The color-dual fate of $\mathcal{N} = 4$ supergravity

John Joseph M. Carrasco,$^{1,2}$ Matthew Lewandowski,$^1$ and Nicolas H. Pavao$^1$

$^1$Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA
$^2$Institut de Physique Théorique, Université Paris Saclay, CEA, CNRS, F-91191 Gif-sur-Yvette, France

We conjecture that the massless gauge theory of Yang-Mills deformed by a higher-derivative $\text{Tr}(F^3)$ operator will not be tree-level color-dual without an infinite number of additional counterterms. The requirement of color-dual kinematics and consistent factorization between four- and five-points induces a tower of increasingly higher-dimensional operators. We find through explicit calculation that their amplitudes are consistent with the $\alpha'$ expansion of those generated by the $(DF)^2 + \text{YM}$ theory, a known color-dual theory where the $\text{Tr}(F^2)$ term has been given a mass squared proportional to $1/\alpha'$. Considering consistent double-copy construction as a physical principle would suggest that a $\text{Tr}(F^3)$-based attempt at a color-dual resolution of the UV divergence of $\mathcal{N} = 4$ supergravity may be able to tame the bad ultraviolet behavior of the theory, but at the cost of field-theoretic locality.

INTRODUCTION

Recent years have demonstrated that perturbative calculation in quantum gravity theories is not as prohibitive as Feynman diagram approaches, in generic gauges, might suggest. The duality between color and kinematics [1], and associated double-copy construction [1–3], reduces the complexity of calculations in many gravity theories to understanding predictions in much simpler gauge theories.

On the other hand, identifying consistent high-energy, or ultraviolet (UV), completions of quantum gravity theories still remains challenging from a point-like quantum field theory perspective. The only proven UV completion to quantum gravity, the closed string, requires an infinite number of higher-derivative corrections from the QFT perspective—arguably rendering the theory non-local. The best candidate for a perturbatively finite local quantum field theory of gravity is the maximally supersymmetric theory [4], $\mathcal{N} = 8$ supergravity (SG), which remains finite in four dimensions at least through the five-loop correction [5–7]. Counterterms compatible with known linearly-realized symmetries have been identified which would be relevant starting at seven loops [8, 9], although their coefficients have not been determined and could vanish in four dimensions. The ultimate fate of $\mathcal{N} = 8$ SG awaits explicit calculation.

The pure half-maximal $\mathcal{N} = 4$ SG survives [10] the analogous test at three-loops [9], only to diverge at four loops [11]. The double-copy structure of pure $\mathcal{N} = 4$ SG and its corresponding spectrum follows from

$$\left(\mathcal{N} = 4 \text{ SG}\right) = \left(\mathcal{N} = 4 \text{ sYM}\right) \otimes \text{YM}. \quad (1)$$

One copy is the maximally supersymmetric $\mathcal{N} = 4$ super Yang-Mills theory (sYM) in four dimensions, and the other pure non-supersymmetric Yang-Mills. The observed divergence at four loops has been linked to the $U(1)$ anomalous behavior [12, 14] of $\mathcal{N} = 4$ SG amplitudes. Such anomalous behavior at one loop can be removed with a simple local counterterm whose double-copy description involves adding the $\text{Tr}(F^3)$ operator to the pure Yang-Mills theory. Does including this counterterm render the theory finite? This too awaits explicit calculation—but investigation at one and two loops [15,16] has verified that the addition of this counterterm, appropriately tuned, does indeed remove the identified anomalous behavior.

Let us now consider a notion of double-copy consistency that applies equally well to gauge theories and gravity theories. Often gauge theories are called color-dual if all color-ordered tree-level amplitudes can be made compatible with the duality between color and kinematics. Gravity theories are called double-copy constructable if their tree-level amplitudes can be constructed from the product of kinematics of color-dual gauge theories. We will describe a quantum field theory as double-copy consistent if every tree-level amplitude can be written as a generalized product of factors that obey the same algebraic relations, and the separate factors can be organized into ordered amplitudes which satisfy all unitarity constraints, i.e. factorizing on physical poles to lower-multiplicity ordered amplitudes from the same theory. We will show in this Letter that the consequence of requiring double-copy consistency of $\text{YM} + F^3$ appears to be an infinite tower of counterterms. The double-copy of this tower, with $\mathcal{N} = 4 \text{ sYM}$, may well be able to tame the UV behavior of the $\mathcal{N} = 4$ SG theory, but at the ultimate cost of field-theory locality.

DOUBLE-COPY CONSISTENCY OF $\text{Tr}(F^3)$

The idea of double-copy consistency gets to the heart of an open question regarding double-copy construction. Should we regard the double copy as a technical trick to be employed piecemeal, amplitude by amplitude as necessary, or rather as a physical principle pointing to the presence of an as-yet-unrecognized physical mechanism braiding together factors of two otherwise consistent theories? In the context of appreciating excellent
UV behavior, we note that the scattering amplitudes of both open and closed tree-level string theory, for external massless states, demand field-theoretic double-copy consistency [17–21].

As a precursor to the more complicated gauge theory involving $\text{Tr}(F^3)$, let us consider a simple effective field theory that requires an infinite number of counterterms to be double-copy consistent. Consider a theory of massless scalars with only the following four-field interaction term,

$$\mathcal{L}_4 = A_{\alpha} f^{abc} f^{ecd} (\partial_\mu \phi_a) \phi_b \phi_c (\partial_\mu \phi_d).$$

All odd-multiplicity amplitudes vanish. The adjoint color-weights (sometimes called flavor) satisfy Jacobi and antisymmetry at every even multiplicity. The color-stripped amplitudes for this theory, however, admit adjoint color-dual representations only at four-points. This theory requires including an additional six-field two-derivative operator in order for the theory’s six-point amplitude to satisfy the adjoint-type duality between color and kinematics. The form of the operator and its coefficient are entirely fixed by the duality between color and kinematics as well as unitarity via consistent factorization to lower multiplicity [22]. Indeed requiring double-copy consistency of this theory involves adding an infinite sequence of even-field two-derivative operators, which sums to the venerable nonlinear sigma model—a theory whose kinematic weights are known to be color-dual to all multiplicity at tree-level [23]. Demanding double-copy consistency for this massless scalar theory involving only a two-derivative four-point interaction encodes the same physical Nambu-Goldstone symmetry as imposing the constraint of the famous Adler’s zero.

Let us now consider Yang-Mills theory with a straight-forward inclusion of the simple $\text{Tr}(F^3)$ operator. For this theory,

$$\mathcal{L}_{YM+F^3} = -\frac{1}{4} \text{Tr}(F^2) + \alpha' \text{Tr}(F^3),$$

all three-point amplitudes satisfy the duality between color and kinematics, as does the four-point tree-level amplitude through $\mathcal{O}(\alpha')$ [23]. As noted in the introduction, this $\mathcal{O}(\alpha')$ contribution is particularly important: when double-copied with $N = 4$ sYM, it describes the local-counterterm required to remove the one-loop $U(1)$ anomalous behavior from $N = 4$ SG. In contrast, the $\mathcal{O}(\alpha'^2)$ contribution to the four-point amplitude, while gauge invariant, does not satisfy the duality between color and kinematics. One must modify $\mathcal{L}_{YM+F^3}$ with an additional $\text{Tr}(F^4)$ operator [24] if one wishes for the four-point amplitudes generated at $\mathcal{O}(\alpha'^2)$ to be color-dual. We would say that the unmodified $\mathcal{L}_{YM+F^3}$ theory is not double-copy consistent.

In fact, it appears that no finite number of local operators is sufficient to render $\mathcal{L}_{YM+F^3}$ double-copy consistent. The situation may be even worse than the scalar example of eq. (2), which at least requires only a finite number of operators to allow any particular multiplicity amplitude to be color-dual. For the theory of $YM + F^3$ we may need to climb to all orders in mass-dimension even at relatively small multiplicity. Indeed, we need not look any further than factorization consistency between four-points and five-points in the theory.

Consider the generalized unitarity cut of a five-point tree-level amplitude at $\mathcal{O}(\alpha'^{n+1})$. The three-point contribution is decomposed into the $F^2$ and $\alpha'F^3$ building blocks.

FIG. 1. Generalized unitarity cut on five-point tree-level amplitude at $\mathcal{O}(\alpha'^{n+1})$. The three-point contribution is decomposed into the $F^2$ and $\alpha'F^3$ building blocks.

$$A_5(12345)(k_4+k_5)^2=\sum_\text{states} A_4(123|\ell) A_3(-\ell 45).$$

Considering the decomposition of the three-point amplitude into Yang-Mills and $\text{Tr}(F^3)$ contributions, $A_3 = A_3^{YM} + \alpha' A_3^{F^3}$, the presence of a purely local four-point contribution at order $\mathcal{O}(\alpha'^n)$ means the existence of a non-vanishing and factorizing five-point amplitude of $\mathcal{O}(\alpha'^{n+1})$. An $\mathcal{O}(\alpha'^{n+1})$ contribution at five-points may not itself be compatible with the duality between color-and kinematics without the additional factorizing contribution from some specific four-point contact term of order $\mathcal{O}(\alpha'^{n+1})$ contracted with the $\text{Tr}(F^2)$ terms. Contraction of any such new four-point contact with the $\alpha' \text{Tr}(F^3)$ terms now forces consideration of the color-dual requirements of a non-vanishing $\mathcal{O}(\alpha'^{n+2})$ contribution at five-points and so forth.

Indeed, this is precisely what we find, by explicit calculation, through $\mathcal{O}(\alpha'^4)$. To tease out this inductive ladder, we start with four-points where adjoint color-dual amplitudes can always be written in terms of cubic, or trivalent, graphs:

$$A_4 = \frac{n \alpha' c_s}{s} + \frac{n \alpha' c_t}{t} + \frac{n \alpha' c_u}{u}.$$
u to refer to the following Mandelstam invariants:

\[ s = (k_1 + k_2)^2 = (k_3 + k_4)^2, \quad (6) \]
\[ t = (k_2 + k_3)^2 = (k_1 + k_4)^2, \quad (7) \]
\[ u = (k_1 + k_3)^2 = (k_2 + k_4)^2 = -s - t. \quad (8) \]

Each graph has a color weight arising from dressing every vertex with a color-structure constant, e.g., \( c_s = c(1, 2, 3, 4) = f^{ab} f^{ba} a_4 \). Similarly, keeping polarization vectors formal, we require a functional map from the labeled graph to kinematic weight such that, e.g.,

\[ n_s = n(1, 2, 3, 4), \quad (9) \]
\[ n_t = n(1, 4, 3, 2), \quad (10) \]
\[ n_u = n(1, 3, 2, 4). \quad (11) \]

We can give this color-dual numerator a generic ansatz through some mass-dimension. We impose the requisite parameters, e.g., \( c_s = c(1, 2, 3, 4) = f^{ab} f^{ba} a_4 \). Similarly, keeping polarization vectors formal, we require a functional map from the labeled graph to kinematic weight such that, e.g.,

\[ n_s = n(1, 2, 3, 4), \quad (9) \]
\[ n_t = n(1, 4, 3, 2), \quad (10) \]
\[ n_u = n(1, 3, 2, 4). \quad (11) \]


d to refer to the normalized scalar permutation invariant at higher points fixed the \( a_i \) to take on any particular non-vanishing value we would not be forced to admit a tower of higher-derivative operators. On the contrary, however, requiring that color-dual five-point amplitudes factorize correctly, entirely fixes \( a_3, a_4, 1, \) and \( a_{42}, 2 \) and relates \( a_{4,F3} \) to \( a_{3,YM} \).

Since we know there are only eight distinct color-dual gauge-invariant vector building blocks at four-points, the most efficient four-point ansatz to arrive at eq. (12) is small–exposing how powers of scalar permutation invariants can map the vector basis to span all relevant mass-dimensions. No such vector basis has yet been clarified at five-points. At five-points we must build an ansatz in terms of bare Lorentz invariants so as to identify the constraints from manifesting color-dual kinematics and factorization.

At five-points there are up to 30 distinct Lorentz products to contend with, and like four-points, there is only one cubic topology to dress. To complete the calculation involves consideration of a 58,923 parameter ansatz spanning from order four in dot products for the Yang-Mills numerator to order eight for a mass-dimension \( \alpha' \) above Yang-Mills. Imposing the duality between color and kinematics without consideration of factorization simply relates these parameters to each other, while setting some to vanish. Generalized unitarity then fixes the coefficients of monomials to appear in a double-copy consistent theory in terms of couplings introduced at lower multiplicity.

We find that requiring color-dual amplitudes at five-points places the following constraints on the free parameters in eq. (12),

\[ a_{4,F3} = 1 + a_{3,YM}, \quad (13) \]
\[ a_3 = 1, \quad (14) \]
\[ a_{4,1} = a_{4,2} = 1. \quad (15) \]

In summary, the five-point vector amplitude requires consideration of kinematics at \( O(\alpha'^2) \), simply by virtue of three insertions of the Tr\((F^3)\) operator. Evidently such a factorizing vector amplitude at five-point can only be color-dual if it also includes terms from a now-required order \( O(\alpha'^3) \) contact term at four-points, eq. (14). This new contact at four-points means we must consider five-points through order \( O(\alpha'^4) \). Requiring a factorizable amplitude of that kinematic mass-dimension to be color-dual in turn demands inclusion of the order \( O(\alpha'^4) \) contact term at four-points, eq. (15). We conjecture that if we are to have a double-copy consistent theory, this tower will continue to all orders in higher-derivative corrections.

A natural question is if the results found so far through five-points and \( O(\alpha'^4) \) can resum to a known theory. We apparently have the freedom to set \( a_{3,YM} = 0 \). If we do so, both our four-point and five-point amplitudes through \( O(\alpha'^4) \) are consistent with the \( \alpha' \) expansion of...
the $B(1, \ldots, n)$ amplitudes of [27]. These $B$ amplitudes [23] belong to the $(DF)^2 + \text{YM}$ theory of ref. [29], where the $(DF)^2$ has been deformed by a massive gauge-theory, with mass scale set by $1/\alpha'$. Indeed, in ref. [23], the four-point amplitude of $(DF)^2 + \text{YM}$ theory was recovered in terms of the above identified color-dual building blocks:

$$n_s(DF)^2 + \text{YM} = n_s^{\text{YM}} + \alpha n_s^{(F)^2 + F^4} + \alpha' n_s^{DF + F^4},$$

$$1 - \alpha'^2 \sigma_2 - \alpha'^3 \sigma_3,$$  \hspace{0.5cm} (16)

with $n_s^{(DF)^4} \equiv n_s^{(DF)^4} + n_s^{(DF)^2}$. The $(DF)^2 + \text{YM}$ theory that generates the $B$ amplitudes is a fascinating color-dual dimension-six theory involving the $\text{Tr}(F^3)$ operator with higher-order propagators. It was first written down by Johansson and Nohle [29] with the explicit aim of finding a double-copy description of conformal supergravity.

Here we come to the most important consequence of our analysis. If one regards double-copy construction as a mathematical trick to simplify the calculation of the kinematic form of specific amplitudes of interest, there is no motivation to introduce a tower of higher-derivative operators. If, on the other hand, we require double-copy consistency as a matter of principle, and we wish to grapple with the UV behavior of $\mathcal{N} = 4$ supergravity by adding the $\text{Tr}(F^3)$ operator to the Yang-Mills copy, we find ourselves lifting the Poincaré supergravity theory to Berkovits-Witten conformal supergravity [30, 31].

$$\mathcal{N} = 4 \text{ Berkovits-Witten} \ldots =$$

$$\mathcal{N} = 4 \text{ sYM} \otimes ((DF)^2 + \text{YM} + \ldots).$$  \hspace{0.5cm} (17)

We include ellipses to emphasize the potential addition of higher-derivative operators unhinged by solely requiring the double-copy consistency of YM + $F^3$. Such freedom is exposed through $O(k^{10})$ by the entirely unconstrained parameter $a_3, \text{YM}$. This freedom can be fixed with particular single-valued multiple zeta values (MZV) [32, 33] so that $((DF)^2 + \text{YM} + \ldots) \rightarrow [(DF)^2 + \text{YM}]^{\alpha'}$, where the superscript denotes a particular mapping to all orders in $\alpha'$ that we will discuss. Such a choice will promote $\mathcal{N} = 4$ SG to the gravitational heterotic string at tree-level.

Using eq. (16) to rewrite our constrained ansatz in eq. (12), without requiring that $a_3, \text{YM}$ vanishes offers a revised form of the four-point numerator for our double-copy consistent theory:

$$n^{\text{dec}} = n^{(DF)^2 + \text{YM}} + O(\alpha'^5)$$

$$+ a_3, \text{YM} \alpha'^3 \sigma_3 \left(n^{\text{YM}} + \alpha' n^{F^3} + O(\alpha'^2)\right).$$  \hspace{0.5cm} (18)

Note that the terms from the second line, $n^{\text{YM}} + \alpha' n^{F^3}$, are the first two terms of the $\alpha'$ expansion of $n^{(DF)^2 + \text{YM}}$ given in eq. (16). This suggests that color-dual consistent amplitudes can be promoted to higher-order contact terms via a product of their color-dual numerators with scalar permutation invariants, and that this information can be consistently propagated to higher multiplicity color-dual amplitudes. Indeed, as we shall discuss shortly, there is a known map from string-theory considerations which offers not only a proof of concept but a prescriptive understanding of how such higher multiplicity color-dual amplitudes may be consistently included. In light of this behavior, it is not hard to imagine that the following resummation spans the full set of double-copy consistent four-point amplitudes, $A_4^{\text{dec}}$, for $\text{Tr}(F^3)$ theories:

$$A_4^{\text{dec}} = B(1, 2, 3, 4) + \sum_{x \geq 1, y} c_{(x,y)} \sigma_3^x \sigma_2^y \alpha'^{3x + 2y},$$  \hspace{0.5cm} (19)

where $\sigma_3$ and $\sigma_2$ are the four-point scalar permutation invariants, and all remaining freedom is given in terms of numeric ansatz parameters $c_{(x,y)}$, which encode the Wilson coefficients of higher-derivative corrections.

**HETEROTIC STRING AND THE SV PROMOTION**

We will now show that the additional UV freedom to add operators to the $(DF)^2 + \text{YM}$ theory will allow us to lift the $\mathcal{N} = 4$ Berkovits-Witten supergravity theories to the tree-level graviton amplitudes of the heterotic string. First, let us recall the field-theoretic double-copy structures appearing in tree-level string-theory amplitudes. The most important clarifying example is the double-copy construction of ordered open superstring amplitudes from ordered Yang-Mills amplitudes and doubly-ordered $Z$-theory amplitudes [17, 21]:

$$A_A^{\text{OSS}} = A_a^{\text{YM}} \otimes ab Z_{ab},$$  \hspace{0.5cm} (20)

where the indices, $a, b$ and $A$, refer to various orderings of kinematic labels, and the outer product is taken to mean a field-theoretic double-copy – expressed particularly with a field-theory KLT kernel acting on ordered amplitudes, and an implied sum over all repeated indices. The doubly-ordered $Z$-theory disc amplitudes encode the string-theoretic scalar corrections at each order in $\alpha'$.

For $n$-point amplitudes, we use capital $A$ to refer to orderings of external legs that satisfy string-theoretic monodromy relations [34] and the lowercase $a$ to index orderings that satisfy field-theoretic amplitude relations [11, 33]. So dressing the $A$ ordering of $Z_{Aa}$ with Chan-Paton color-weights and the $a$ ordering with field-theory color-weights yields color-dressed $n$-point amplitudes in $Z$-theory. This higher-derivative bi-colored scalar theory has the property that its $a$-stripped amplitudes satisfies adjoint-type color-kinematics duality order
We realize the low-energy expansion of and posit in our conjectured resummation of eq. (19).

Field-theoretic double-copy construction, as per eq. (20) and eq. (21), is equally applicable to closed superstring construction. This can be seen by first noting the construction of closed supersymmetric string amplitudes via the string KLT kernel [2], represented here by

$$\text{CSS} = A_A^{\text{OSS}} \otimes_{\alpha'} A_B^{\text{OSS}}.$$  \hfill (22)

Plugging in the expressions for open superstring amplitudes, given in eq. (20), reveals a field-theory double-copy,

$$\text{CSS} = (A_a^{\text{YM}} \otimes_{\alpha'} Z_{Ab}) \otimes_{\alpha'} (Z_{Bc} \otimes_{\alpha'} A_B^{\text{YM}}) = A_a^{\text{YM}} \otimes_{\alpha'} (Z_{Ab}) \otimes_{\alpha'} (A_B^{\text{YM}}) = A_a^{\text{YM}} \otimes_{\alpha'} (A_a^{\text{YM}})_b^c,$$  \hfill (23)

where in the last line we introduced the single-valued promotion of field-theory amplitudes,

$$\left( Y^i \right)_a^b \equiv \left( Z_{Ab} \right) \otimes_{\alpha'} (Z_{Bc} \otimes_{\alpha'} Y_c).$$  \hfill (24)

This operation is called ‘single-valued’ because all the coefficients of $\alpha'$ introduced by the promotion come with only single-valued multiple zeta values. All resulting amplitudes satisfy color-kinematics as well as factorization. At four-points this can be understood as multiplying the $Y$ theory color-dual numerators by scalar permutation invariants at each order in $\alpha'$. The existence of such a map that is consistent to all multiplicity means that we are free to conjecture the most general double-copy consistent UV completion of $B(1, 2, 3, 4)$ to be that of eq. (19).

It was pointed out in ref. [28] that the $B$-theory amplitudes generated by the $(DF)^2 + YM$ Lagrangian also play a critical role in the field-theoretic construction of the gravitational heterotic string amplitude,

$$\text{HS} = B_a \otimes_{\alpha'} (A_a^{YM})_b^c.$$  \hfill (25)

It is clear from the above construction that one could equally well describe the heterotic string as

$$\text{HS} = (B)_a^b \otimes_{\alpha'} A_b^{YM}.$$  \hfill (26)

We see that the remaining freedom in the set of consistent double-copy completions to $YM + F^3$ must allow for the single-valued promotion of $(DF)^2 + YM$. Indeed such is the remaining freedom we find through $O(\alpha'^4)$ in eq. (18), and posit in our conjectured resummation of eq. (19).

We realize the low-energy expansion of $[(DF)^2 + YM]^\nu$ by setting $\alpha_{YM} = c_{1,0} = \zeta(3)$ in eqs. (18) and (19), rather than zero. This means that in the sea of potential double-copy consistent attempts to tame the UV behavior of $\mathcal{N} = 4$ SG by engaging with its anomalous behavior via Tr($F^3$), one of them, with particular single-valued MZV coefficients of additional higher-derivative operators, results in the gravitational amplitudes of the heterotic string.

**NEXT STEPS**

We have presented evidence that demanding double-copy consistency of a gauge theory with the Tr($F^3$) operator induces an all-order tower of $\alpha'$ corrections, which seems to require at a minimum all higher-derivative corrections associated with $(DF)^2 + YM$. There exists a small basis of color-dual vector building blocks, up to trivial scalar permutation invariants, at four-points [25]. Using this basis reduces the complexity of four-point color-dual vector amplitudes to simple considerations of what permutation invariant scalars are required for a given mass-dimension. Developing a similar basis for vector building-blocks at five-points, as has already been done for higher-derivative color-weights [26], could allow a simple derivation that YM + $F^3$ must close to $(DF)^2 + YM$ under double-copy consistency. If we require double-copy consistency of an $\mathcal{N} = 4$ SG theory modified by $(\mathcal{N} = 4 sYM) \otimes \text{Tr}(F^3)$ it appears that the fate of the theory lies in a family of Berkovitz-Witten conformal supergravity theories with freedom to add an additional tower of higher-derivative corrections. A particular choice of their Wilson coefficients takes the theory to the graviton amplitudes of the heterotic string. We look forward to verifying the interplay of such higher-derivative operators on the UV behavior of the theory through explicit calculation.

**Acknowledgements:** We gratefully acknowledge related collaboration and insightful conversations with Zvi Bern, James Mangen, Laurentiu Rodina, Radu Roiban, Aslan Seifi, and Suna Zekioğlu. We would like to thank Henrik Johansson, James Mangen, Olver Schlotterer, and Suna Zekioğlu for very helpful feedback regarding earlier drafts. This work was supported by the DOE under contract DE-SC0021485 and by the Alfred P. Sloan Foundation. M.L. and N.H.P. would additionally like to acknowledge the Northwestern University Amplitudes and Insight group, Department of Physics and Astronomy, and Weinberg College for support.

[1] Z. Bern, J. J. M. Carrasco, and H. Johansson, Phys. Rev. D78, 085011 (2008), arXiv:0805.3993 [hep-ph]

[2] H. Kawai, D. C. Lewellen, and S. H. H. Tye, Nucl. Phys. B269, 1 (1986)
