s_f} \left[ \text{tr}(I) + \sum_{CP} g_p G_{t_1}^{t_1} G_{i_2}^{i_2} \cdots G_{i_1}^{i_1} \right],$$.  \hspace{1cm} (C6)

where $g_p$ depends on $p$, the number of partitions. For $n \leq 7$, we always have $|I| \leq 5$, such that the only relevant $g_p$'s are $g_1 = 1$ and $g_2 = \frac{5}{7}$. In the following, we give explicitly the building blocks of $Q_n$ up to seven points.

a. **Five points**: Besides Eq. (C5), we also need

$$\langle \Psi(12) \Psi(345) \rangle = \frac{1}{4} T_{12} \langle \Psi(345) + \langle \Psi(12) T_{345} \rangle.$$  \hspace{1cm} (C7)

The overall normalization here is different from that in Eq. (C5) because of the different normalizations adopted for $T_{ab}$ and longer cycles, see Eq. (16) and (17).

b. **Six points**: In addition to Eq. (C5) and (C7), there are three more building blocks:

$$\langle \Psi(12) \Psi(3456) \rangle = \frac{1}{4} T_{12} \langle \Psi(3456) + \langle \Psi(12) T_{3456} \rangle,$$  \hspace{1cm} (C8)

$$\langle \Psi(12) \Psi(34567) \rangle = \frac{1}{8} T_{12} \langle \Psi(34567) + \langle \Psi(12) T_{34567} \rangle,$$  \hspace{1cm} (C9)

$$\langle \Psi(12) \Psi(345678) \rangle = \frac{1}{16} T_{12} \langle \Psi(345678) + \langle \Psi(12) T_{345678} \rangle.$$  \hspace{1cm} (C10)

c. **Seven points**: There are again three new building blocks emerge at $n = 7$:

$$\langle \Psi(12) \Psi(345678) \rangle = \frac{1}{8} T_{12} \langle \Psi(345678) + \langle \Psi(12) T_{345678} \rangle,$$  \hspace{1cm} (C11)

$$\langle \Psi(12) \Psi(3456789) \rangle = \frac{1}{16} T_{12} \langle \Psi(3456789) + \langle \Psi(12) T_{3456789} \rangle,$$  \hspace{1cm} (C12)

$$\langle \Psi(12) \Psi(3456789) \rangle = \frac{1}{32} T_{12} \langle \Psi(3456789) + \langle \Psi(12) T_{3456789} \rangle.$$  \hspace{1cm} (C13)

Up to seven points, we find that $\langle \cdots \rangle$ simply averages over $\text{tr}(I)$ and $\hat{T_I}$. We leave the study of generic cases to future works.

---

[1] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, A. B. Goncharov, A. Postnikov, and J. Trnka, Grasmannian Geometry of Scattering Amplitudes (Cambridge University Press, 2016) arXiv:1212.5605 [hep-th].
[63] M. Chiodaroli, M. Gunaydin, H. Johansson, and R. Roiban, JHEP 01, 081 (2015) arXiv:1408.0764 [hep-th].

[64] M. Chiodaroli, M. Gunaydin, H. Johansson, and R. Roiban, JHEP 07, 002 (2017) arXiv:1703.00421 [hep-th].

[65] H. Johansson, G. Mogull, and F. Teng, JHEP 09, 080 (2018) arXiv:1806.05124 [hep-th].

[66] C. Baadsgaard, N. E. J. Bjerrum-Bohr, J. L. Bourjaily, and P. H. Damgaard, JHEP 12, 019 (2016) arXiv:1610.04228 [hep-th].

[67] C. Cardona, B. Feng, H. Gomez, and R. Huang, JHEP 09, 133 (2016) arXiv:1606.00670 [hep-th].

[68] N. E. J. Bjerrum-Bohr, J. L. Bourjaily, P. H. Damgaard, and B. Feng, JHEP 09, 094 (2016) arXiv:1608.00006 [hep-th].

[69] R. Huang, Y.-J. Du, and B. Feng, JHEP 06, 133 (2017) arXiv:1702.05840 [hep-th].

[70] S. He and Y. Zhang, JHEP 02, 019 (2017) arXiv:1608.08448 [hep-th].