From building blocks to theories: EFThedron and a Haagerup TFT

Thesis by
Tzu-Chen Huang

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended May 25, 2022

Tzu-Chen Huang
ORCID: 0000-0002-8738-7695

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Hirosi Ooguri for his guidance. I learned a lot about approaching research problems and research rigor from him. He has given me great freedom to explore and work on different topics. When I was struggling with certain events in my life, he offered me his wisdom.

I'm deeply indebted to Yu-tin Huang. I cherished the time we spent together working on problems, often while drunk. Had I not met him in my undergraduate study and started a research project under his supervision, I wouldn't have been able to appreciate the beauty of theoretical physics and the thrill of doing research about it.

It wouldn't be possible for me to finish this thesis without Ying-Hsuan Lin. His passion for research lit a candle for me right at the time when I doubted my direction in life most severely. What was initially a candlelight burns so brightly now that it has become my own passion.

I also want to thank my wonderful collaborators: Nima Arkani-Hamed, Gary Shiu, Toshifumi Noumi, Stefano Andriolo, Yuji Tachikawa, Kantaro Ohmori, and Masaki Tezuka.

I'm grateful to my thesis committee David Simmons-Duffin, Clifford Cheung, and Frank Porter, who have helped me so much during the whole process. I have also learnt a lot in David's and Cliff's group meetings, and would like to thank them for welcoming me.

Thanks to friends in ACT for helping me with countless things from the very beginning of my graduate school life. I will forever remember the time we spent together. Special thanks to Yu-Li Ni for offering me helps in critical moments in my life, and for involving me in a wonderful side project during the pandemics. And thanks to Han-Hsin for accompanying me on gourmet adventures. Finally, I have to thank my roommate Hao-Hsuan for putting up with me for so long.

Thanks to my office mates Petr Kravchuk and Nikita Sopenko for being good company. Also thanks to Matthew Heydeman, Murat Kologlu, Lev Spodyneiko, Junyu Liu, Julio Parra-Martinez, Nathan Benjamin, and Cyuan-Han Chang for memorable moments during my time at the high energy theory group.

If I ever achieve anything in my life, it's only because of the constant support and love from my parents and my wife Liang.

ABSTRACT

This thesis is dedicated to the study of certain building blocks of scattering amplitudes in $(3+1)d$ Minkowskian spacetime and that of topological field theory in $(1+1)d$, together with the constraints which result from the properties of these building blocks.

The first part of the thesis is concerned with the introduction of an on-shell formalism for massless and massive particles. We identify all possible three-point tensor structures compatible with the little group symmetry and overall mass dimension, and use them to arrive at a new description of various scattering amplitudes through unitarity and locality. One of the objects that result from this construction, the spinning polynomial, is then fed into the dispersion relation to derive a convex hull constraining the EFT coefficients. We further investigate the intersection of the convex hull resulting from the positive expansion of residue and the half moment curve.

In the second part, we turn our attention to topological defect lines in $(1+1)d$ topological field theory with Haagerup fusion ring. We first solve for the F-symbols of fusion categories in the Haagerup-Izumi family under the assumption of transparency. The purpose of transparency is twofold: it allows for a simple formula for F-symbols while at the same time tremendously simplifies the diagrammatic calculus with topological defect lines. Finally, we construct a topological field theory with 15 pointlike operators and demonstrate that it satisfies the four-point crossing constraints and torus one-point modular invariance constraints.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-Tin Huang. "The EFT-Hedron". In: *JHEP* 05 (2021), p. 259. DOI: `10.1007/JHEP05(2021)259`. arXiv: `2012.15849 [hep-th]`.
T.-C. H. participated in the conception of the project, analysis, and the writing of the paper.

[2] Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-tin Huang. "Scattering amplitudes for all masses and spins". In: *JHEP* 11 (2021), p. 070. DOI: `10.1007/JHEP11(2021)070`. arXiv: `1709.04891 [hep-th]`.
T.-C. H. participated in the conception of the project, the analysis, and the writing of the paper.

[3] Tzu-Chen Huang and Ying-Hsuan Lin. "Topological Field Theory with Haagerup Symmetry". In: (Feb. 2021). DOI: `10.1063/5.0079062`. arXiv: `2102.05664 [hep-th]`.
T.-C. H. participated in the conception of the project, the analysis, and the writing of the paper.

[4] Tzu-Chen Huang and Ying-Hsuan Lin. "The F-Symbols for Transparent Haagerup-Izumi Categories with $G = \mathbb{Z}_{2n+1}$". In: (July 2020). arXiv: `2007.00670 [math.CT]`.
T.-C. H. participated in the conception of the project, wrote the program for solving the pentagon equations, and participated in the writing of the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Symmetry has been a very powerful organizing principle throughout the history of modern physics. Whenever a physical system possesses certain symmetries, it serves as a hint at extra degrees of freedoms as well as a constraint on the possible degrees of freedoms the system can have. At a more fundamental level, symmetries constrain the forms of basic building blocks in a theory. Whether these constraints manifest themselves in the physics they describe is an interesting and important question, and will be the central focus of this thesis.

This thesis consists of two parts. In the first part, which consists of Chapter 2 and Chapter 3, we will be concerned with physics in $3+1$d Minkowskian spacetime. Through the lens of little group symmetry, we will construct an on-shell formalism to describe the local structure of massive and massless particles with arbitrary spins. Since the ingredients of the construction will be solely from physical quantities, local structures that are not constructable using the formalism will be forbidden in a physical theory under our assumptions, reproducing various no-go theorems in the literature. Together with unitarity and locality, this determines the structure of residues for any tree-level $2 \to 2$ scattering process. Finally, properties encoded in these singularities can be translated into statements about low energy observables in effective field theories using dispersion relations.

In the second part, we will look at extended objects living in $1+1$d topological field theory. These objects carry a particular algebraic structure known as the fusion category, which endows them with the ability to fuse together to form new objects like a ordinary symmetry operator, but without the requirement that it must be invertible. We will be interested in a particular family of fusion categories: the Haagerup-Izumi fusion categories. In Chapter 4, we describe the algebraic data specifying the Haagerup-Izumi fusion categories, and we introduce the notion of a transparent fusion category. The property of transparency greatly facilitates the process of solving for the categorical data of those fusion categories. Finally, in Chapter 5, we formulate a $1+1$d topological field theory with local operators and topological defect lines whose fusion structure are given by the $\mathcal{H}_3$ fusion category in the Haagerup-Izumi family.

1.1 On-shell formalism for $(3+1)d$ Quantum Field Theory

Quantum Field Theory(QFT) describes the interactions of local and extended objects via quantum fields. When the QFT is invariant under a spacetime symmetry, we can classify the states by the irreducible representations of that symmetry. In the case where the spacetime is $(3+1)d$ Minkowskian spacetime, we have the Lorentz group $O(1,3)$. If we include translations as well, this becomes the Poincare group.

Irreducible representations of Poincare group are given in Wigner's classification [173], in which one labels a particle by the eigenvalues of the square of energy-momentum four vector and that of the Pauli-Lubanski vector. In the fixed momentum subspace, the action of the Pauli-Lubanski vector forms a Lie algebra whose corresponding Lie group is called the little group. For massless representations, this little group is $ISO(2)$. For massive ones, it is $SO(3)$.

To describe particles with more than one degree of freedom, we use multiple quantum fields that together form Lorentz covariant vectors/tensors. However, there is a discrepancy in degrees of freedom between a physical particle and Lorentz covariant vector/tensor, which leads to the introduction of redundancies in the gauge theory description of particles with spins.

To calculate a scattering process using these gauge fields, one needs to introduce a set of polarization vectors in the plane wave expansion, which should contain the physical degrees of freedom of the particles involved in the scattering while at the same time account for the aforementioned discrepancy. For a massless spin-1 particle with four momentum p^μ, this is implemented by taking the equivalence classes

$$\{\epsilon^\mu | \epsilon^\mu \sim \epsilon^\mu + a p^\mu\} \tag{1.1}$$

as the polarization vectors.

On the other hand, we can exploit the isomorphism between the restricted Lorentz group $SO^+(1,3)$ and $PSL(2,\mathbb{C})$ to map any four-vector p^μ to

$$(\sigma^\mu p_\mu)_{a\dot{b}} = \begin{pmatrix} -p^0 + p^3 & p^1 - ip^2 \\ p^1 + ip^2 & -p^0 - p^3 \end{pmatrix}_{a\dot{b}}. \tag{1.2}$$

The determinant of this matrix is the square of the four vector p^μ, and so for massless p^μ the matrix is singular. We can then introduce *spinor helicity variables*

$$\{\lambda_a, \tilde{\lambda}_{\dot{a}}\}, \tag{1.3}$$

such that

$$\lambda_a \tilde{\lambda}_{\dot{b}} = (\sigma^\mu p_\mu)_{a\dot{b}}. \tag{1.4}$$

These objects are Weyl spinors which individually transform under the $U(1)$ part of the massless little group. They have unit charge with opposite under this transformation, so their contributions cancel in the above formula, as is expected for a function of four momentum.

Using these variables, one can form tensors with desired little group scaling for massless particles straightforwardly. Suppose we fix the convention so that λ_i has charge -1 for particle i. A tensor like

$$\lambda_1^2 \lambda_2 \tilde{\lambda}_3 \tag{1.5}$$

will then have little group charges $(-2, -1, +1)$. Using these variables, the process of determining physical tensor structures gets rid of gauge redundancies completely: any structure one can write down is free of gauge redundancy.

One important property to make the most of the on-shell formalism is that of *unitarity*. In the standard QFT treatment, we separate the trivial part and the interacting part in a scattering matrix $S = 1 + iT$. Unitarity of the scattering process then requires

$$(1 + iT)(1 - iT^\dagger) = SS^\dagger = 1 \Rightarrow T - T^\dagger = iTT^\dagger. \tag{1.6}$$

For initial state $|i\rangle$ and final $|f\rangle$, this reads

$$\langle f|T - T^\dagger|i\rangle = i\langle f|TT^\dagger|i\rangle = \sum_\psi i\langle f|T|\psi\rangle\langle\psi|T^\dagger|i\rangle. \tag{1.7}$$

In the last identity, we inserted a complete basis in the Hilbert space. To tree level in perturbation theory, this can be replaced by a complete basis of one particle states.

Thus, the residues at the simple poles of a tree-level scattering amplitude will be determined by its factorization into lower-point amplitudes. That is, amplitudes involving less particles. In the case of a four point amplitude, since we know the three point amplitudes completely—they are just the tensor structures formed by the on-shell variables—we will be able to construct the residues directly if we know about the one particle state $|\psi\rangle$ that contributes to the above sum.

Here, *locality* comes into play. In terms of scattering amplitudes, it means that the amplitude, considered as a complex function, will only have singularities when sum of the momenta of a subset of particles participating in this scattering becomes on-shell.

Combining the inputs from unitarity and locality, we know the locations of the simple poles of tree-level $2 \to 2$ scattering amplitudes and the form of residue on this pole in terms of three point amplitudes. The knowledge of all possible three point amplitudes through the on-shell formalism then provides us with a way to build/constrain the original $2 \to 2$ scattering amplitudes.

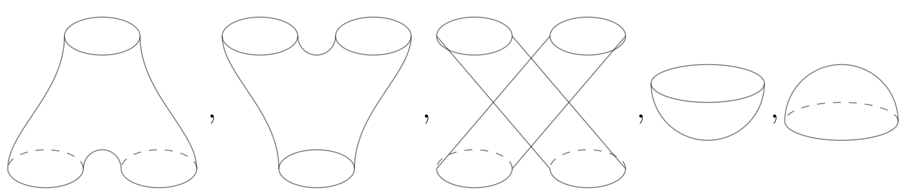
1.2 $1+1d$ topological quantum field theory

Topological quantum field theories, or TQFTs, are QFTs with some special properties. In short, correlation functions of operators in TQFT do not depend on the geometry of the manifold on which it is defined. Early examples include [175] and [151]. For a more recent exposition on the classification of TQFTs, see [124].

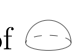
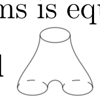
An axiomatic formulation is developed for TQFTs by Michael Atiyah [15] based on an earlier set of proposed axioms by Graeme Segal. In order to describe it, we need to introduce the notion of cobordisms: a cobordism between two manifolds X and Y is a manifold which is one dimension higher and has disjoint boundaries X and Y.

Roughly speaking, an n dimensional TQFT assigns a vector space to each closed oriented $n-1$ dimensional manifold, and a linear map between vector spaces to each cobordism between closed oriented $n-1$ dimensional manifolds. The linear map will be subject to some axioms. This assignment is mathematically known as a *symmetric monoidal functor* from the category of cobordisms to the category of vector spaces.

For TQFTs on two dimensional manifolds, one can construct all cobordisms using the fundamental cobordisms:



(1.8)

More complicated cobordisms can be obtained by composing these fundamental cobordisms, and the resulting image under TQFT are just compositions of the respective linear maps. It turns out that the structure described by these cobordisms is equivalent to that of a *commutative Frobenius algebra*: the composition of  and  gives a pairing $f : A \times A \to k$ between algebra on vector space assigned to S^1 that satisfies

$$\begin{aligned} f(ab, c) &= f(a, bc) \\ f(a, b) &= f(b, a). \end{aligned} \tag{1.9}$$

Together with the existence of a unit through , this defines a commutative Frobenius algebra, completing the description for a $1+1$d TQFT without any topological defect lines.

1.3 Monoidal structure of topological defect lines in $1+1$d

Topological defect lines are extended objects that can be thought of as generalizations of ordinary symmetry operators. Therefore, we will review the basic properties of the ordinary symmetry operators. For a continuous global symmetry transformation G with Noether current j^μ, we can define the extended operator

$$Q[\Sigma] = \int_\Sigma dn^\mu j_\mu. \tag{1.10}$$

When inserted into a correlation function, we can continuously deform the support of the operator Σ without changing the value of the correlation function, as long as we do not pass any local operator charged under the symmetry. When there is a non-vanishing contribution, it will come from the part of Σ that surrounds the charged local operator.

Since the extended operators are defined with respect to elements of a symmetry group, we can fuse two symmetry operators labeled by f and g together to form a new operator that implements transformation fg. Finally, there will be a trivial line operator that does nothing and commutes with everything.

Sometimes the operations implemented by topological lines do not have inverses. We still require there to be a fusion structure associated with the lines. Algebraically, this fusion structure can be captured by a monoidal category.[1]

A *category* C consists a collection of objects $X, Y, \cdots$ and sets of morphisms $\mathrm{Hom}_C(X,Y)$ associated with pairs of objects (X,Y). For every element in the set $\mathrm{Hom}_C(X,Y)$, X is called the domain of the morphism and Y the codomain. If the domain of a morphism is the same as the codomain of another morphism, we can define a new morphism through composition. The composition of morphisms is associative. Finally, there is an identity morphism 1_X for every object X that composes trivially with any morphism involving X.

In order to add more structure to a category, we will need to define a functor. *A functor* F from category C to D assigns an object $F(X) \in \mathrm{Obj}(D)$ to every $X \in \mathrm{Obj}(C)$. The

[1]To properly define the algebraic structure associated with the topological defect lines, we need to be able to take duals, which are interpreted as orientation reversal of lines. This makes a category *rigid*. With some finiteness properties of lines and fusion results, we arrive at the full definition of a fusion category as a rigid semisimple linear monoidal category with finitely many isomorphism classes of simple objects and a simple unit.

simplest example of a functor is an identity functor 1_C, defined for every category C: it sends objects and morphisms to themselves.

A *monoidal category* is a category C together with a unit object I, a functor $\otimes$ from $C \times C$ to C, and three families of natural isomorphisms that are subject to certain axioms. Being a functor from $C \times C$ to C, $\otimes$ specifies the following binary operations:

$$\begin{gathered}\otimes : \mathrm{Ob}(C) \times \mathrm{Ob}(C) \to \mathrm{Ob}(C) \\ \otimes : \mathrm{Hom}_C(X_1, Y_1) \times \mathrm{Hom}_C(X_2, Y_2) \to \mathrm{Hom}_C(X_1 \otimes X_2, Y_1 \otimes Y_2).\end{gathered} \tag{1.11}$$

The identity object and $\otimes$ functor categorizes the notion of identity and product in a monoid. In the language of topological line operators, the $\otimes$ functor represents fusion, while the identity object corresponds to the trivial line. The natural isomorphisms are

- The left(right) unitor assigns an isomorphism from functor $X \to I \otimes X (X \to X \otimes I)$ to the identity functor for each $X \in \mathrm{Ob}(C)$.
- The associator assigns an isomorphism from functor $\otimes(\otimes \times 1_C)$ to functor $\otimes(1_C \times \otimes)$.

These definitions need to be supplemented by two constraints: the triangle axiom, which requires the commutativity of

$$\begin{tikzcd} (X \otimes 1) \otimes X \arrow[rr] \arrow[dr] & & X \otimes (1 \otimes Y) \arrow[dl] \\ & X \otimes Y & \end{tikzcd} , \tag{1.12}$$

and the pentagon axiom, which requires the commutativity of

$$\begin{tikzcd} & ((X \otimes Y) \otimes Z) \otimes W \arrow[dl] \arrow[dr] & \\ (X \otimes (Y \otimes Z)) \otimes W \arrow[d] & & (X \otimes Y) \otimes (Z \otimes W) \arrow[d] \\ X \otimes ((Y \otimes Z) \otimes W) \arrow[rr] & & X \otimes (Y \otimes (Z \otimes W)) \end{tikzcd} . \tag{1.13}$$

The arrows in the diagrams are constructed using the associator, the unitors, and the identity functor. Mac Lane's coherence theorem [125] guarantees that every diagram consists of moves assembled from the natural isomorphisms commutes if these axioms are satisfied.

Chapter 2

SCATTERING AMPLITUDES FOR ALL MASSES AND SPINS

[1]

2.1 Scattering Amplitudes in the Real World

Recent years have seen an explosion of progress in our understanding of scattering amplitudes in gauge theories and gravity. Infinite classes of amplitudes, whose computation would have seemed unthinkable even ten years ago, can now be derived with pen and paper on the back of an envelope using a set of ideas broadly referred to as "on-shell methods" [30–32, 34, 35, 39, 40, 42, 44]. This has enabled the determination of scattering amplitudes of direct interest to collider physics experiments, while at the same time opening up novel directions of theoretical research into the foundations of quantum field theory, amongst other things revealing surprising and deep connections of this basic physics with areas of mathematics ranging from algebraic geometry to combinatorics to number theory.

Almost all of the major progress in this field has been in understanding scattering amplitudes for massless particles. There are seemingly good reasons for this, both technically and conceptually. Technically, almost all treatments of the subject, especially in four dimensions, involve the introduction of special variables (such as spinor-helicity, twistor or momentum-twistor variables) to trivialise the kinematical on-shell constraints for massless particles (see [62, 69, 99] for a comprehensive review). And conceptually, while it is clear that the conventional field-theoretic description of *massless* particles with spin, which involves the introduction of huge gauge redundancy, leaves ample room for improvement—provided by on-shell methods that directly describe particles, eliminating any reference to quantum fields and their attendant redundancies—the advantage of "on-shell physics" seems to disappear for the case of massive particles where no gauge redundancies are needed.

As we will see, the technical issue about massless kinematics is just that—the transition to describing massive particles is a triviality—while the conceptual issue is not an obstacle but rather an invitation to understand the both the physics of "infrared deformation"

[1] This chapter is adapted from Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-tin Huang. "Scattering amplitudes for all masses and spins". In: *JHEP* 11 (2021), p. 070. DOI: 10.1007/JHEP11(2021)070. arXiv: 1709.04891 [hep-th].

of massless theories (by the Higgs mechanism and confinement), as well that of UV completion (such as with perturbative string theory), from a new on-shell perspective (see sec.2.6).

But before getting too far ahead of ourselves it suffices to remember that the only exactly massless particles we know of in the real world are photons and gravitons; even the spectacular success of on-shell methods applied to collider physics are for high energy gluon collisions, which are ultimately confined into massive hadrons at long distances. Even if we consider the weakly coupled scattering amplitudes for Standard Model particles above the QCD scale, almost all the particles are massive. If the amazing structures unearthed in the study of gauge and gravity scattering amplitudes are indeed an indication of a radical new way of thinking about quantum particle interactions in space-time, they must naturally extend beyond photons, gravitons and gluons to electrons, W, Z particles and top quarks as well.

Keeping this central motivation in mind, in this paper we initiate a systematic exploration of the physics of scattering amplitudes in four dimensions, for particles of general masses and spins. We proceed in sec.2.2 with an on-shell formalism where the amplitude is manifestly covariant under the massive SU(2) little group. This approach allows us to cleanly categorize all distinct three-couplings for a given set of helicities or masses and spins. When constructing four-point amplitudes, this formalism sharply pinpoints the tension between locality and consistent factorization, which, in turn provides a portal into the difficulty of having higher-spin massive particles that is fundamental. As we will see, everything that is typically taught in an introductory courses on QFT and the Standard Model—including classic computations of the electron $(g-2)$ and the QCD β function (sec.2.7)—can be transparently reproduced from an on-shell perspective directly following from the physics of Poincare invariance, locality and unitarity, without ever encountering quantum fields, Lagrangians, gauge and diff invariance, or Feynman rules.

There are a number of other motivations for developing this formalism. For instance, much of the remarkable progress in our understanding of the dynamics of supersymmetric gauge theories came from exploring their moduli spaces of vacua [7]. From this point of view the study of massless scattering amplitudes has been stuck on a desert island at the origin of moduli space; we should now be able to study how the S-matrix varies on moduli space in general supersymmetric theories, especially beginning with the Coulomb branch of $\mathcal{N}=4$ SYM in the planar limit (see [57] for early surveys).

Another motivation, alluded to above, is the physics of UV completion for gravity scattering amplitudes. It is easy to show on general grounds that any weakly coupled UV completion for gravity amplitudes must involve an infinite tower of particles with infinitely increasing spins (as of course seen in string theory) [9]. This raises the possibility that string theory might be derivable from the bottom-up, as the unique weakly-coupled UV completion of gravity. But it has become clear that consistency conditions for massless graviton scattering alone are not enough to uniquely fix amplitudes—deformations of the graviton scattering amplitudes compatible with all the standard rules have been identified (eq.(12.6) in [9]). This is not surprising, since the most extreme tension in this physics is the coexistence of gravitons with massive higher-spin particles. Indeed (as we will review in 2.3 from an on-shell perspective) the presence of gravity makes the existence of *massless* higher-spin particles impossible. We should therefore expect the strongest consistency conditions on perturbative UV completion to involve the scattering of massless gravitons and massive higher-spin particles, the study of which calls for a good general formalism for treating amplitudes for general mass and spin.

Finally, an understanding of amplitudes for general mass and spin removes the distinction between "on-shell" observables like scattering amplitudes and "off-shell" observables like correlation functions [71]. After all, loosely speaking the way experimentalists actually measure correlation functions of some system is to weakly couple the system to massive detectors, and effectively measure the scattering amplitudes for the detectors thought of as massive particles with general mass and spin! More precisely, as we demonstrate in sex.2.8, to compute the correlation functions for (say) the stress tensor (in momentum-space), we need only imagine weakly coupling a continuum of massive spin 2 particle to the system with a universal (and arbitrarily weak) coupling; the leading scattering amplitudes for these massive particles is then literally the correlation function for the stress tensor in momentum space. This should allow us to explore both on- and off-shell physics in a uniform "on-shell" way.

2.2 The Little Group

Much of the non-trivial physics of scattering amplitudes traces back to the simple question

—"what is a particle?"—and the attendant concept of Wigner's "little group" governing the kinematics of particle scattering. Let us review this standard story. Following Wigner (and Weinberg's exposition and notation) [20, 171, 173], we think of "particles" as irreducible unitary representations of the Poincare group. We diagonalize the translation operator by labelling particles with their momentum p^μ; any other labels a particle state can carry are labelled by σ. In order to systematically label all one-particle states, we start with some reference momentum k_μ and the states $|k, \sigma\rangle$. Now, we can write any momentum p as a specified Lorentz-transformation $L(p;k)$ acting on k, i.e. $p_\mu = L_\mu^\nu(p;k)k_\nu$. Note that $L(p;k)$ is not unique since there are clearly Lorentz transformations that leave p invariant—these "little group" transformations will figure prominently in what follows, for now we simply emphasize that we pick some specific $L(p;k)$ for which $p = L(p;k)k$. We also assume that we have a unitary representation of the Lorentz group, i.e. for every Lorentz transformation Λ there is an associated unitary operator $U(\Lambda)$ acting on the Hilbert space, such that $U(\Lambda_1\Lambda_2) = U(\Lambda_1)U(\Lambda_2)$. Then we simply *define* one-particle states $|p,\sigma\rangle$ as

$$|p,\sigma\rangle \equiv U(L(p;k))|k,\sigma\rangle\,. \tag{2.1}$$

Note that the σ index is the same on the left and the right, this is the sense in which we are *defining* $|p,\sigma\rangle$. Having made this definition, we can ask how $|p,\sigma\rangle$ transforms under a general Lorentz transformation

$$U(\Lambda)|p,\sigma\rangle = U(\Lambda)U(L(p;k))|k,\sigma\rangle = U(L(\Lambda p;k))U(L^{-1}(\Lambda p;k)\Lambda L(p;k))|k,\sigma\rangle\,. \tag{2.2}$$

Now, $W(\Lambda,p,k) = L^{-1}(\Lambda p;k)\Lambda L(p;k)$ is not in general a trivial Lorentz transformation, it is only a transformation that leave k invariant since clearly $(Wk) = k$. This subgroup of the Lorentz group is the "little group". Thus, we must have that

$$U(W(\Lambda,p;k))|k,\sigma\rangle = D_{\sigma\sigma'}(W(\Lambda,p;k))|k,\sigma'\rangle\,, \tag{2.3}$$

where $D_{\sigma\sigma'}(W)$ is a representation of the little group. We have therefore found the desired transformation property

$$U(\Lambda)|p,\sigma\rangle = D_{\sigma\sigma'}(W(\Lambda,p;k))|\Lambda p,\sigma'\rangle\,. \tag{2.4}$$

We conclude that a particle is labeled by its momentum and transforms under some representation of the little group.

Scattering amplitudes for n particles are thus labeled by (p_a,σ_a) for $a = 1,\cdots,n$. The Poincare invariance of the S-matrix —translation and Lorentz invariance—then tells us that

$$\begin{aligned}\mathcal{M}(p_a,\sigma_a) &= \delta^D(p^\mu_{a_1} + \cdots p^\mu_{a_n})M(p_a,\sigma_a)\\ M^\Lambda(p_a,\sigma_a) &= \prod_a \left(D_{\sigma_a\sigma'_a}(W)\right) M((\Lambda p)_a,\sigma'_a)\,.\end{aligned} \tag{2.5}$$

In D spacetime dimensions, the little group for massive particles is $SO(D-1)$. For massless particles the little group is the the group of Euclidean symmetries in $(D-2)$ dimensions, which is $SO(D-2)$ augmented by $(D-2)$ translations. Finite-dimensional representations require choosing all states to have vanishing eigenvalues under these translations, and hence the little group is just $SO(D{-}2)$.

So much for the basic kinematics of particle scattering amplitudes. It is when we come to dynamics, and in particular to the crucial question of guaranteeing that the physics of particle interactions is compatible with the most minimal notion of locality encoded in the principle of cluster decomposition, that a fateful decision is made to choose a particular description of particle scattering, thereby introducing the idea of quantum fields. Beyond particles of spin zero (and their associated scalar fields), there is a basic kinematical awkwardness associated with introducing fields: fields are manifestly "off-shell", and transform as Lorentz tensors (or spinors), while particle states transform instead under the little group. The objects we compute directly with Feynman diagrams in quantum field theory, which are Lorentz tensors, have the wrong transformation properties to be called "amplitudes". This is why we introduce the idea of "polarisation vectors", that are meant to transform as bi-fundamentals under the Lorentz and little group, to convert "Feynman amplitudes" to the actual "scattering amplitudes". For instance in the case of spin 1 particles, we introduce $\epsilon^\mu_\sigma(p)$, with the property that $\epsilon^\mu_\sigma(\Lambda p) = \Lambda^\mu_\nu \epsilon^\nu_{\sigma'}(p) D_{\sigma\sigma'}(W)$, so that $\epsilon^\mu_\sigma(p) M_\mu(p, \cdots)$ transforms properly. For massive particles, such polarization vectors certainly exist, though they have to satisfy constraints. For instance we must have $p_\mu \epsilon^\mu_\sigma = 0$ for massive spin 1, or for massive spin $1/2$, we use a Dirac spinor Ψ^A_σ with $(\Gamma^\mu p_\mu - m)^A_B \Psi^B = 0$. These constraints are an artifact of using *fields* as auxiliary objects to describe the interactions of the more fundamental *particles*. For massless particles with spin ≥ 1 the situation is worse, since "polarisation vectors" transforming as bi-fundamentals under the Lorentz and little groups don't exist. Say for massless particles in four dimensions, if we make some choice for the $\epsilon^\mu_\pm$ for photons of helicity ± 1, we find that for Lorentz transformations $(\Lambda p) = p$, $(\Lambda \epsilon_\pm)^\mu = e^{\pm i\theta} \epsilon^\mu_\pm + \alpha(\Lambda, p) p^\mu$. So polarisation vectors don't genuinely transform as vectors under Lorentz transformations, only the "gauge equivalence class" $\{\epsilon^\mu_\pm | \epsilon^\mu_\pm + \alpha p^\mu\}$ is invariant under Lorentz transformations. This infinite redundancy is hard-wired into the usual field-theoretic description of scattering amplitudes for gauge bosons and gravitons, and is largely responsible for the apparent enormous complexity of amplitudes in these theories, obscuring the remarkable simplicity and hidden infinite-dimensional symmetries actually found in the physics.

The modern on-shell approach to scattering amplitudes departs from the conventional approach to field theory already at this early kinematical stage, by directly working with objects that transform properly under the little group (and so at least kinematically deserve to be called "scattering amplitudes") from the get-go. Auxiliary objects such as "quantum fields" are never introduced and no polarization vectors are needed. It is maximally easy to do this in the $D = 4$ spacetime dimensions of our world, where the kinematics is as simple as possible. Here the little groups are $SO(2) = U(1)$ for massless particles, and $SO(3) = SU(2)$ for massive particles, which are the simplest and most familiar Lie groups.

In four dimensions, we label massless particles by their helicity h. Massive particles transform as some spin S representation of $SU(2)$. The conventional way of labelling spin states familiar from introductory quantum mechanics is by picking a spin axis $\hat{z}$. and giving the eigenvalue of J_z in that direction. This is inconvenient for our purposes, since the introduction of the reference direction $\hat{z}$ breaks manifest rotational (not to speak of Lorentz) invariance. We will find it more convenient instead to label states of spin S as a symmetric tensor of $SU(2)$ with rank $2S$; this entirely elementary group theory is reviewed in appendix A.2. Let's illustrate the labelling of states by considering a four-particle amplitudes where particles $1, 2$ are massive with spin $1/2$ and 2, and particles $3, 4$ are massless with helicities $+3/2$ and -1. This would be represented as an object

$$M^{\{I_1\},\{J_1,J_2,J_3,J_4\},\{+\frac{3}{2}\},\{-1\}}(p_1, p_2, p_3, p_4) \tag{2.6}$$

where $\{I_1\}, \{J_i\}$ are the little group indices of particle 1 and 2 respectively, and the amplitude transforms as

$$M^{\{I_1\},\{J_1,J_2,J_3,J_4\},\{+\frac{3}{2}\},\{-1\}} \to (W^{I_1}_{1K_1})(W^{J_1}_{2L_1} \cdots W^{J_4}_{2L_4})(w_3)^3(w_4)^{-2} M^{\{K_1\},\{L_1,L_2,L_3,L_4\},\{+\frac{3}{2}\},\{-1\}}, \tag{2.7}$$

where the W matrices are $SU(2)$ transformation in the spin $1/2$ representation and $w = e^{i\theta}$ is the massless little group phase factor for helicity $+1/2$.

Massless and Massive Spinor-Helicity Variables

Our next item of business is to find variables for the kinematics that hardwire these little group transformation laws, this will be simultaneously associated with convenient representations of the on-shell momenta. As usual we will use the $\sigma^\mu_{\alpha\dot{\alpha}}$ matrices to convert between four-momenta p^μ and the 2×2 matrix $p_{\alpha\dot{\alpha}} = p_\mu \sigma^\mu_{\alpha\dot{\alpha}}$[2]. Note that $\det p_{\alpha\dot{\alpha}} = m^2$, so that there is an obvious difference between massless and massive particles.

For massless particles, we have $\det p_{\alpha\dot{\alpha}} = 0$ and thus the matrix $p_{\alpha\dot{\alpha}}$ has rank 1. Thus we can write it as the direct product of two, 2-vectors $\lambda, \tilde{\lambda}$ as [28, 60, 95, 177]

$$p_{\alpha\dot{\alpha}} = \lambda_\alpha \tilde{\lambda}_{\dot{\alpha}}. \tag{2.8}$$

[2]For our conventions of signature and spinor indices, see appendix A.1.

For general complex momenta the $\lambda_\alpha, \tilde{\lambda}_{\dot{\alpha}}$ are independent two-dimensional complex vectors. For real momenta in Minkowski space $p_{\alpha\dot{\alpha}}$ is Hermitian and so we have $\tilde{\lambda}_{\dot{\alpha}} = \pm(\lambda_\alpha)^*$, (with the sign determined by whether the energy is taken to be positive or negative).

Often the introduction of these "spinor-helicity" variables is motivated by the desire to explicitly represent the (on-shell constrained) four-momentum $p_{\alpha\dot{\alpha}}$ by the unconstrained $\lambda_\alpha, \tilde{\lambda}_{\dot{\alpha}}$. But the spinor-helicity variables also have another conceptually important role to play: they are the objects that transform nicely under both the Lorentz and Little groups. Thus while amplitudes for massless particles are *not* functions of momenta and polarization vectors (or better yet, are only redundantly represented in this way), they *are* directly functions of spinor-helicity variables.

The relation to the little group is clearly suggested by the fact that it is impossible to uniquely associate a pair $\lambda_\alpha, \tilde{\lambda}_{\dot{\alpha}}$ with some $p_{\alpha\dot{\alpha}}$, since we can always rescale $\lambda_\alpha \to w^{-1}\lambda_\alpha, \tilde{\lambda}_{\dot{\alpha}} \to w\tilde{\lambda}_{\dot{\alpha}}$ keeping $p_{\alpha\dot{\alpha}}$ invariant. The connection can be made completely explicit by attempting to give some specific prescription for picking $\lambda_\alpha^{(p)}, \tilde{\lambda}_{\dot{\alpha}}^{(p)}$, which leads us through an exercise completely parallel to our discussion of the little group. We first choose some reference massless momentum $k_{\alpha\dot{\alpha}}$ and also choose some fixed $\lambda_\alpha^{(k)}, \tilde{\lambda}_{\dot{\alpha}}^{(k)}$ so that $k_{\alpha\dot{\alpha}} = \lambda_\alpha^{(k)}\tilde{\lambda}_{\dot{\alpha}}^{(k)}$. For every other null momentum, we choose a Lorentz transformation $\mathcal{L}(p;k)_\alpha^\beta, \tilde{\mathcal{L}}(p;k)_{\dot{\alpha}}^{\dot{\beta}}$ such that $p_{\alpha\dot{\alpha}} = \mathcal{L}(p;k)_\alpha^\beta \tilde{\mathcal{L}}(p;k)_{\dot{\alpha}}^{\dot{\beta}} k_{\beta\dot{\beta}}$, and we then *define* $\lambda_\alpha^{(p)} \equiv \mathcal{L}(p;k)_\alpha^\beta \lambda_\beta^{(k)}, \tilde{\lambda}_{\dot{\alpha}}^{(p)} \equiv \tilde{\mathcal{L}}(p;k)_{\dot{\alpha}}^{\dot{\beta}} \tilde{\lambda}_{\dot{\beta}}^{(k)}$. Having now picked a way of associating some $\lambda_\alpha^{(p)}, \tilde{\lambda}_{\dot{\alpha}}^{(p)}$ with $p_{\alpha\dot{\alpha}}$, we can ask for the relationship between, for example, $\lambda_\alpha^{(\Lambda p)}$ and $\lambda_\alpha^{(p)}$ for some Lorentz transformation Λ; what we find is

$$\lambda_\alpha^{(\Lambda p)} = w^{-1}(\Lambda, p, k)\,\Lambda_\alpha^\beta \lambda_\beta^{(p)}. \tag{2.9}$$

For general complex momenta w is simply a complex number and we have the action of $GL(1)$, for real Lorentzian momenta we must have $w^{-1} = \pm(w)^*$ so $w = e^{i\theta}$ is a phase representing the $U(1)$ little group. Most obviously we can perform a Lorentz transformation W for which $Wk = k$, we simply find $\lambda \to w^{-1}\lambda$. To be explicit, let

$$k_{\alpha\dot{\alpha}} = \begin{pmatrix} 2E & 0 \\ 0 & 0 \end{pmatrix},\ \lambda_\alpha = \sqrt{2E}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tilde{\lambda}_{\dot{\alpha}} = \sqrt{2E}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.10}$$

represent a massless momentum in the z direction. Then a rotation around the z axis (which leaves k invariant) is

$$\Lambda_\alpha^\beta = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix}, \tilde{\Lambda}_{\dot{\alpha}}^{\dot{\beta}} = \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix} \tag{2.11}$$

under which obviously $\lambda_\alpha \to e^{i\phi/2}\lambda_\alpha, \tilde{\lambda}_{\dot{\alpha}} \to e^{-i\phi/2}\tilde{\lambda}_{\dot{\alpha}}$.

To summarize, amplitudes for massless particles are Lorentz-invariant functions of $\lambda_\alpha, \tilde{\lambda}_{\dot{\alpha}}$ with the correct little-group helicity weights,

$$M(w^{-1}\lambda, w\tilde{\lambda}) = w^{2h} M(\lambda, \tilde{\lambda}). \tag{2.12}$$

We now turn to the case of massive particles. There is no essential difference with the massless case; we simply have that $p_{\alpha\dot{\alpha}}$ has rank two instead of rank one, and so can be written as the sum of two rank one matrices as

$$p_{\alpha\dot{\alpha}} = \lambda_\alpha^I \tilde{\lambda}_{\dot{\alpha}I}. \tag{2.13}$$

where $I = 1, 2$. Note that

$$p^2 = m^2 \to \mathrm{d}et\lambda \times \mathrm{d}et\tilde{\lambda} = m^2. \tag{2.14}$$

We can use this to set $\mathrm{d}et\lambda = M, \mathrm{d}et\tilde{\lambda} = \tilde{M}$ with $M\tilde{M} = m^2$. It is sometimes useful to keep the distinction between $M, \tilde{M}$, but for our purposes in this paper we will simply take $M = \tilde{M} = m$. Of course $\lambda^I, \tilde{\lambda}_I$ can't uniquely be associated with a given p, we can perform an $SL(2)$ transformation $\lambda^I \to W^I_J \lambda^J, \tilde{\lambda}_I \to (W^{-1})^J_I \tilde{\lambda}_J$. Note that we could extend this $SL(2)$ to a $GL(2)$ if we also allowed (opposite) rephrasings of the mass parameters $M, \tilde{M}$, but by making the choice $M = \tilde{M} = m$ does not allow this. This is not a disadvantage for our purposes, since the object $M/\tilde{M}$ transforms only under the $GL(1)$ part of the $GL(2)$ and can be used to uplift any $SL(2)$ invariant into a $GL(2)$ invariant if desired.

For real Lorentzian momenta we have W should be in the $SU(2)$ subgroup of $SL(2)$ and gives us the action of the little group. We can make the connection explicit just as we did for the massless case, by defining $\lambda_\alpha^I, \tilde{\lambda}_{\dot{\alpha}I}$ for a reference momentum $k_{\alpha\dot{\alpha}}$ and boosting to define them for all momenta. A summary of this elementary kinematics is given in appendix B.

We conclude that that the amplitudes for massive particles are Lorentz-invariant functions for $\lambda^I, \tilde{\lambda}_I$ which are symmetric rank $2S$ tensors $\{I_1, \cdots, I_{2S}\}$ for spin S particles. Note that we can obviously use $\epsilon^{IJ}, \epsilon_{IJ}$ to raise and lower indices so that we can, for example, write $p_{\alpha\dot{\alpha}} = \lambda_\alpha^I \tilde{\lambda}_{\dot{\alpha}}^J \epsilon_{IJ}$. Also note that clearly

$$p_{\alpha\dot{\alpha}} \tilde{\lambda}^{\dot{\alpha}I} = m\, \lambda_\alpha^I\,,\; p_{\alpha\dot{\alpha}} \lambda^{\alpha I} = -m\, \tilde{\lambda}_{\dot{\alpha}}^I. \tag{2.15}$$

If we combine $(\lambda_\alpha^I, \tilde{\lambda}^{\dot{\alpha}I})$ into a Dirac spinor Ψ_A^I, this is of course the Dirac equation $(\Gamma^\mu p_\mu - m)_A^B \Psi_B^I = 0$. But there is no particular reason for doing this in our formalism: even the usual (good) reason for introducing Dirac spinors—making parity manifest in theories which have a parity symmetry—can be more easily accomplished without using Dirac spinors in our approach. We will thus not encounter any Γ matrices in our discussion. Note also that using $(p_{\alpha\dot{\alpha}}/m)$ allows to freely convert between λ^I_α and $\tilde{\lambda}^I_{\dot{\alpha}}$ variables. We will sometimes find it useful, especially in the context of the systematic classification of amplitude structures, to use this freedom in order to use e.g. only λ^I_α to describe a given massive particle. Then we can write the symmetric tensor as

$$M^{\{I_1\cdots I_{2S}\}} = \lambda^{I_1}_{\alpha_1}\cdots\lambda^{I_{2S}}_{\alpha_{2S}} M^{\{\alpha_1\cdots\alpha_{2S}\}} \tag{2.16}$$

where $M^{\{\alpha_1\cdots\alpha_{2S}\}}$ is totally symmetric in the α indices.[3]

Let us illustrate our notation for writing amplitudes by returning to the example of a four-particle amplitude with $(1,2)$ being massive with spin $(1/2,2)$, and $(3,4)$ massless with helicity $(+3/2)$ and (-1). Let's give examples of "legal" expressions for these amplitudes, that is objects with the correct little group transformation properties. Two possible terms are

$$[2^{J_1}3][2^{J_2}3][2^{J_3}3]\left(\kappa\langle 1^{I_1}2^{J_4}\rangle\langle 4|(p_1p_2)|4\rangle + \kappa'\langle 41^{I_1}\rangle\langle 2^{J_4}4\rangle\right) + symmetrize\, in\,\{J_{1,2,3,4}\}. \tag{2.17}$$

It would clearly be notationally cumbersome to have our formulas littered with explicit $SU(2)$ little group indices; fortunately it is also entirely un-necessary to do so. We will simply denote the massive spinor helicity variables in **BOLD**, and suppress the $SU(2)$ little group indices. Since these indices are completely symmetrized, putting them back in is completely trivial and unambiguous. In this way, we re-write the above expressions as

$$[\mathbf{2}3]^3\left(\kappa\langle\mathbf{12}\rangle\langle 4|p_1p_2|4\rangle + \kappa'\langle 4\mathbf{1}\rangle\langle 4\mathbf{2}\rangle\right). \tag{2.18}$$

We stress again that there is no notion of the usual "helicity weight" little group for the massive particles; we can freely have expressions (as in the above) that from the viewpoint of massless amplitudes look like they are "illegally" combining terms with different helicity weight. As we will later see this reflects a beautiful feature of this formalism, making it trivial to see how massive amplitudes decompose into the massless helicity amplitudes at very high energies.

We pause to note the relation between our discussion here and a route to massive spinor-helicity variables taken by a number of other authors [17, 18, 97, 112, 113,

[3] The amplitude as a function of massive spinors can be viewed as a natural consequence of choosing the space-cone gauge Feynman rules [48].

141, 152, 153]. This approach begins by noting that we can always represent $p_{\alpha\dot{\alpha}} = \lambda_\alpha \tilde{\lambda}_{\dot{\alpha}} - (m^2/\langle\lambda\eta\rangle[\tilde{\lambda}\tilde{\eta}])\eta_\alpha\tilde{\eta}_{\dot{\alpha}}$, for some reference spinors $\eta, \tilde{\eta}$.[4] The states are then labelled by giving the spin in the direction picked out by the lightlike directon $\eta\tilde{\eta}$. Of course this corresponds to a particular choice for our $(\lambda^I_\alpha, \tilde{\lambda}^I_{\dot{\alpha}})$, but making this choice at the very outset obscures the Lorentz and little group transformation properties of the amplitude. Practically speaking, given some formula written in terms of the $\lambda, \tilde{\lambda}, \eta, \tilde{\eta}$, this makes it difficult to ascertain whether or not it is kinematically a legal expression for an amplitude, and thus the program of systematically classifying and constructing on-shell amplitudes is difficult to pursue in this formalism.

Let us further illustrate our notation by presenting some classic scattering amplitudes in these variables. We will simply state the results here and derive them from first-principles later in the paper; here we are only illustrating the notation and its utility for understanding the physics. Consider for instance the result for tree-level Compton scattering (12^-3^+4) where particles $2, 3$ are photons of helicity $(-, +)$ while $1, 4$ are charged massive particles of spin $0, 1/2, 1$. The amplitudes are given by

$$M(12^-3^+4) = \frac{g^2}{(s-m^2)(u-m^2)} \times \left\{ \begin{array}{ll} \langle 2|(p_1 - p_4)|3]^2 & [spin\, 0] \\ \langle 2|(p_1 - p_4)|3] \left(\langle \mathbf{12}\rangle[\mathbf{4}3] + \langle \mathbf{4}2\rangle[\mathbf{13}]\right) & [spin\, \tfrac{1}{2}] \\ \left(\langle \mathbf{12}\rangle[\mathbf{4}3] + \langle \mathbf{4}2\rangle[\mathbf{13}]\right)^2 & [spin\, 1] \end{array} \right\}. \tag{2.19}$$

Note the absence of γ matrices for the spin 1/2 case—the common complaint amongst students first doing these computations—"why are we dragging around four-component objects when the electron has only two spin degrees of freedom?"—is entirely absent here. Similarly for the spin 1 case there are no polarization vectors. Indeed these expressions are the most compact representation for these amplitudes possible, directly in terms of the physical degrees of freedom of the actual particles, with no reference to fields as auxiliary objects.

The high-energy limit

It is very easy to relate the massive and massless spinor-helicity variables, and especially to take the high-energy limit of scattering amplitudes and see how massive amplitudes for particles with spin decompose into the different helicity components. To do so, we note that it is convenient to expand λ^I_α in a basis of two-dimensional vectors $\zeta^{\pm I}$ in the

[4]The formalism here obviously have some parallels with the 6D spinor-helicity formalism [33, 52], but here the little group is a single SU(2) instead of SU(2)×SU(2) as in six-dimensions, and thus there are no "unnecessary" symmetries.

little-group space. In other words, we can expand

$$\begin{aligned}\lambda^I_\alpha &= \lambda_\alpha \zeta^{-I} + \eta_\alpha \zeta^{+I} \\ \tilde{\lambda}^I_{\dot\alpha} &= \tilde{\lambda}_{\dot\alpha} \zeta^{+I} + \tilde{\eta}_{\dot\alpha} \zeta^{-I},\end{aligned} \tag{2.20}$$

where

$$\epsilon_{IJ}\zeta^{+I}\zeta^{-J} = 1, \langle\lambda\eta\rangle = m, [\tilde{\lambda}\tilde{\eta}] = m. \tag{2.21}$$

Note, as explicitly given in the kinematics Appendix C, in a given frame we naturally have $\zeta^{\pm I}$ as the eigenstates of spin $1/2$ in the direction of the spatial momentum $\vec{p}$, and we can identify $\lambda_\alpha = \sqrt{E+p}\zeta^+_\alpha, \eta_\alpha = \sqrt{E-p}\zeta^-_\alpha$ and similarly $\tilde{\lambda}_{\dot\alpha} = \sqrt{E+p}\tilde{\zeta}^-_{\dot\alpha}, \tilde{\eta}_{\dot\alpha} = \sqrt{E-p}\tilde{\zeta}^+_{\dot\alpha}$. Clearly, in the high energy limit $\sqrt{E+p} \to \sqrt{2E}$ while $\sqrt{E-p} \to m/\sqrt{2E}$, so that both $\eta, \tilde{\eta}$ are proportional to m and vanish relative to $\lambda, \tilde{\lambda}$. Said in a more Lorentz-invariant way, to take the high-energy limit we take

$$\eta_\alpha = m\hat{\eta}_\alpha,\ \tilde{\eta}_{\dot\alpha} = m\hat{\tilde{\eta}}_{\dot\alpha}; \quad with\, \langle\lambda\hat{\eta}\rangle = [\tilde{\lambda}\hat{\tilde{\eta}}] = 1 \tag{2.22}$$

with all dimensionless ratios of the form

$$\frac{m}{\langle\lambda_a\lambda_b\rangle},\ \frac{m}{[\tilde{\lambda}_a\tilde{\lambda}_b]} \to 0. \tag{2.23}$$

Note that any scattering amplitude naturally decomposes into different spins states in the spatial direction of motion, via

$$M^{I_1\cdots I_{2S}} = \sum_h \left((\zeta^+)^{S+h}(\zeta^-)^{S-h}\right)^{I_1\cdots I_{2S}} M_h(\lambda, \tilde{\lambda}; \eta, \tilde{\eta}), \tag{2.24}$$

where trivially

$$M_h(w^{-1}\lambda, w\tilde{\lambda}; w\eta, w^{-1}\tilde{\eta}) = w^{2h}M_h(\lambda, \tilde{\lambda}; \eta, \tilde{\eta}). \tag{2.25}$$

Thus, the different helicity components in the high-energy limit are just given by

$$\mathrm{H}elicity\, h\, component = \mathrm{L}im_{m\to 0} M_h(\lambda, \tilde{\lambda}; \eta = m\hat{\eta}, \tilde{\eta} = m\hat{\tilde{\eta}}). \tag{2.26}$$

As a simple exercise for taking the high-energy limit, let's consider the coupling of a massive vector to two massless scalars. This amplitude is simply

$$\frac{\langle\mathbf{3}1\rangle\langle\mathbf{3}2\rangle}{\langle 21\rangle}\,. \tag{2.27}$$

Let us consider the high-energy limit of this amplitude. Substituting eq.(2.20), the $(-, 0, +)$ component of the vector are separately given as

$$\begin{aligned}
- :&\quad \frac{\langle\mathbf{3}1\rangle\langle\mathbf{3}2\rangle}{\langle 21\rangle} \xrightarrow{H.E.} \frac{\langle 31\rangle\langle 32\rangle}{\langle 21\rangle} \\
0 :&\quad \frac{\langle\mathbf{3}1\rangle\langle\mathbf{3}2\rangle}{\langle 21\rangle} \xrightarrow{H.E.} \frac{(\langle\eta_3 1\rangle\langle 32\rangle + \langle\eta_3 2\rangle\langle 31\rangle)}{2\langle 21\rangle} \\
+ :&\quad \frac{\langle\mathbf{3}1\rangle\langle\mathbf{3}2\rangle}{\langle 21\rangle} \xrightarrow{H.E.} \frac{\langle\eta_3 1\rangle\langle\eta_3 2\rangle}{\langle 21\rangle} = \frac{[3|p_2|1\rangle[3|p_1|2\rangle}{m^2\langle 21\rangle} = \frac{[32][31]}{[21]}.
\end{aligned} \tag{2.28}$$

We see that only the plus and minus helicity amplitude survives, and as η_3 scales as m, the longitudinal mode is sub-leading in m.[5]

Especially in the context of the rather degenerate kinematics of three particle amplitudes, simply setting the $\eta, \tilde{\eta} \to 0$ can give rise to $0/0$ ambiguities, and this proper definition of the high-energy limit we have specified should be used. But for more generic situations, and for any expressions that is manifestly smooth as $m \to 0$, we can simply set $\eta, \tilde{\eta} \to 0$ to take the high-energy limit. There is an especially easy way of doing this with the "**BOLD**" notation we have introduced above, that shortcuts the need for any explicit expansion in terms of $\zeta^{\pm I}$ as we have indicated above. We simply unbold the characters![6] Let us illustrate how this works for the case of Compton scattering of a charged spin one particle in eq.(2.19), and see how the massive amplitude decomposes into its helicity constituents. Expanding out the square of the numerators we find

$$\underset{(1,4)\,\text{have}\,hel.(-1,+1)}{\langle 12\rangle^2[34]^2} \quad + \quad \underset{(1,4)\,\text{have}\,hel.(+1,-1)}{\langle 42\rangle^2[31]^2} \quad + \quad \underset{(1,4)\,\text{have}\,hel.(0,0)}{2\langle 12\rangle[34]\langle 42\rangle[31]} \tag{2.31}$$

Note that as helicity amplitudes "adding" the components in this way would be illegal, but this is exactly how we can pick out the different pieces of the massive amplitude that unifies the different helicity amplitudes together into a single object, in the high-energy limit! Note also that quite nicely the $(0,0)$ helicity components reproduce the HE limit of the scalar Compton amplitude, reflecting the fact that the longitudinal component of the charged massive spin 1 particle is just a charged scalar at high energies.

2.3 Massless Three- and Four-Particle Amplitudes

Having dispensed with kinematics, we now move on to determining dynamics. We will follow a familiar strategy, starting by determining the structure of all possible three-particle amplitudes:

[5] These results can also be obtained by converting the conventional polarization vector representation of the three particle amplitude to the massive spinor helicity basis. First, being a Lorentz vector and a symmetric tensor in SU(2), the on-shell form of the polarization vector is fixed to (see also [94])

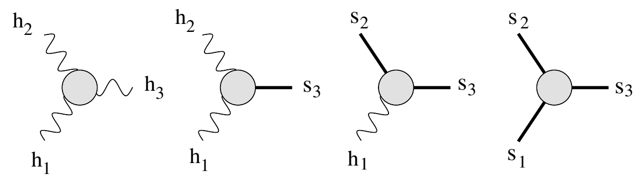
$$\epsilon_{\alpha\dot{\alpha}} = \frac{\lambda_{\alpha}^{\{I_1}\tilde{\lambda}_{\dot{\alpha}}^{I_2\}}}{m}\,. \tag{2.29}$$

Contracting with the momenta then converts the polarization vector to pure chiral indices, $\epsilon_{\alpha\beta} = \epsilon_{\alpha\dot{\alpha}}\frac{p^{\dot{\alpha}}{}_{\beta}}{m}$. Taking the high energy limit, one straight forwardly obtains the three helicity sectors:

$$\epsilon^{-}_{\alpha\beta} = \frac{\lambda_\alpha\lambda_\beta}{m},\ \epsilon^{0}_{\alpha\beta} = \frac{\lambda_\alpha\eta_\beta + \eta_\alpha\lambda_\beta}{2m},\ \epsilon^{+}_{\alpha\beta} = \frac{\eta_\alpha\eta_\beta}{m}\,, \tag{2.30}$$

in the chiral representation.

[6] This is analogous to the replacement of $k \to k^\flat$ in the massive spinor helicity formalism of [116].



(2.32)

When many species $N_{s,m}$ of particle of the identical mass and spin/helicity, we will label them with an index "a".We will always think of these as real particles, and assume that the "free propagation" does not change the a index, i.e. that free propagation has an $SO(N_{s,m})$ symmetry. This choice is hardwiring the most basic physics of unitarity. Note that it is trivial to have (non-unitary) Lagrangian theories that violate this rule, for instance we can have grassmann scalar fields ψ_a with free action $J^{ab}\partial_\mu\psi_a\partial^\mu\psi_b$ with antisymmetric J_{ab}. Here the free propagation is proportional to J^{-1}_{ab} which vanishes for $a = b$, and the free theory has an $Sp(N)$ rather than $SO(N)$ symmetry.

Moving beyond three particles, the central constraint on higher-point tree amplitudes is unitarity, in the form of consistent factorization. For massless or massive internal particles goes on shell, spin s goes on-shell, we must have

$$M \to \frac{M_L^{a\,h} M_R^{a\,-h}}{P^2}\,[massless],\; M \to \frac{M_L^{a\,\{I_1\cdots I_{2s}\}} M^a_{R\,\{I_1\cdots I_{2s}\}}}{P^2 - M^2}\,[massive]\,. \tag{2.33}$$

We will impose this consistency condition at 4 points, which must factorize onto a product of three-particle amplitudes.

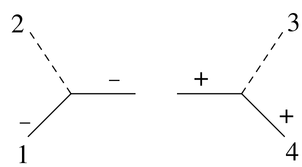
As is by now well-known, these conditions are incredibly restrictive for massless particles. The kinematics of three-particle momentum conservation forces either $\lambda_1, \lambda_2, \lambda_3$ to be all proportional, or $\tilde\lambda_1, \tilde\lambda_2, \tilde\lambda_3$ to all be proportional. Thus the three-particle amplitudes must either be of the form $[12]^a[23]^b[31]^c$ or $\langle 12\rangle^a\langle 23\rangle^b\langle 31\rangle^c$ in these two cases respectively, and the powers are fixed by the helicities of the three particles. The amplitudes are given by

$$M^{h_1h_2h_3} = \begin{array}{c} \tilde g[12]^{h_1+h_2-h_3}[23]^{h_2+h_3-h_1}[31]^{h_3+h_1-h_2}\;when\, h_1+h_2+h_3>0 \\ g\langle 12\rangle^{h_3-h_1-h_2}\langle 23\rangle^{h_1-h_2-h_3}\langle 31\rangle^{h_2-h_3-h_1}\;when\, h_1+h_2+h_3<0 \end{array}\,. \tag{2.34}$$

Note that only by symmetries we could use either of the two expression regardless of the sign of $h_1 + h_2 + h_3$, but we also demand that the amplitudes have a smooth limit in Minkowski signature where the brackets also go to zero. We see that, up to the overall couplings $g, \tilde g$, the three-particle amplitudes are entirely fixed by Poincare symmetry.

We now move on to determining four-particle amplitudes from consistent factorization. The obvious strategy for doing this is to simply compute the residue in, for example, the s-channel by gluing together the three particle amplitudes on the two sides of the channel; then multiply this residue by $1/s$. Adding over the channels should then give us an object that factors correctly in all the channels. This trivially works for ϕ^3 theory where the coupling is simply a constant g, and the residue in each channel is simply g^2. Then an object with the correct poles in all channels is $g^2(1/s+1/t+1/u)$. Of course in addition to this we may have contact terms with no poles at all, and whose form is not fixed by the three-particle amplitudes. But we will only be concerning ourselves with the parts of the four particle amplitudes that are forced to exist by consistent factorization given the three-particle amplitudes.

Let's repeat this exercise for the slightly more interesting case of Yukawa theory, where the three-particle amplitude for fermions 1,2 of helicity $-1/2$ to a scalar 3 is simply $y\langle 12\rangle$. Let us compute the s-channel



$$, \quad R_s = \langle 1I\rangle[I4] = \langle 1|p_I|4] \,, \tag{2.35}$$

where here and in what follows we will suppress the trivial coupling constant dependence. This can be simplified using that $p_I = p_1 + p_2 = -p_3 - p_4$, to $\langle 1|p_2|4] = -\langle 1|p_3|4] = \frac{1}{2}\langle 1|(p_2-p_3)|4]$. The residue in the u channel is the same swapping $2, 3$. So finally the consistently factorizing amplitude is

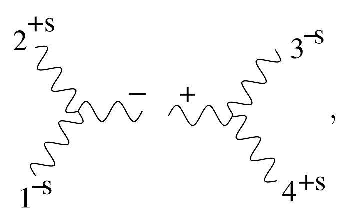
$$\frac{\langle 1|(p_2-p_3)|4]}{s} + \frac{\langle 1|(p_3-p_2)|4]}{u} \,. \tag{2.36}$$

Self-interactions

Let's now try a different example: consider a theory of a single self-interacting particle of spin $\mathbf{s}$. The three particle amplitude for $(1^{-\mathbf{s}}2^{-\mathbf{s}}3^{+\mathbf{s}})$ is $\frac{\langle 12\rangle^{3\mathbf{s}}}{\langle 13\rangle^{\mathbf{s}}\langle 23\rangle^{\mathbf{s}}}$. Note a remarkable feature of this expression, which we did not encounter in either the ϕ^3 or Yukawa theory cases: already the 3 particle amplitude appears to have poles! Thus in a sense these amplitudes are not as "local" as we might have expected. Now of course this peculiarity is un-noticed in the usual Minkowski space, since the three-particle amplitude vanishes in the Lorentzian limit. It is not a coincidence that this subtle sort of "non-locality" appears for precisely the same theories that, in a conventional Lagrangian description, must introduce gauge redundancies for consistency. But returning to our problem of determining four-particle amplitudes by imposing consistent factorization, this feature introduces an important obstruction. The strategy of computing the residue in the s-

channel, multiplying by $1/s$, then summing over channels, is no longer guaranteed to work; as we will see because of the poles in the three-particle amplitudes, the residue in the s channel will itself have poles in the the other channels, making it non-trivial to be able to find an object that consistently factorizes in all channels. Indeed, while we can define massless three-particle amplitudes for any helicities, it will be impossible to find consistent four-point amplitudes for all but the familiar interacting theories of massless spin $0, 1/2, 1, 3/2$ and 2 particles. This exercise has been carried out in systematically in [25, 128], here we highlight some aspects of this story before moving on to carrying out the similar analysis with massive particles.

Let us return to the theory of self-interacting massless particles of spin $\mathbf{s}$; we will consider the four-particle amplitude $(1^{-\mathbf{s}}2^{+\mathbf{s}}3^{-\mathbf{s}}4^{+\mathbf{s}})$. The residue in the s-channel, reached when $[12] \to 0$ and $\langle 34\rangle \to 0$, is



$$, \quad R_s = \left(\frac{\langle I1\rangle^3}{\langle 12\rangle\langle 2I\rangle}\right)^{\mathbf{s}} \left(\frac{[4I]^3}{[I3][34]}\right)^{\mathbf{s}} = \left(\frac{\langle 13\rangle^2[24]^2}{t}\right)^{\mathbf{s}} \tag{2.37}$$

which, again using that e.g. $\langle 1I\rangle[I4] = \langle 12\rangle[24] = -\langle 13\rangle[34]$, can be simplified to $(\frac{\langle 13\rangle^2[24]^2}{t})^{\mathbf{s}}$. We can similarly compute the t, u channel residues, and we find

$$R_s = \left(\frac{\langle 13\rangle^2[24]^2}{t}\right)^{\mathbf{s}}, R_t = \left(\frac{\langle 13\rangle^2[24]^2}{s}\right)^{\mathbf{s}}, R_u = \left(\frac{\langle 13\rangle^2[24]^2}{t}\right)^{\mathbf{s}} \tag{2.38}$$

For $\mathbf{s} \geq 1$, we encountered the challenge alluded to above: the residue in one channel itself has a pole in another channel. Let us start with $\mathbf{s} = 1$. Given the structure of the residues, any consistent amplitude must have the form

$$\langle 13\rangle^2[24]^2\left(\frac{A}{st} + \frac{B}{tu} + \frac{C}{us}\right). \tag{2.39}$$

Note that as $s \to 0$, we have $t = -u$, e.g. the residue in s is $A/t + C/u = (A - C)/t$. In this way, we find that matching the residues in s, t, u demands that $(A-C) = 1, (B-A) = -1, (B - C) = 1$, which is impossible since the sum of the three terms would have to vanish. We conclude that it is impossible to a single self-interacting massless spin 1 particle! But suppose we have many of these particles labelled by the index a; thus the self-interaction of a_1, a_2, a_3 is further proportional to a coupling constant $f^{a_1a_2a_3}$. Note that for $\mathbf{s} = 1$ the three particle amplitude $(1^{-1}2^{-1}3^{+1}) = \frac{\langle 12\rangle^3}{\langle 13\rangle\langle 23\rangle}$ is anti-symmetric in exchanging $1 \leftrightarrow 2$, implying $f^{a_1a_2a_3}$ taking on the same property. Extending to all helicity configurations one can conclude that $f^{a_1a_2a_3}$ must be totally anti-symmetric.

Next consider the four particle amplitude with labels a_1, a_2, a_3, a_4, the residues in the s, t, u channels have additional factors of $f^{a_1 a_2 e} f^{e a_3 a_4}$ and similarly in the t, u channels. Now the ansatz for the four-particle amplitude has the form

$$\langle 13\rangle^2 [24]^2 \left(\frac{A^{a_1a_2a_3a_4}}{st} + \frac{B^{a_1a_2a_3a_4}}{tu} + \frac{C^{a_1a_2a_3a_4}}{us} \right) \tag{2.40}$$

and matching the residues in s, t, u tells us that

$$\begin{aligned} C^{a_1a_2a_3a_4} - A^{a_1a_2a_3a_4} &= f^{a_1a_2e} f^{ea_3a_4} \\ A^{a_1a_2a_3a_4} - B^{a_1a_2a_3a_4} &= f^{a_2a_3e} f^{ea_4a_1} \\ B^{a_1a_2a_3a_4} - C^{a_1a_2a_3a_4} &= f^{a_1a_3e} f^{ea_4a_2} \end{aligned} \tag{2.41}$$

and now, we can solve for $A^{a_1a_2a_3a_4}, B^{a_1a_2a_3a_4}, C^{a_1a_2a_3a_4}$ if and only if the $f^{a_1a_2a_3}$ satisfies the Jacobi identity

$$f^{a_1a_2e} f^{ea_3a_4} + f^{a_2a_3e} f^{ea_1a_4} + f^{a_1a_3e} f^{ea_4a_2} = 0. \tag{2.42}$$

Let's now move on to a single particle with $\mathbf{s} = 2$. Naively, since the residue in the $s-$channel is proportional to $1/u^2$, we might think that it is impossible for the four-particle amplitude to have crucial properties of having only single poles! However, this $1/u^2$ is the residue just as $s \to 0$, and so it could also be represented as $-\frac{1}{tu}$. Thus there is a unique possibility for the four-particle amplitude for a single massless spin two particle:

$$-\frac{\langle 13\rangle^4 [24]^4}{stu} \tag{2.43}$$

which evidently has all the correct residues in all three channels! We can further investigate the possibility on several massless spin two particles, with a coupling constant $g^{a_1a_2a_3}$; the same analysis as for spin one then gives us quadratic constraints on the $g^{a_1a_2a_3}$ that are solved only by g's that, up to change of basis, are only non-vanishing for $a_1 = a_2 = a_3$, i.e. which are mutually non-interacting.

We have thus seen that the only consistently interacting massless spin one particles must have a Yang-Mills structure, and the only consistent massless spin 2 particles does not non-trivially allow more than one such particle, and gives us the standard gravity amplitude. Of course we have done more than simply show the amplitudes are consistent—we have computed them!

For spin $\mathbf{s} > 2$, the residue in the s-channel is at least $1/u^3$, and so there is no way to have a consistent four particle amplitude with only simple poles in s, t, u. We thus conclude that there are no consistent theories of self-interacting massless particles of spin higher than two.

Interactions with other particles

Let's move on to determine what sorts of self-consistent interactions other particles can have with massless spin 1, 2 particles. Let's start with the coupling of a spin $\mathbf{s}$ particles to spin one particle, for which the three particle amplitude is $\langle 12\rangle^{2\mathbf{s}+1}\langle 23\rangle^{1-2\mathbf{s}}\langle 13\rangle^{-1}$. Let us now consider the residues for the $(1^{-s}2^{+}3^{-}4^{+s})$ amplitude; we get residues in the s and u channels from gluing these three-particle amplitudes together. These residues are trivially computed to be

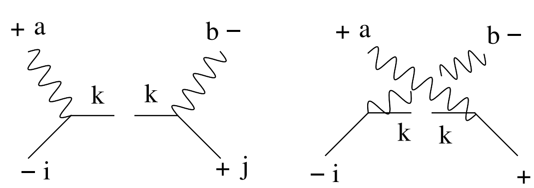
$$R_s = \frac{1}{u}(\langle 13\rangle[24])^{2\mathbf{s}}[2|(p_1-p_4)|3\rangle^{2-2\mathbf{s}},\; R_u = \frac{1}{s}(\langle 13\rangle[24])^{2\mathbf{s}}[2|(p_1-p_4)|3\rangle^{2-2\mathbf{s}}. \tag{2.44}$$

We see there is a qualitative difference between $\mathbf{s} \leq 1$ and $\mathbf{s} \geq 3/2$. For $\mathbf{s} = 0, 1/2, 1$, while the residues in one channel have poles in the other, we can write down a consistently factorizing four-particle amplitude:

$$\frac{(\langle 13\rangle[24])^{2\mathbf{s}}[2|(p_1-p_4)|3\rangle^{2-2\mathbf{s}}}{su}. \tag{2.45}$$

But for $\mathbf{s} \geq 3/2$, the residues have (increasing powers of) the spurious pole in $[2|(p_4 - p_1)|3\rangle$, and so no consistent four particle amplitude is possible. Thus we recover the correct Compton-scattering expressions for particles of spin $0, 1/2, 1$ scattering off photons, while also seeing that it is impossible to have a consistent theory of massless charged particles with spin $\geq 3/2$.

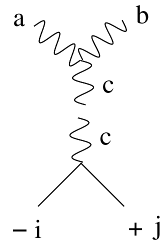
When there are several species of spin $\mathbf{s}$ particles i coupling with several spin one particles a, we attach an extra coupling T^a_{ij} to the vertex. Consider $(1^-_i 2^+_a 3^-_b 4^+_j)$ scattering; writing the residues R in any channel as $R = (\langle 13\rangle[24])^{2\mathbf{s}}[2|p_1|3\rangle^{2-2\mathbf{s}} \times r$, we have



$$,\quad r_s = \frac{1}{u}\,(T^aT^b)_{ij}, \quad r_u = \frac{1}{s}\,(T^bT^a)_{ij}\,, \tag{2.46}$$

where (r_s, r_u) satisfies $s = 0$ and $u = 0$ kinematics respectively. Note that if $(T^aT^b)_{ij} = (T^bT^a)_{ij}$, or the commutator $[T^a, T^b]$ vanishes, we can get a consistent amplitude as with our Compton scattering example, with poles only in these s and u channels, but this is not possible if $[T^a, T^b] \neq 0$. This means that the $1/u$ in r_s and the $1/s$ in r_u must secretly be $1/t$ instead, i.e. must also include a pole in the t channel. Of course fortunately we can have a residue in the t channel, using the cubic self-interaction for gluons. Quite nicely the same kinematical factor appears in R_t, and we find (writing this residue in an

s, u symmetric way):



$$, \quad r_t = (\frac{1}{s}) \times f^{abc} T^c_{ij}. \tag{2.47}$$

Thus, if we have

$$f^{abc} T^c_{ij} = [T^a, T^b]_{ij} \tag{2.48}$$

and using the fact that when $t = 0$, $s = -u$, we find that the following amplitude indeed consistently factorizes in all channels:

$$(\langle 13 \rangle [24])^{2\mathbf{s}} [2|p_1|3\rangle^{2-2\mathbf{s}} \times \left(\frac{(T^a T^b)_{ij}}{ts} + \frac{(T^b T^a)_{ij}}{tu} \right) . \tag{2.49}$$

This agrees with the result in [106]. Also, clearly once again no consistent amplitudes are possible for spin $\mathbf{s} \geq 3/2$. Thus we have discovered the familiar structure of Yang-Mills theories for particles of spin $0, 1/2, 1$.

The same sort of analysis extends to gravity, since the details are virtually identical we will leave them as enjoyable exercises for the reader. We can consider the coupling of two particles of spin $\mathbf{s}$ to a graviton, with strength g. The residues in the s, u channels are no longer equal, and the only way to make a consistent four-particle amplitude is to also have a pole in the t channel, using the graviton self-interaction $\kappa = \frac{1}{M_{Pl}}$. Thus once again the poles for the amplitude are forced to come in the combination $1/stu$. This implies that the coupling constant appearing in the spin-$\mathbf{s}$ exchange channel must be identified with that of the graviton exchange. That is, consistency between the three factorization channel forces the universality of couplings to gravity, $g = \kappa$, with the following form for Compton scattering:

$$\kappa^2 \frac{(\langle 13 \rangle [24])^{2\mathbf{s}} [2|(p_1 - p_4)|3\rangle^{4-2\mathbf{s}}}{stu} . \tag{2.50}$$

Now we see that for $\mathbf{s} \geq 2$ one again develops a spurious pole, and one reaches the conclusion that for spin greater than 2, the particle cannot consistently couple to gravity. In other words, even if higher spin particles are non self-interacting and free, the moment one turns on gravity it ceases to be consistent in flat space. Thus we find that the only possible consistent theories that can couple to gravity can only have spins $(0, 1/2, 1, 3/2)$.[7]

[7] As we remarked in our discussion above on self-interacting spin 2, via a basis change it is always possible to say that the spin 2 particles are effectively in different universes with no mutual interactions; in each one of these decoupled sectors the gravitons can be coupled to their own spectrum of particles with spin $(0, 1/2, 1, 3/2)$.

We can also discover the need for supersymmetry when massless particles of spin 3/2 are present. Consider for simplicity the case with a single spin 3/2 particle ψ. Now let's imagine we also have a massless scalar ϕ. Both of these particles have a universal coupling to gravity, so there is inevitably an amplitude for $\psi_1\psi_1\phi_2\phi_2$ scattering mediated by gravity. We can again compute the residue in the s-channel, and find that it has a pole in the t channel. But since there is no (ψ, ϕ, graviton) coupling (amplitudes must be grassmann even), we can't have any t-channel poles, and so this theory is inconsistent. The only way to have a consistent amplitude is if we *also* introduce a massless fermion χ, now we can have a (ψ, ϕ, χ) interaction with the same gravitational strength $1/M_{Pl}$, which provides the needed pole in the t-channel. The full amplitude is then given as:

$$(1, 2, 3^{-\frac{3}{2}}, 4^{+\frac{3}{2}}) = \kappa^2 \frac{\langle 3|(p_1{-}p_2)|4]^3}{st}\,. \tag{2.51}$$

Thus we see that we must have a bose-fermi degenerate spectrum, with the couplings of the "gravitino" ψ to particles and their superpartners of universal gravitational strength.

We have given a lightning tour of some of the arguments leading to the determination of all consistent theories of massless particles via the "four-particle scattering" test. It is remarkable to see the architecture of fundamental physics emerge from these concrete algebraic consistency conditions in such a simple way. A more complete and systematic treatment can be found in [25, 128].

Before moving on to considering massive amplitudes, let us briefly comment the (in)consistency of theories with three-particle amplitudes for helicities satisfying $h_1{+}h_2{+}h_3 = 0$. Apart from the case of all scalars $h_1 = h_2 = h_3 = 0$, we have "phase" singularities in the couplings, for instance we have a coupling of the form $\langle 13\rangle/\langle 12\rangle$ or $[12]/[13]$ for a spin zero particle 1 to particles 2, 3 of helicity $\pm 1/2$. This peculiar interaction is unfamiliar, and does not arise from Lagrangian couplings. But, as expected, it is also impossible to find a correctly factorizing four-particle amplitude with these couplings [25, 128], so consistency forces the couplings to vanish.

2.4 General Three Particle Amplitudes

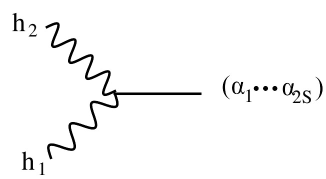
In this section we will categorize the most general three-point amplitude with arbitrary masses. As discussed in section 2.2, the amplitude will be labeled by the spin-S representation of the SU(2) little group for massive legs and helicities for the massless legs. For amplitudes involving massive legs, it will be convenient to expand in terms of λ_α^I, since any dependence on $\tilde{\lambda}_{\dot\alpha}^I$ can be converted using eq.(2.15). For example for a general one massive two massless amplitude, with leg 3 being a massive spin-S state, we have:

$$M_3^{\{I_1\cdots I_{2S}\},h_1,h_2} = \lambda_{3,\alpha_1}^{I_1}\cdots\lambda_{3,\alpha_{2S}}^{I_{2S}} M_3^{\{\alpha_1\cdots\alpha_{2S}\},h_1,h_2}\,, \tag{2.52}$$

where (h_1, h_2) are the helicity. We will be interested in the most general form of the stripped $M_3^{\{\alpha_1\cdots\alpha_{2S}\},h_2,h_3}$, which is now a tensor in the $SL(2,C)$ Lorentz indices. The problem thus reduces to finding two linear independent 2-component spinors that span this space, which we will denote as (v_α, u_α). The convenient choice of (v_α, u_α) will depend on the number of massive legs in a given set up and we will analyze each case separately. We note that a similar classification of three-point interactions using a different basis can be found in [55, 56].

Two-massless one-massive

Let's first begin with the two massless and one massive interaction:



$$M^{h_1h_2}{}_{\{\alpha_1\alpha_2\cdots\alpha_{2S}\}}.$$

Since both legs $1,2$ are massless, their spinors can serve as a natural basis:

$$(v_\alpha, u_\alpha) = (\lambda_{1\alpha}, \lambda_{2\alpha}). \tag{2.53}$$

The helicity weight (h_1, h_2) then completely fixes the degree-$2S$ polynomial in λ_1, λ_2 up to an overall coupling constant:

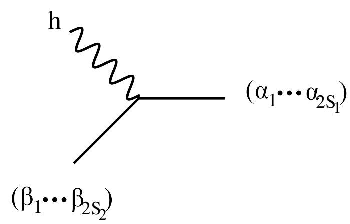
$$M^{h_1h_2}{}_{\{\alpha_1\alpha_2\cdots\alpha_{2S}\}} = \frac{g}{m^{2S+h_1+h_2-1}}\left(\lambda_1^{S+h_2-h_1}\lambda_2^{S+h_1-h_2}\right)_{\{\alpha_1\alpha_2\cdots\alpha_{2S}\}}[12]^{S+h_1+h_2}\,, \tag{2.54}$$

where with appropriate factors of m such that it has the correct mass-dimension. Note that we can trade $[12]$ for $\langle 12\rangle$ using $[12] = \frac{m^2}{\langle 21\rangle}$. When the massive leg is a fermion, i.e. $S \in \frac{1}{2}\mathbb{Z}$, we must then require precisely one of the massless legs to be a fermion as well.

The fact that the structure of this three-point amplitude is unique implies no go theorems for certain interactions. For example, for identical helicities the factor $[12]^{S+2h_1}$ will attain an extra factor of $(-1)^{1+2h_1}$ under $1,2$ exchange for odd spins. This will result in the wrong spin-statistics, thus a particle of odd spin S cannot decay to identical particles with the same helicity. Now suppose the particles have opposite helicity, namely $h_1 = -h_2 = h$. If we take into account that the exponents of λ_1 and λ_2 must both be positive, we conclude that the amplitude vanishes if $|h| > S/2$. For massive spin one states, this is Yang's theorem—that a massive spin one particle cannot decay to a pair of photons. We also learn that a massive spin three particle cannot decay to a pair of gravitons. Note that we have invoked spin-statistics without giving its on-shell origin. As we will see in the coming subsection 2.4, when considering the three-point amplitude of identical massive spin-S states to gravity, spin-statistics is immediately forced upon us.

One-massless two-massive

For two massive legs, the three-point amplitude is now labeled by (h, S_1, S_2)



$$M^h{}_{\{\alpha_1\alpha_2\cdots\alpha_{2S_1}\},\{\beta_1\beta_2\cdots\beta_{2S_2}\}} \tag{2.55}$$

The analysis depends on whether or not the masses are identical. For equal mass, the kinematics becomes degenerate and one expects some form of superficial non-locality. The reason is that the equal mass kinematics occurs precisely for minimal coupling, where its massless limit contain inverse power of spinor brackets as discussed in the previous section. As we will see, for this case we need to introduce a new variable x that encodes this non-locality.

Unequal mass

For unequal mass, one of the basis spinor can be λ of the massless leg, while the remaining can be chosen to be $\tilde{\lambda}$ contracted with one of the massive momentum. For example one can choose

$$(v_\alpha, u_\alpha) = \left(\lambda_\alpha,\ \frac{p_{1\alpha\dot{\beta}}}{m_1}\tilde{\lambda}^{\dot{\beta}}\right). \tag{2.56}$$

Unlike the one massive case, here the amplitude is not unique. The helicity constraint only fixes the polynomial degree in u and v to differ by $2h$. For $S_1 \neq S_2$ there are then a total of $C = S_1+S_2-|S_1-S_2|+1$ different tensor structures, and the general three-point amplitude is given by:

$$M^h{}_{\{\alpha_1\alpha_2\cdots\alpha_{2S_1}\},\{\beta_1\beta_2\cdots\beta_{2S_2}\}} = \sum_{i=1}^{C} g_i (u^{S_1+S_2+h} v^{S_1+S_2-h})^{(i)}_{\{\alpha_1\alpha_2\cdots\alpha_{2S_1}\},\{\beta_1\beta_2\cdots\beta_{2S_2}\}}, \tag{2.57}$$

where i labels the different structure and g_i is the coupling constant for the different tensor structures. Note that the number of possible tensor structures is determined by the lowest spin. For example for one $S_1 = 1$ $S_2 = 2$, we have three tensor structures. For a minus helicity photon these are given by

$$(vvvv)(uu),\ (vvvu)(vu),\ (vvuu)(vv)\,, \tag{2.58}$$

where the parenthesis indicates the grouping of the symmetrized SU(2) little group index. One can also compare this with a Feynman diagram vertex $F_{3,\mu\nu}\epsilon_2^{\nu\rho}\partial_\rho\epsilon_1^\mu$, where ϵ_1, ϵ_2 are the polarization vectors for the massive particles. Again, substituting the on-shell form of the massless polarization vectors $\epsilon_i^- = \frac{|i\rangle[\tilde{\mu}|}{[i\tilde{\mu}]}$, $\epsilon_i^+ = \frac{|\mu\rangle[i|}{\langle\mu i\rangle}$, where $|\tilde{\mu}], |\mu\rangle$ are reference spinors, and massive ones in eq.(2.29), one finds

$$\begin{aligned} M_{3\{\alpha_1\alpha_2\}\{\beta_1\beta_2\beta_3\beta_4\}} &= \frac{m_1^2}{m_2^4}\frac{1}{m_1^2-m_2^2}\left[m_1(uu)_{\{\alpha_1\alpha_2\}}(uuvv)_{\{\beta_1\beta_2\beta_3\beta_4\}}\right. \\ &- \left. m_2(uv)_{\{\alpha_1\alpha_2\}}(uuuv)_{\{\beta_1\beta_2\beta_3\beta_4\}}\right]. \end{aligned} \tag{2.59}$$

Indeed the three-point amplitude for the vertex can be expanded on the basis in eq.(2.58), as it should.

Equal mass: the x-factor

If the masses are identical, then u and v are no longer independent, since

$$v^\alpha u_\alpha = \frac{\langle 3|p_1|3]}{m} = 0\,. \tag{2.60}$$

Thus (u^α, v^α) are parallel to each other and pick out just one direction in the SL(2,C) space. There is however a crucial piece of additional data in the constant of proportionality between u and v, which we will call "x":

$$x\lambda_{3\alpha} = \frac{p_{1\alpha\dot{\alpha}}}{m}\tilde{\lambda}_3^{\dot{\alpha}}\,, \qquad \frac{\tilde{\lambda}_3^{\dot{\alpha}}}{x} = \frac{p_1^{\dot{\alpha}\alpha}\lambda_{3\alpha}}{m}\,. \tag{2.61}$$

Note that x carries $+1$ little group weight of the massless leg. Furthermore, x cannot be expressed in a manifestly local way. Indeed contracting both sides of the above equation with a reference spinor ζ yields

$$x = \frac{\langle\zeta|p_1|3]}{m\langle\zeta 3\rangle}, \tag{2.62}$$

so while x is independent of ζ, any concrete expression for it has an apparent, spurious pole in ζ. In the next section, as we glue the three-point amplitudes to get the four-point, it will be convenient to choose ζ to be the spinor of the external legs on the other side. The denominator then yields a pole in other channels! This yields non-trivial constraint for the four-point amplitude to have consistent factorisation in all channels.

Now the only objects we have carrying SL(2,C) indices are λ_3, as well as the the anti-symmetric tensor $\varepsilon_{\alpha\beta}$.[8] We can then express the three-point amplitude as

$$\begin{aligned} M^h{}_{\{\alpha_1\alpha_2\cdots\alpha_{2S_1}\},\{\beta_1\beta_2\cdots\beta_{2S_2}\}} &= \sum_{i=|S_1-S_2|}^{(S_1+S_2)} g_i x^{h+i} (\lambda_3^{2i} \varepsilon^{S_1+S_2-i})_{\{\alpha_1\alpha_2\cdots\alpha_{2S_1}\},\{\beta_1\beta_2\cdots\beta_{2S_2}\}} \\ &= \sum_{i=|S_1-S_2|}^{(S_1+S_2)} g_i\, x^h \left[\lambda_3^i \left(\frac{p_1\tilde{\lambda}_3}{m}\right)^i \varepsilon^{S_1+S_2-i}\right]_{\{\alpha_1\alpha_2\cdots\alpha_{2S_1}\},\{\beta_1\beta_2\cdots\beta_{2S_2}\}}, \end{aligned} \tag{2.63}$$

where the superscript on $\lambda, \varepsilon, p\tilde{\lambda}/m$ indicates its power. For later purpose we present it in two equivalent representations.

Minimal Coupling for Photons, Gluons, Gravitons

We have seen that while there is a unique structure for massless three-particle amplitudes once the helicities are specified, for couplings of, for example, two equal mass particles of spin S to a massless particle there are $(2S{+}1)$ independent structures, each term with n factors of ε with $n = 0, \cdots, 2S$. Let us take the massless particle to be a graviton. Note that ε is antisymmetric with respect to the exchange $1 \leftrightarrow 2$. Furthermore while the definition of x in eq.(2.61) implies that it picks up a minus sign under the $1 \leftrightarrow 2$, this is irrelevant for gravitational couplings which are proportional to x^2. Thus we see that one gravitation two identical spin S amplitude will have a factor of $(-)^{2S+1}$ under the exchange of the spin-S states. This is nothing but the spin-statistic theorem!

Now one of the $(2S{+}1)$ structures is special, and corresponds to what we usually think of as "minimal coupling" to photons, gluons and gravitons. The defining characteristic of "minimal coupling" is physically very clear. For massless particles, the mass dimension of the couplings is given by $1 - |h_1 + h_2 + h_3|$, and so the leading low-energy interactions with photons, gluons and gravitons—those with dimensionless gauge couplings e, g or gravitational coupling $1/M_{Pl}$, involve massless particles of opposite helicity. The definition of "minimal coupling" for massive particles is then simply the interaction whose leading high-energy limit is dominated by precisely this helicity configuration. As we will see the remaining $(2S+1)-1 = 2S$ interactions represent the various multipole-moment couplings (such as the magnetic dipole moment in the coupling to photons.)

In our undotted SL(2,C) basis, the amplitude with a positive helicity state can be viewed as an expansion in λ. The leading piece in this expansion, namely that where the SL(2,C)

[8] Note in the unequal mass case, since u, v provided a basis, we didn't need to separately introduce $\varepsilon_{\alpha\beta}$ since $(u_\alpha v_\beta - u_\beta v_\alpha) = \langle uv\rangle\varepsilon_{\alpha\beta}$. However as $m_1 \to m_2$ these invariants vanish. This also shows the absence of a singularity in eq.(2.59) as $m_1 \to m_2$.

indices are completely carried by the Levi-Cevita tensors, precisely corresponds to minimal coupling! It is instructive to see why this is the case. Using the simplest example, a photon coupled to two fermions, we find:

$$xm\varepsilon_{\alpha_1\alpha_2} \to x\langle\mathbf{12}\rangle = \langle\mathbf{12}\rangle\frac{\langle\zeta|p_1|3]}{m\langle\zeta 3\rangle} = \frac{\langle\mathbf{2}\zeta\rangle[3\mathbf{1}] + \langle\mathbf{1}\zeta\rangle[3\mathbf{2}]}{\langle\zeta 3\rangle}. \tag{2.64}$$

Taking the high energy limit, we see that the leading term indeed correspond two possible pairs of opposite helicity fermion,

$$\frac{\langle\mathbf{2}\zeta\rangle[3\mathbf{1}] + \langle\mathbf{1}\zeta\rangle[3\mathbf{2}]}{\langle\zeta 3\rangle} \xrightarrow{H.E.} \frac{[13]^2}{[12]} + \frac{[23]^2}{[12]} + \mathcal{O}(m)\,. \tag{2.65}$$

In general the the minimal coupling between photon and two spin-S states is simply:

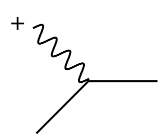
$$\begin{aligned} M^{\mathrm{min},+1}{}_{\{\alpha_1\cdots\alpha_{2S}\},\{\beta_1\cdots\beta_{2S}\}} &= xm\left(\prod_{i=1}^{2S}\varepsilon_{\alpha_i\beta_i} + sym\right), \\ M^{\mathrm{min},-1}{}_{\{\dot\alpha_1\cdots\dot\alpha_{2S}\},\{\dot\beta_1\cdots\dot\beta_{2S}\}} &= \frac{m}{x}\left(\prod_{i=1}^{2S}\varepsilon_{\dot\alpha_i\dot\beta_i} + sym\right), \end{aligned} \tag{2.66}$$

where we've also included the negative helicity photon in its simplest dotted representation. The proper amplitude (with little group indices) is then given as:

$$M^{\mathrm{min},+1} = x\frac{\langle\mathbf{12}\rangle^{2S}}{m^{2S-1}}, \quad M^{\mathrm{min},-1} = \frac{1}{x}\frac{[\mathbf{12}]^{2S}}{m^{2S-1}}. \tag{2.67}$$

For gravitons, we simply introduce an extra power of $\frac{m}{M_{pl}}x$. The fact that in this formalism, minimal coupling is as simple as $\lambda\phi^3$ heralds its potential for simplification. It is also instructive to see how such simple representation emerges from the usual vertices in Feynman rules. Here we present examples for scalar, spinor, and vector at three points:

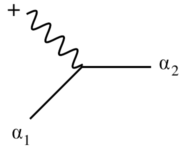
Scalars :



$$\epsilon_3\cdot p_1 = \frac{\langle\xi|p_1|3]}{\langle 3\xi\rangle} = -mx\,, \tag{2.68}$$

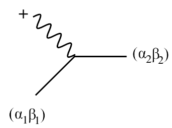
where we've used the identity $xm\lambda_3 = p_1|3]$. Similarly for spin-$\frac{1}{2}$ and 1, we have:

Fermions :



$$\begin{aligned} \bar u_1\,\not\epsilon_3 v_2 &= \left(\frac{p_2^{\dot\gamma\alpha_2}}{m}, \delta_\gamma^{\alpha_2}\right)\begin{pmatrix} 0 & \frac{\tilde\lambda_{3\dot\gamma}\xi_\beta}{\langle 3\xi\rangle} \\ -\frac{\tilde\lambda_3^{\dot\beta}\xi^\gamma}{\langle 3\xi\rangle} & 0\end{pmatrix}\begin{pmatrix}\frac{p_{2\dot\beta}{}^{\alpha_1}}{m} \\ \varepsilon^{\alpha_1\beta}\end{pmatrix} \\ &= x\varepsilon^{\alpha_1\alpha_2} \end{aligned} \tag{2.69}$$

Vectors :



$$\frac{p_1^{\beta_1\dot\alpha_1}}{m}\left[\epsilon_3\cdot p_1\varepsilon_{\alpha_1\alpha_2}\varepsilon_{\dot\alpha_1\dot\alpha_2}+p_{2\alpha_1\dot\alpha_1}\epsilon_{3\alpha_2\dot\alpha_2}-p_{1\alpha_2\dot\alpha_2}\epsilon_{3\alpha_1\dot\alpha_1}\right]\frac{p_2^{\beta_2\dot\alpha_2}}{m}$$
$$= -mx\left(\varepsilon^{\alpha_1\alpha_2}\varepsilon^{\beta_1\beta_2}+sym(\alpha\leftrightarrow\beta)\right). \tag{2.70}$$

The fact that minimal coupling is literally the "minimal" interaction in the undotted SL(2,C) representation indicates the λ expansion should directly correspond to the presence of couplings through higher-dimensional operators. These precisely are the magnetic and electric moments. Let us begin with the magnetic dipole moment. Since this corresponds to a coupling of the particle with $F^{\mu\nu}$, it can only occur for particles with spin. Thus we can extract the electric dipole moment by separating the minimal coupling into a piece that is universal, and pieces that only exists for spinning particles.

Recall that the field strength in momentum space becomes $F_{\mu\nu}\to\lambda_\alpha\lambda_\beta\varepsilon_{\dot\alpha\dot\beta}+\tilde\lambda_{\dot\alpha}\tilde\lambda_{\dot\beta}\varepsilon_{\alpha\beta}$. This implies that couplings through the field strength will be transparent in the undotted frame for negative helicity photon, and dotted frame for the positive photon. With this in mind we convert the minimal coupling for spin-$\frac{1}{2}$ and negative helicity photon into the dotted frame:

$$\frac{p_1^{\alpha\dot\alpha}}{m}\left(\frac{m}{x}\varepsilon_{\dot\alpha\dot\beta}\right)\frac{p_2^{\beta\dot\beta}}{m}=\frac{m}{x}\left(\varepsilon^{\alpha\beta}+x\frac{\lambda_3^\alpha\lambda_3^\beta}{m}\right). \tag{2.71}$$

Here the piece $\frac{m}{x}\varepsilon^{\alpha\beta}$ is the same as that for scalars, sans the $\varepsilon^{\alpha\beta}$ factor which is necessary to carry the SL(2,C) indices, and thus a universal term. The extra piece $\lambda_3^\alpha\lambda_3^\beta$ then represents the magnetic moment coupling, with the amplitude given by

$$\frac{\langle\mathbf{1}3\rangle\langle 3\mathbf{2}\rangle}{m}. \tag{2.72}$$

Thus we immediately see that $g=2$ for the magnetic dipole moment.[9] Thus for minus helicity photon, the general spin-$\frac{1}{2}$ amplitude has the simple expansion:

$$M^{-1}{}_{\dot\alpha_1\dot\alpha_2}=\frac{1}{x}m\varepsilon_{\dot\alpha_1\dot\alpha_2}-\frac{(g-2)}{4}\frac{(\tilde\lambda_3\tilde\lambda_3)_{\dot\alpha_1\dot\alpha_2}}{x^2}, \tag{2.73}$$

where we've manifestly separated the minimal coupling and the $(g-2)$ part of the magnetic dipole moment. It is straight forward to see that $\frac{(\tilde\lambda_3\tilde\lambda_3)_{\dot\alpha_1\dot\alpha_2}}{x^2}$ in the undotted

[9] As a comparison, for the positive helicity and insisting on the undotted frame, we can make the separation after contracting λ^Is. More precisely:

$$x\varepsilon_{\alpha\beta}\ \to\ x\frac{\langle\mathbf{12}\rangle}{m}=\frac{1}{m}\left(x[\mathbf{12}]smatrix/+\frac{[\mathbf{1}3][3\mathbf{2}]}{m}\right).$$

frame is simply $\lambda_3\lambda_3$. For the plus helicity, one has:

$$M^{+1}{}_{\alpha_1\alpha_2} = mx\varepsilon_{\alpha_1\alpha_2} + \frac{(g-2)}{4}x^2(\lambda_3\lambda_3)_{\alpha_1\alpha_2}\,. \tag{2.74}$$

One can trivially extend this to higher spin. For example for spin-1, the minimal coupling now contains both the magnetic dipole moment and electric quadrupole moment. The minimal coupling yields:

$$\begin{aligned}
&\frac{m}{x}\left(\varepsilon^{\alpha_1\alpha_2} - x\frac{\lambda_3^{\alpha_1}\lambda_3^{\alpha_2}}{m}\right)\left(\varepsilon^{\beta_1\beta_2} - x\frac{\lambda_3^{\beta_1}\lambda_3^{\beta_2}}{m}\right) + (\alpha_1 \leftrightarrow \beta_1) \\
= \quad & -\frac{m}{x}\varepsilon^{\alpha_1\{\alpha_2}\varepsilon^{\beta_2\}\beta_1} - \varepsilon^{\alpha_1\{\alpha_2}\lambda_3^{\beta_2\}}\lambda_3^{\beta_1} - \varepsilon^{\beta_1\{\beta_2}\lambda_3^{\alpha_2\}}\lambda_3^{\alpha_1} + 2x\frac{\lambda_3^{\alpha_1}\lambda_3^{\alpha_2}\lambda_3^{\beta_1}\lambda_3^{\beta_2}}{m}\,.
\end{aligned} \tag{2.75}$$

We again see that the first term is the universal piece, and the terms quadratic in λ is the dipole moment where as the terms quartic in λ is the electric quadrupole moment. Thus the general three-point amplitude for the charged vector and a photon is:

$$\begin{aligned}
M^{-1}{}_{\{\dot\alpha_1\dot\beta_1\}\{\dot\alpha_2\dot\beta_2\}} \quad = \quad & \frac{1}{x}m\varepsilon_{\{\dot\alpha_1\dot\alpha_2}\varepsilon_{\dot\beta_1\}\dot\beta_2} + (g-2)\left(\frac{\varepsilon_{\dot\alpha_1\{\dot\alpha_2}\tilde\lambda_{3\dot\beta_2\}}\tilde\lambda_{3\dot\beta_1}}{x^2} + \frac{\varepsilon_{\dot\beta_1\{\dot\beta_2}\tilde\lambda_{3\dot\alpha_2\}}\tilde\lambda_{3\dot\alpha_1}}{x^2}\right) \\
+ \quad & 2(g'+1)\frac{\tilde\lambda_{3\dot\alpha_1}\tilde\lambda_{3\dot\alpha_2}\tilde\lambda_{3\dot\beta_1}\tilde\lambda_{3\dot\beta_1}}{mx^3},
\end{aligned} \tag{2.76}$$

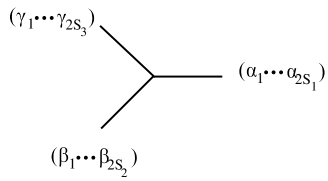
where $(g-2)$ and $(g'+1)$ is the anomalous magnetic dipole and electric quadrupole moment respectively.

Three massive

For all massive legs, we no longer have massless spinors to span the $SL(2,C)$ space. This implies that the space has to be spanned by tensors instead. The fundamental building blocks are now

$$\mathcal{O}_{\alpha\beta} = p_{1\{\alpha\dot\beta}p_{2\beta\}}{}^{\dot\beta}, \quad \varepsilon_{\alpha\beta}\,. \tag{2.77}$$

Note that since $\mathcal{O}_{\alpha\beta}\mathcal{O}_{\gamma\delta} - \mathcal{O}_{\gamma\beta}\mathcal{O}_{\alpha\delta} \sim \varepsilon_{\alpha\gamma}\varepsilon_{\beta\delta}$, pairs of $\varepsilon_{\alpha\beta}$ can be traded for products of $\mathcal{O}_{\alpha\beta}$. The general form of the three-point amplitude is:



$$M_{\alpha_1\cdots\alpha_{2S_1},\beta_1\cdots\beta_{2S_2},\gamma_1\cdots\gamma_{2S_3}} \quad = \quad \sum_{i=0}^{1}\sum_{\sigma_i} g_{\sigma_i}\left(\mathcal{O}^{S_1+S_2+S_3-i}\varepsilon^i\right)^{\sigma_i}_{\{\alpha_1\cdots\alpha_{2S_1}\},\{\beta_1\cdots\beta_{2S_2}\},\{\gamma_1\cdots\gamma_{2S_3}\}}\,, \tag{2.78}$$

where $i = 0, 1$ represents the number of εs and σ_i labels all distinct ways the SU(2) indices can be distributed on $\mathcal{O}$s and should be summed over. It will be interesting to see whether the higher spin interactions from string theory(see [79] for recent results) span the space of all interaction allowed.

2.5 Four Particle Amplitudes For Massive Particles

Now that we have determined the structure of all possible three-particle interactions, we would like to proceed to investigating the consistency of four-particle amplitudes. Just as we did for all massless particles, we ask: given a spectrum of particles, and a set of three-particle interactions, is it possible to find a four-particle amplitude that consistently factorizes in all possible channels? We stress that this is a completely sharply defined and straightforward algebraic problem. To be maximally pedantic, suppose we have a set of particles with masses (zero or non-zero) given by m_i. Then the most general ansatz for the four-particle amplitude has the form

$$\frac{N}{\prod_i (s-m_i^2)(t-m_i^2)(u-m_i^2)} \tag{2.79}$$

and we simply wish to determine whether there is a consistent numerator N that allows this function to factorize correctly in the s, t, u channels[10]

$$M \to \frac{M^a_{L,\{I_1\cdots I_{2s}\}}\varepsilon^{I_1J_1}\cdots\varepsilon^{I_{2s}J_{2s}}M^a_{R\{J_1\cdots J_{2s}\}}}{P^2-M^2}\,. \tag{2.80}$$

As we've shown before, it is convenient to expand the amplitude on the λ^I_α basis, in which case the contraction of little group indices now translates to the contraction of undotted SL(2,C) indices:

$$\frac{\lambda^I_\alpha\lambda^J_\beta}{m}\varepsilon_{IJ} = \varepsilon_{\alpha\beta}\,. \tag{2.81}$$

To make contact with the usual Feynman rules, the numerator of the vector propagator is $G_{\mu\nu} \equiv \eta_{\mu\nu} - \frac{p_\mu p_\nu}{m^2}$, which in SL(2,C) undotted representation is:

$$G_{\alpha\dot\alpha,\beta\dot\beta} = 2\varepsilon_{\alpha\beta}\varepsilon_{\dot\alpha\dot\beta} - \frac{p_{\alpha\dot\alpha}p_{\beta\dot\beta}}{m^2} \;\to\; \frac{p^{\dot\alpha}_\gamma p^{\dot\beta}_\delta}{m^2}G_{\alpha\dot\alpha,\beta\dot\beta} = \varepsilon_{\alpha\{\beta}\varepsilon_{\gamma\delta\}}\,, \tag{2.82}$$

as expected. This is not surprising, as we've discussed in the introduction, the transverse traceless-ness, which determines the numerator of the propagator, simply translates to symmetrization of the SL(2,C) indices.

[10] Of course the amplitude cannot be uniquely determined in this way, since we can always simply have contact terms that are simply polynomials with no poles at all (corresponding to piece in N that cancels all the poles). To avoid clutter, we will suppress the possible contact terms in what follows.

In practice, we don't need to work with this slavishly systematic ansatz for the amplitude with the giant denominator consisting of all possible simple poles. Instead, following the same steps as in the all massless case, given the spectrum and the three-particle amplitudes, we will first simply compute the residues $R(i)_s, R(i)_t, R(i)_u$ in the s, t, u channels from the exchange of the i'th particle. If these residues are local, we are trivially done, since the object

$$\sum_i \left(\frac{R_s^{(i)}}{s - m_i^2} + \frac{R_t^{(i)}}{t - m_i^2} + \frac{R_u^{(i)}}{u - m_i^2} \right) \tag{2.83}$$

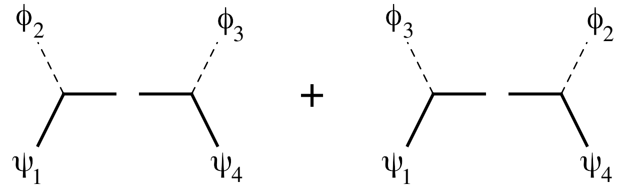
manifestly matches the poles in all the channels. This is the case for the massive $g\phi^3$ theory where these residues are all simply $R_s = R_t = R_u = g^2$. But as we already saw in the massless case, there are more interesting cases where the residues in one channel themselves have poles in another channel. With massive particles this will occur whenever we have minimal coupling and the "x" factor. In this case an ansatz separately summing the channels cannot work, and we must use building blocks that have simple poles in more than one channel. For massless particles, the requirement of four-particle consistency was so strong as to simply make certain theories (of high-enough spin charged or gravitating massless particles) impossible. It also enforced universality of the couplings to gravitons and the usual Yang-Mills structure for coupling to photons and gluons. We will see the analogue of these statements for massive amplitudes. Once again, consistent factorization will demand the standard couplings to photons, gluons and gravitons, will also see that any self-interactions have to be invariant under the (global part) of the gauge symmetry. But with these restriction met, it *is* possible to find consistently factorizing four-particle amplitudes for any masses and spins. This is of course expected, since almost all interesting objects in the real world are massive particles of high spin! But of course as we will also see, the impossibility of consistent amplitudes for massless particles of high spin shows up in a singularity of the massive high spin amplitudes in the high-energy (or $m \to 0$) limit, giving a very concrete sense in which particles of high spin cannot be "elementary".

Manifest local gluing

We first begin with the construction of amplitudes without any x-factor non-localities. Let's begin with Yukawa amplitude, i.e. one massless scalar two massive fermion amplitude. The three-point amplitude is simply

$$\frac{g}{m_f}\langle \mathbf{1}3\rangle[\mathbf{2}3] + \frac{g'}{m_f}\langle \mathbf{2}3\rangle[\mathbf{1}3]. \tag{2.84}$$

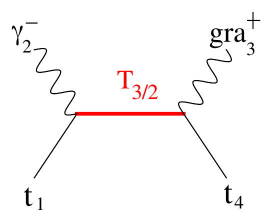
where m_f is the mass of the fermion. The gluing in the s- and u-channel yields:



$$= \frac{g^2\langle\mathbf{12}\rangle\langle\mathbf{34}\rangle[32] + g'g[\mathbf{12}]\langle\mathbf{43}\rangle\langle 2|p_4|3] + c.c.}{m_f(s-m_f^2)} + (2 \leftrightarrow 3)\,, \tag{2.85}$$

where by $c.c.$ we are exchanging $\lambda \leftrightarrow \tilde{\lambda}$ and $g \leftrightarrow g'$. As one can see, since the three-point amplitude was local, the resulting four-point amplitude can be written in a manifest local way with two separate channels.

A "slightly" more complicated example would be the process $\gamma^- + t \rightarrow gra^+ + t$, via a massive spin-$\frac{3}{2}$ exchange:



$$. \tag{2.86}$$

Here, $t_{1,4}$ are the massive top quarks with their mass denoted by m_t. The three-point amplitude on both sides are:

$$\begin{aligned} V_L &= \frac{g}{m_t^3}\langle\mathbf{p}2\rangle^3[\mathbf{1}2] + \frac{g'}{m_t^3}\langle\mathbf{1}2\rangle\langle\mathbf{p}2\rangle^2[\mathbf{p}2], \\ V_R &= \frac{g''}{m_t^3}[\mathbf{4}3][\mathbf{p}3]^3\,. \end{aligned} \tag{2.87}$$

There are two tensor structures for V_L, reflecting the two distinct way the SL(2,C) indices can distribute. The resulting four-point amplitude is then,

$$\frac{gg''[\mathbf{4}3][2\mathbf{1}]\langle 2|p_4|3]^3 + g'g''[\mathbf{4}3]\langle\mathbf{1}2\rangle\langle 2|p_4|3]^2[32]m_t}{(s-m_T^2)m_t^6} + (2 \leftrightarrow 3)\,, \tag{2.88}$$

where m_T is the mass of the spin-3/2 particle.

In the above examples, the residues are manifestly local as it is inherited from the three-point amplitude. The only place potential non-locality can occur is when factors of x appear for the three-point amplitude, for example the minimal coupling. Thus in the next section we will focus on minimal coupling for massless spin-1 and 2 particles.

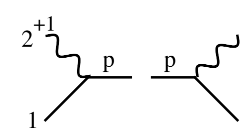
Minimal Coupling

In this subsection we will consider the gluing of minimally coupled higher spin particles. We will first begin with charged particles, which entails the three-point coupling of two massive spin-S state and a positive or negative helicity photon. The three point amplitude is given in eq.(2.66), and after dressing with external spinors the complete amplitude is:

$$M_S^{+h} = x^h \frac{\langle \mathbf{12}\rangle^{2S}}{m^{2S-1}}, \quad M_S^{-h} = \frac{1}{x^h}\frac{[\mathbf{12}]^{2S}}{m^{2S-1}}\,. \tag{2.89}$$

Compton Scattering For S $\leq$ 1

Let us begin with scalar. Here one simply has:



$$\sim m^2 \frac{x_{12}}{x_{34}}. \tag{2.90}$$

Here the subscripts on x serve to distinguish between different three point vertices. Now since

$$x_{12}\lambda_2 = \frac{p_1|2]}{m} \rightarrow x_{12} = \frac{\langle 3|p_1|2]}{\langle 32\rangle m}, \quad \frac{p_4|3\rangle}{m} = \frac{\tilde{\lambda}_3}{x_{34}} \rightarrow \frac{1}{x_{34}} = \frac{\langle 3|p_4|2]}{[23]m}\,, \tag{2.91}$$

we see that the residue is given by:

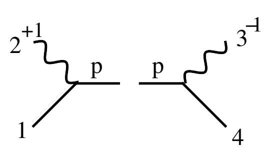
$$m^2\frac{x_{12}}{x_{34}} = -\frac{\langle 3|p_1|2]^2}{t}\,. \tag{2.92}$$

Again the s-channel residue is non-local and must be interpreted as a pole from the other channel! We now have a choice, it can either be interpreted as a massless particle in the t-channel, or an u-channel massive particle since $-t = u - m^2$ when $s = m^2$. For there to be a t-channel massless pole, the vectors must be gluons instead of photons, and we leave this possibility to the later part of this subsection. For the case where one has a u-channel massive pole, the amplitude is simply:

$$M(\phi_1\gamma_2^+\gamma_3^-\phi_4) = \frac{\langle 3|p_1 - p_4|2]^2}{4(s-m^2)(u-m^2)}\,, \tag{2.93}$$

As the amplitude is symmetric under $1 \leftrightarrow 4$ exchange, it is guaranteed to be consistent with the u-channel factorisation. It is straight forward to see that at H.E. one obtains the usual two adjoint-scalar two gluon, and two charged scalar two photon amplitude.

Let us now consider Compton scattering for general spin. The s-channel gluing yields,



$$\sim \frac{1}{m^{2(S-1)}}\frac{x_{12}}{x_{34}}(\langle \mathbf{1}P^I\rangle[P_I\mathbf{4}])^{2S}\,. \tag{2.94}$$

Recall that $\frac{x_{12}}{x_{34}}m^2 = -\langle 3|p_1|2]^2/t$, if we rewrite t as $u - m^2$ and put back the s-channel propagator, this has the property that it is symmetric under $1 \leftrightarrow 4$ (it is the scalar amplitude after all). This means that if $\frac{P^{2S}}{m^{2S}}$ matches to the u-channel residue then we are done! Finally using the identity:[11]

$$\frac{\langle \mathbf{1}|P_I|\mathbf{4}]}{m} = m\frac{\langle \mathbf{4}3\rangle[\mathbf{1}2] + \langle \mathbf{1}3\rangle[\mathbf{4}2]}{\langle 3|p_1|2]}\,, \tag{2.97}$$

one derives the following ansatz for the four-point amplitude of minimally coupled general spin-S amplitude:

$$\frac{\langle 3|p_1|2]^{2-2S}}{(s-m^2)(u-m^2)}\left(\langle \mathbf{4}3\rangle[\mathbf{1}2] + \langle \mathbf{1}3\rangle[\mathbf{4}2]\right)^{2S}\,. \tag{2.98}$$

Note that the final result has an extra $(-)^{2S}$ sign for spin-S under $1 \leftrightarrow 4$ exchange. This tells us that charged half integer spins must be fermions, while integer spins are bosons. Thus we've recovered spin-statistics from the principles of Poincare symmetry and unitarity. The result in eq.(2.98) contains spurious singularities which cancel for $S \leq 1$. This signals that there is something fundamentally different for charged particles of $S \leq 1$ and $S > 1$. For $S = 1/2, 1$ we recover the Compton scattering:

$$\begin{aligned} M(\mathbf{1}^{\frac{1}{2}}, \gamma_2^{+1}, \gamma_3^{-1}, \mathbf{4}^{\frac{1}{2}}) &= \frac{\langle 3|p_1{-}p_4|2]}{2(s-m^2)(u-m^2)}\left(\langle \mathbf{4}3\rangle[\mathbf{1}2] + \langle \mathbf{1}3\rangle[\mathbf{4}2]\right) \\ M(\mathbf{1}^{1}, \gamma_2^{+1}, \gamma_3^{-1}, \mathbf{4}^{1}) &= \frac{\left(\langle \mathbf{4}3\rangle[\mathbf{1}2] + \langle \mathbf{1}3\rangle[\mathbf{4}2]\right)^2}{(s-m^2)(u-m^2)}\,. \end{aligned} \tag{2.99}$$

In appendix A.4 we reproduce this result using Feynman diagrams for fermions. By studying the H. E. limit, one can easily verify that this is correct. At H.E. for $S = 1$ one obtains three terms, two of which are contributions where legs 1 and 4 are opposite helicity gluons, and a final one which is when they are both scalars, which are the Goldstone bosons that were eaten in the Higgs mechanism! Note that this is telling us that the Higgs mechanism provides a way to "unify" the independent massless amplitudes in the IR. We will discuss this phenomenon in more detail in section 2.6.

Now in the above discussion the result from the s-channel gluing can be matched to the u-channel if we have a single species of spin-S. If there are multiple species, then

[11]This identity can be derived as follows: $|P^I\rangle[P_I|$ is the internal momentum that satisfies the s-channel on-shell constraint,

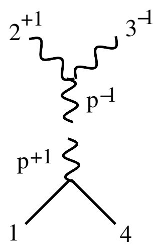
$$P_{\alpha\dot\alpha}\tilde\lambda_2^{\dot\alpha} = -mx_{12}\lambda_{2\alpha}, \quad P_{\alpha\dot\alpha}\tilde\lambda_3^{\dot\alpha} = mx_{34}\lambda_{3\alpha}, \quad P^2 = m^2 \tag{2.95}$$

The solution is given by:

$$P_{\alpha\dot\alpha} = \frac{-m^2\lambda_{3\alpha}\tilde\lambda_{2\dot\alpha} + (p_{1\alpha\dot\beta}\tilde\lambda_2^{\dot\beta})(p_{4\dot\alpha\beta}\zeta_3^{\beta})}{\langle 3|p_1|2]}\,. \tag{2.96}$$

Contracting with λ_4^I and $\tilde\lambda_1^I$ yields eq.(2.97).

similar to the massless discussion in section 2.3, we should assign a matrix T^a_{ij} to each vertex, and due to $[T^a, T^b] \neq 0$, the matching to the u-channel will be off by a piece that is proportional to $f^{abc}T^c_{ij}$. This mismatch is a sign that the t-channel pole from the s-channel factorisation should be assigned into a physical massless pole, i.e. revealing the presence of an non-abelian vector. For this to hold we should show that the s-channel residue admits this interpretation. Indeed taking a scalar for example, $\langle 3|p_1|2]^2/(s-m^2)t$ can be matched to the t-channel residue since



$$x_{14}\frac{\langle 3P\rangle^3}{\langle P2\rangle\langle 23\rangle} = -f^{ab}{}_c T^c \frac{\langle 3|p_1|3]\langle 3|p_1|P]}{x_{14}m\langle P2\rangle} = \langle 3|p_1|2]^2\left(\frac{T^aT^b}{s-m^2}+\frac{T^bT^a}{u-m^2}\right). \tag{2.100}$$

The last equality utilizes the fact that when $t=0$, $s-m^2 = -(u-m^2)$. Thus the final amplitude is given by:

$$\langle 3|p_1|2]^{2-2S}\left(\langle \mathbf{43}\rangle[\mathbf{12}]smatrix/+\langle \mathbf{13}\rangle[\mathbf{42}]\right)^{2S}\frac{1}{t}\left[\frac{(T^aT^b)_{ij}}{s-m^2}+\frac{(T^bT^a)_{ij}}{u-m^2}\right]. \tag{2.101}$$

Compton scattering for $S > 1$

The ansatz for general minimal coupling in eq.(2.98) appears to contain non-physical poles for $S > 1$. Of course this cannot be the final story since there's an abundance of charged higher spin-states in nature, and although we know that they are not fundamental, it has no bearing on the existence of S-matrix for low energy scattering. In deriving eq.(2.98), we started from the s-channel residue and analytically continued P_I to a form that is manifestly $2 \leftrightarrow 3$ and $+ \leftrightarrow -$ symmetric, and thus can be directly matched to u-channel residues. This is not entirely necessary, since the full amplitude can contain terms that only contain s and not u-channel pole. Thus the very fact that eq.(2.98) gives us non-physical poles for $S > 1$ is precisely telling us that such terms must be present.

To see this subtlety in detail, let's consider minimal coupling for spin-3/2, for which the gluing from s-channel yields:

$$-\frac{\langle 3|p_1|2]^2}{t}\left(\frac{\langle \mathbf{43}\rangle[\mathbf{12}]+\langle \mathbf{13}\rangle[\mathbf{42}]}{\langle 3|p_1|2]}\right)^3. \tag{2.102}$$

First note that by using eq.(2.15) one can rewrite the internal propagator in a mostly local form:

$$\frac{\langle\mathbf{43}\rangle[\mathbf{12}]+\langle\mathbf{13}\rangle[\mathbf{42}]}{\langle 3|p_1|2]} = \left(\frac{[\mathbf{14}]}{m}+\frac{\langle\mathbf{42}\rangle[2\mathbf{1}]-\langle\mathbf{12}\rangle[2\mathbf{4}]}{2m^2}\right)+\frac{t[2\mathbf{1}]\langle 3\mathbf{4}\rangle}{2m^2\langle 3|p_1|2]}\equiv A+B \tag{2.103}$$

The first two terms, denoted as A, are local at the expense of introducing extra inverse powers of m and are anti-symmetric under $1\leftrightarrow 4$, inheriting the symmetry properties of its parent. This guarantees that the local terms can be combined with the pre-factor and reproduce the correct residue on the u-channel pole. The last term, denoted as B, while being spurious, does not contribute to the u-channel residue and thus we are free to rewrite it in a local form using s-channel kinematics:

$$\left.\frac{t[2\mathbf{1}]\langle 3\mathbf{4}\rangle}{2m^2\langle 3|p_1|2]}\right|_{s=m^2} = -\frac{\langle\mathbf{4}3\rangle[32]\langle 2\mathbf{1}\rangle}{2m^3}. \tag{2.104}$$

Now expanding $(A+B)^3$, only the A^3 term will contribute to both s- and u- propagators, while terms with B will contribute solely to s-channel propagators. Putting everything together, one finds the following local form for the amplitude:

$$\begin{aligned} M(\mathbf{1}^{\frac{3}{2}},\gamma_2^{+1},\gamma_3^{-1},\mathbf{4}^{\frac{3}{2}}) \quad = \quad & \frac{\langle 3|p_1|2]^2}{(u-m^2)(s-m^2)}A^3-\left\{\frac{\langle 3|p_1|2][2\mathbf{1}]\langle 3\mathbf{4}\rangle}{2m^2(s-m^2)}\times\right. \\ & \left.\left(3A^2-3A\frac{\langle\mathbf{4}3\rangle[32]\langle 2\mathbf{1}\rangle}{2m^3}+\frac{\langle\mathbf{4}3\rangle^2[32]^2\langle 2\mathbf{1}\rangle^2}{4m^6}\right)+(1\leftrightarrow 4)\right\}. \end{aligned} \tag{2.105}$$

We now see that in the final local form, all terms contain $1/m$ factors and becomes singular in the H.E. limit. In other words, the obstruction of taking $m\to 0$ reflects the absence of a consistent massless high energy amplitude. For example the leading term in $1/m$ that will contribute to $M(1^{+\frac{3}{2}},\gamma_2^{+1},\gamma_3^{-1},4^{+\frac{3}{2}})$ at high energies is given by:

$$\frac{\langle 3|p_1|2]^2}{(u-m^2)(s-m^2)}\frac{[\mathbf{14}]^3}{m^3}\to\frac{\langle 31\rangle^2[12]^2}{us}\frac{[14]^3}{m^3}. \tag{2.106}$$

As we will elaborate below, this is the concrete sense in which charged particles with spin S $\geq 3/2$ cannot be "elementary", the same conclusion holds for any particles at all of spin S $\geq 5/2$ that can consistently couple to gravity.

Graviton Compton Scattering

Let us again begin with scalars, with the massive scalars are on legs $1,4$, a positive and negative helicity graviton on legs $2,3$ respectively. The s-channel residue is given as:

$$\frac{m^4}{M_{pl}^2}\frac{x_{12}^2}{x_{34}^2}=\frac{\langle 3|p_1|2]^4}{t^2M_{pl}^2}\,, \tag{2.107}$$

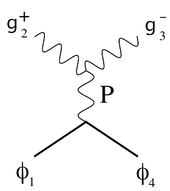
where M_{pl} is the Plank mass. As with the massless discussion we now have double pole in t, which can be identified as the massive pole $1/(u-m^2)$ and a massless $1/t$ pole. Thus the four-point amplitude is simply

$$-\frac{\langle 3|p_1|2]^4}{(s-m^2)(u-m^2)tM_{pl}^2}\,. \tag{2.108}$$

It is instructive to verify that the massless pole is correct. Let us take the residue at $t=0$, in the kinematics where $\langle ij\rangle = 0$. The residue of eq.(2.108) is

$$-\frac{\langle 3|p_1|2]^4}{\langle 3|p_1|3]\langle 2|p_1|2]M_{pl}^2}\,. \tag{2.109}$$

Since $\langle ij\rangle = 0$, the massless three-point amplitude should be $\overline{\text{M}HV}$, and one has



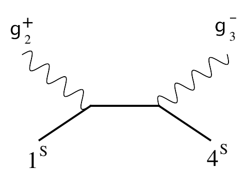
$$\frac{[2P]^6}{[23]^2[3P]^2}\frac{1}{x_{14}^2M_{pl}^2} = \frac{[2P]^4[2|p_1|P\rangle^2}{[23]^2[3P]^2m^2M_{pl}^2} = \frac{[2P]^2[2|p_1|3\rangle^2}{[3P]^2M_{pl}^2}, \tag{2.110}$$

where P is the massless internal momenta. Finally using the identity

$$\frac{[2P]^2}{[3P]^2} = \frac{[2P][2|p_1|P\rangle}{[3P][3|p_1|P\rangle} = -\frac{[2|p_1|3\rangle^2}{[3|p_1|2\rangle[2|p_1|3\rangle} = -\frac{[2|p_1|3\rangle^2}{\langle 3|p_1|3]\langle 2|p_1|2]}, \tag{2.111}$$

where in the last line, we've applied Schouten on the denominator, keeping in mind that $\langle 23\rangle = 0$. Thus we see that eq.(2.108) yields correct factorization in all channels.

For massive higher-spin particles, we again use the mixed representation. The s-channel residue yields:



$$\frac{x_{12}^2}{x_{34}^2}\frac{m^4}{M_{pl}^2}\left(\frac{P_I^{2S}}{m^{2S}}\right)_{\alpha_1\cdots\alpha_{2S},\dot\alpha_1\cdots\dot\alpha_{2S}}\,. \tag{2.112}$$

Using the explicit form for P_I in eq.(2.96), we find that the residue, after contracting with the external $(\lambda_1^I, \lambda_4^I)$ is simply

$$\frac{\langle 3|p_1|2]^4}{t^2M_{pl}^2}\left(\frac{\langle \mathbf{4}3\rangle[\mathbf{1}2]+\langle \mathbf{1}3\rangle[\mathbf{4}2]}{\langle 3|p_1|2]}\right)^{2S}\,. \tag{2.113}$$

Thus for $S \leq 2$ we find a perfectly local four-point amplitude given by:

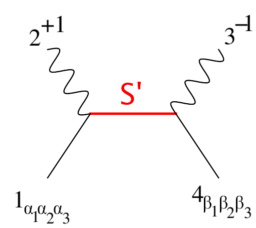
$$-\frac{\langle 3|p_1|2]^4}{(s-m^2)(u-m^2)tM_{pl}^2}\left(\frac{\langle \mathbf{4}3\rangle[\mathbf{1}2]+\langle \mathbf{1}3\rangle[\mathbf{4}2]}{\langle 3|p_1|2]}\right)^{2S}\,. \tag{2.114}$$

For $S>2$, we see that the formula ceases to be local. Similar to our photon coupling analysis, this indicates that the residue of s-channel must be separated into pieces that will combine with other channels and pieces that don't.

Massive higher spins cannot be elementary

We have seen that Compton scattering amplitudes for particles of high enough spin do not have a healthy high-energy limit, growing as powers of (p/m). Of course so long as the gauge/gravitational couplings are small, these amplitudes do not become $O(1)$ till energies parametrically above the particle mass m, so in that sense no inconsistency is encountered in the effective theory of a single massive higher spin particle till a cut-off parametrically above its mass. Nonetheless, the sickness of the $m \to 0$ limit does show that a single massive higher spin particle cannot be "elementary", and that any consistent theory for such particles must also include new particle states with a mass comparable to m. As an example, suppose we have some strongly-interacting QCD-like gauge theory; can such a theory have a spectrum consisting of bound states of high spin, with a parametrically large gap up to higher excited states? Our analysis suggests that this is impossible. We can imagine weakly gauging a global symmetry of the theory, or coupling the system to gravity. The total cross-section, for example, for $\gamma\gamma \to X$ should be bounded by $\sigma < C \times e^4/s$ for some constant C characterizing the current four-point amplitude. But if we have a charged higher spin particle, just the cross-section for its production would grow as $e^4/s \times (s/m^2)^n$, and if there is a parametrically large gap up to other particle states this will exceed the bound when $(s/m^2)^n > C$. Of course this is a somewhat qualitative argument, but we believe it captures the essence of why higher-spin massive particles must be composite. A sharpening of the argument may be able to give a more quantitative bound for the scale beneath which new particles must appear.

We can ask if the presence of new states in the propagator can tame this high-energy behaviour by cancelling the $1/m^6$ singularity in eq.(2.106). In other words consider the case where one has a new spin S' state with the similar mass as the $S = \frac{3}{2}$, then one can include the contribution:



$$. \tag{2.115}$$

If $S' \neq S$, then in the degenerate mass limit, it is easy to see that the three point amplitude cannot involve the pure x dependent pieces and thus the residue must be local. This then tells us that the contribution of such terms in the high energy limit must take the form $\frac{n_s}{sm^\alpha} + \frac{n_u}{um^\alpha}$ for some α, and n_s, n_u is some local function in kinematic invariants. This has a distinct high energy behaviour than eq.(2.106) which behaves as

$1/su$, and thus cannot cancel.[12] For $S' = S$, if the masses are not identical then the residue is again local and we have the same issue. If the masses are the same, then one simply obtains the exact same form as eq.(2.106) with identical signs, and the H.E. behaviour is again untamed.

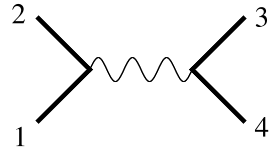
Thus even with finite number of states with comparable mass, the sick H.E. limit still rules out isolated charged higher spin state as a fundamental particle. The above analysis does provide a loophole: one can have an infinite tower of ever increasing higher spin states. While their presence in the propagator only produces terms with single poles in the H.E. limit, an infinite sum of n_s/s terms can produce poles in u if the degree of polynomial for n_s unbounded. That is, if the exchanged state has unbounded spin. This is precisely what happens for string theories which contain massive higher spin states.

All Possible Four Particle Amplitudes

Having discussed the four-particle amplitudes associated with the most familiar and important three-particle interactions, let us finally turn to computing *all* possible four-particle amplitudes. As we have seen when there are no "x" factors involved, we have local residues and the construction of four-particle amplitudes is trivial. We will therefore concentrate on discussing the cases where consistent factorization is non-trivial, which involve having at least one minimal coupling with an "x" factor, but now allowing for the most general set of other couplings. We will see (once again) that consistency demands that the minimal couplings have the standard Yang-Mills/gravitational forms, and that the other interactions have to be (globally) Yang-Mills invariant. But it *is* then possible to find consistently factorizing four-point amplitudes for any choice of three-particle interactions satisfying these conditions.

All Massive amplitude

This is the simplest, since we only need to consider the massless exchange. Consider the exchange of a massless-photon; for external scalars we have:



$$: \quad m^2\left(\frac{x_{12}}{x_{34}}+\frac{x_{34}}{x_{12}}\right)\,, \tag{2.116}$$

[12]Strictly speaking, due to our helicity choice eq.(2.106) really only has an s-channel pole at H.E. The bad H.E. behaviour in both channels will be present for other helicity configurations.

where the two terms correspond to the two different helicities. Using $x_{12}\lambda_P \equiv \frac{p_1}{m}|P]$, $x_{34}\lambda_P \equiv \frac{p_3}{m}|P]$ and $P = p_1 + p_2$, we find:

$$\frac{1}{s}m^2\left(\frac{x_{12}}{x_{34}} + \frac{x_{34}}{x_{12}}\right) = \frac{1}{s}\left(\frac{\langle\eta|p_1|P]\langle P|p_3|\xi]}{\langle\eta P\rangle[P\xi]} + \frac{\langle\eta|p_3|P]\langle P|p_1|\xi]}{\langle\eta P\rangle[P\xi]}\right) = \frac{2(p_1\cdot p_3)}{s}, \quad (2.117)$$

where one uses the fact that $\langle P|p_i|P] = 0$ for any external momenta p_i. This is not the complete answer, as one expects $\frac{(p_1\cdot p_3)-(p_2\cdot p_3)}{s}$ from minimal coupling. The difference is s/s and thus has no factorization poles. The correct answer can be inferred from symmetry arguments under $1 \leftrightarrow 2$ exchange. Thus the correct completion is

$$\frac{1}{s}m^2\left(\frac{x_{12}}{x_{34}} + \frac{x_{34}}{x_{12}}\right) = \frac{(p_1-p_2)\cdot p_3}{s}, \quad (2.118)$$

For the exchange of a general massless spin S state, we simply get a factor of $((p_1-p_2)\cdot p_3)^S$ for the numerator.

Now we let the external particles carry spin. For simplicity we will consider the case where all four particles are of the same spin. Then the residue for the most general coupling is given by:

$$\begin{aligned} &\frac{x_{12}}{m^{2S-2}}\left[\langle\mathbf{12}\rangle^{2S} + \sum_{i=0}^{2S-1}\left(a_i\langle\mathbf{12}\rangle^i\left(\frac{\langle\mathbf{1}P\rangle[\mathbf{2}P]}{m}\right)^{2S-i} + \tilde{a}_i\langle\mathbf{12}\rangle^i\left(\frac{\langle\mathbf{2}P\rangle[\mathbf{1}P]}{m}\right)^{2S-i}\right)\right] \\ \times \quad &\frac{1}{x_{34}}\left[[\mathbf{34}]^{2S} + \sum_{i=0}^{2S-1}\left(b_i[\mathbf{34}]^i\left(\frac{\langle\mathbf{3}P\rangle[\mathbf{4}P]}{m}\right)^{2S-i} + \tilde{b}_i[\mathbf{34}]^i\left(\frac{\langle\mathbf{4}P\rangle[\mathbf{3}P]}{m}\right)^{2S-i}\right)\right]. \end{aligned} \quad (2.119)$$

where $a_i, b_i, \tilde{a}_i, \tilde{b}_i$ parameterize all possible coupling to the photon, and for parity invariant theories we have $a_i = b_i$ and $\tilde{a}_i = \tilde{b}_i$. Since besides the leading term in the square brackets, each of the terms contains $|P]\langle P|$ which can readily convert $\frac{x_{12}}{x_{34}}$ into local forms, and thus we only need to worry about the term

$$\frac{1}{m^{2S-2}}\left(\frac{x_{12}}{x_{34}}\langle\mathbf{12}\rangle^{2S}[\mathbf{34}]^{2S} + \frac{x_{34}}{x_{12}}[\mathbf{12}]^{2S}\langle\mathbf{34}\rangle^{2S}\right), \quad (2.120)$$

where we've included the contribution where the photon helicity is flipped. Finally, using the identity:

$$[\mathbf{12}] = \langle\mathbf{12}\rangle + \frac{\langle\mathbf{1}|P|\mathbf{2}]}{m}, \quad (2.121)$$

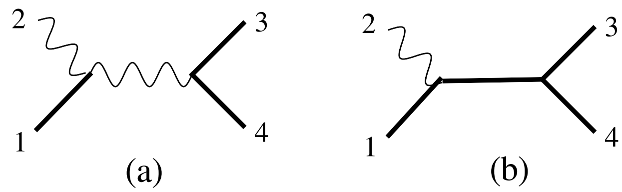
introduces $|P]\langle P|$ that can again be used to absorb the x-factors leaving behind

$$\frac{\langle\mathbf{12}\rangle^{2S}\langle\mathbf{34}\rangle^{2S}}{m^{2S-2}}\left(\frac{x_{12}}{x_{34}} + \frac{x_{34}}{x_{12}}\right) = \frac{\langle\mathbf{12}\rangle^{2S}\langle\mathbf{34}\rangle^{2S}}{m^{2S}}(p_1-p_2)\cdot p_3, \quad (2.122)$$

where we've used eq.(2.118). Thus we see that the massless gluing of any three point vertex can be converted into a local form. For more general external spins, the analysis is the same albeit more complicated.

Three-massive one-massless

If we have three-massive legs, the dangerous x-factors can occur in two types of diagrams for the s-channel residue:



$$. \tag{2.123}$$

Let us first consider the case where the solid lines are massive scalars, and the wavy line is the positive helicity photon. Diagram $(a),(b)$ gives:

$$(a) \quad \frac{[2P]^2}{x_{34}} = \frac{[2P][2|p_3|P\rangle}{m}, (b) \quad m x_{1P} = \frac{[2|p_1|\xi\rangle}{\langle 2\xi\rangle}\,. \tag{2.124}$$

The first is manifestly local. For the second, let's consider the all massive vertex being $\phi\phi'\phi''$ vertex, and the photon only couples to ϕ and ϕ' with coupling e, e'. Then gluing leads to:

$$e\frac{[2|p_1|\xi\rangle}{\langle 2\xi\rangle(s-m^2)} + e'\frac{[2|p_4|\xi\rangle}{\langle 2\xi\rangle(u-m^2)} = (e+e')\frac{[2|p_4|\xi\rangle}{\langle 2\xi\rangle(u-m^2)} + e\frac{[2|p_4p_1|2]}{(s-m^2)(u-m^2)}\,, \tag{2.125}$$

where legs $1,4$ are ϕ,ϕ' respectively. We see that only when the charge is conserved, i.e. $e+e'=0$ does the $\langle 2\xi\rangle$ pole cancels and the amplitude becomes local. If the scalars were all charged with charges e,e',e'', the same analysis would tell us that $e+e'+e''=0$. Next suppose the photon was instead a gluon, with the scalars carry indices i,i',i'' and the three-point amplitude given by $c_{ii'i''}$. We have already seen that consistency demands the couplings to the gluons $T^a_{ij}, T^a_{i'j'}, T^a_{i''j''}$ be generators in some representation of the Yang-Mills group. Then we discover that we must have $T^a_{ij}c_{ji'i''} + T^a_{i'j'}c_{ij'i''} + T^a_{i''j''}c^{ii'j''} = 0$, in other words the cubic interaction must be invariant under the (global) Yang-Mills symmetry. Finally, for graviton, gluing to a ϕ^3 vertex leads to:

$$g_1\frac{[2|p_1|\xi\rangle^2}{M_{pl}\langle 2\xi\rangle^2(s-m^2)} + g_3\frac{[2|p_3|\xi\rangle^2}{M_{pl}\langle 2\xi\rangle^2(t-m^2)} + g_3\frac{[2|p_4|\xi\rangle^2}{M_{pl}\langle 2\xi\rangle^2(u-m^2)}\,, \tag{2.126}$$

where we've let all three scalars couple to gravity. Again after rearranging the terms, one finds that the auxiliary spinor drops out only if $g_1=g_2=g_3$, and one arrives at:

$$\frac{g_1}{M_{pl}}\frac{[2|p_1p_3|2]^2}{(s-m^2)(u-m^2)(t-m^2)}\,. \tag{2.127}$$

Thus we see that coupling to photons, the consistency of the four-point amplitude requires charge to be conserved, for a gluon it requires the particles to be in the adjoint representation, and finally for a graviton, it leads to the equivalence principle. Note that this discussion does not refer to any gauge redundancy and the independence there of. On the other hand, the astute reader will recognize that the factor $\frac{[2|p_1|\xi\rangle}{\langle 2\xi\rangle}$ can be identified with $\epsilon_2 \cdot p_1$ from Feynman rules, where λ_ξ is the reference spinor for the polarization vector ϵ_2. Indeed from the photon and graviton soft-theorem [123, 170], it is precisely this factor whose gauge invariance (Ward identity) demands the conservation of charges and equivalence principle. Here, there's no gauge redundancy, the auxiliary spinor λ_ξ is simply a projection of eq.(2.61), and the independence thereof is the requirement that factorization is consistent to all solutions of x defined through eq.(2.61).

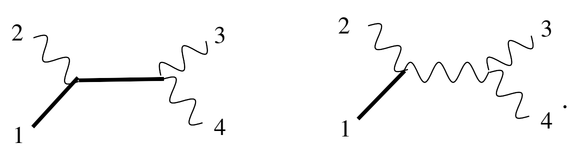
Again the same applies if we consider external spinning particles. For example for massive spin-1, diagram (a) yields,

$$\begin{aligned}
(a) \qquad & \frac{[2P]^3\langle \mathbf{1}2\rangle\langle 1P\rangle}{m^5}\frac{1}{x_{34}}\left(\langle \mathbf{34}\rangle - x_{34}\frac{\langle \mathbf{3}P\rangle\langle \mathbf{4}P\rangle}{m}\right)^2 \\
= \; & [2P]^2\langle \mathbf{1}2\rangle\langle 1P\rangle\frac{[2|p_3|P\rangle}{m^6}\left(\langle \mathbf{34}\rangle - \frac{\langle \mathbf{4}|p_3|P]\langle \mathbf{3}P\rangle}{m^2}\right)^2 , \qquad (2.128)
\end{aligned}$$

where again the residue is local. For diagram (b) the only non-locality originates from the minimal coupling piece, and hence one recovers the same condition as before.

One-massive three-massless

So far we have found that all potential non-localities can be converted into local expressions, and hence the residue of one-channel does not encode information with respect to other channels. For three massless particles things are more interesting. The potential s-channel factorization diagrams are:



(a) (b) . (2.129)

For our purpose, only minimal coupling is relevant for the two massive one massless vertex in (a). We will consider a massive scalar coupled to abelian and non-abelian vectors.

First for the abelian case we only need to consider diagram (a). Taking all vectors to be plus helicity, one finds the s-channel residue given by

$$(a) \quad x_{12}[34]^2 = \frac{\langle 3|p_1|2][34]^2}{m\langle 23\rangle} = \frac{m[42][34]}{\langle 23\rangle} \, . \qquad (2.130)$$

The appearance of $\langle 23\rangle$ seems to indicate an illegal massless pole. However since $s = m^2$, this can be identified as a u-channel massive pole, $u - m^2 = -t$. Thus one can write the amplitude as

$$\frac{m[42][34][23]}{(s-m^2)(u-m^2)}. \tag{2.131}$$

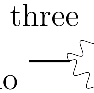
Note however the extra $-$ sign under the $2 \leftrightarrow 3$ exchange will lead to the violation of spin-statistics for identical vectors. Thus we see that minimally coupled scalars are incompatible with a di-photon coupling. Indeed from the action point of view, this is simply the statement that the U(1) symmetry of a charged scalar forbids the appearance of ϕF^2 coupling. Thus there is no such four-particle amplitude for the abelian theory. For the non-abelian case, one must also consider diagram (b), which yields

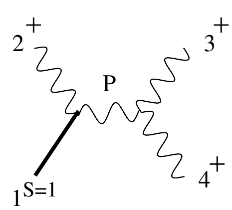
$$(b): \quad \frac{[2P]^2}{m}\frac{[34]^3}{[3P][P4]} = \frac{m^3[34]}{\langle 23\rangle\langle 24\rangle}\,. \tag{2.132}$$

We gain find the illegal pole $1/\langle 24\rangle$. Since we are considering the non-abelian theory we can consider the colour stripped amplitude and convert the spurious pole into a legal t-channel massive pole. This suggests us to write

$$M_4(\mathbf{1}, 2^+, 3^+, 4^+) = \frac{m^3[24][23][34]}{st}\left(\frac{1}{(t-m^2)} + \frac{1}{(s-m^2)} + \frac{1}{m^2}\right), \tag{2.133}$$

where we've added the massless t-channel image, and the extra $1/m^2$ is to guarantee that both massless channels factorises correctly. One can check the s-channel massive residue, which was given in eq.(2.130), matches when taken into account that $\langle 34\rangle = \frac{m^2}{[43]}$. Note that the amplitude vanishes as $m \to 0$ as it should.

Now let's move on to the case where there are external spins. For example, one can consider a massive spin-1 particles couple to three massless vectors. If the vector is abelian, Yang's theorem tells us that there is no  vertex to consider, and thus there are no factorizable four-point amplitude to consider. We instead begin with a massive vector and three gluons. We will start with colour stripped all plus-helicity gluons, whose residue for the massless s-channel is given as:

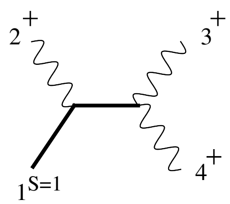


$$\rightarrow \quad \langle \mathbf{1}2\rangle\langle \mathbf{1}P\rangle\frac{[P2]^3}{m^4} \times \frac{[34]^3}{[4P][P3]}\,. \tag{2.134}$$

Now since the vertex on one side contains high power of momenta, there are different ways of rewriting this residue which are equivalent on the kinematics $\langle 34\rangle = 0$. We will choose the representation where one separates the various pieces that contain information on other channels. More precisely, we have:

$$\langle\mathbf{1}2\rangle\langle\mathbf{1}P\rangle\frac{[P2]^3}{m^4}\times\frac{[34]^3}{[4P][P3]}=\langle\mathbf{1}2\rangle\langle\mathbf{1}P\rangle[P2][34]\frac{([P3][42]+[P4][23])^2}{m^4[4P][P3]}$$
$$=[34]\left(\frac{2[\mathbf{1}|p_2|\mathbf{1}\rangle[42][23]}{m^3}+\frac{[42]\langle\mathbf{1}|p_4p_2|\mathbf{1}\rangle}{m^2\langle 23\rangle}+\frac{[23]\langle\mathbf{1}|p_2p_3|\mathbf{1}\rangle}{m^2\langle 24\rangle}\right) \tag{2.135}$$

where the last equality sign is understood to hold on $\langle 34\rangle=0$ kinematics. We see that unavoidably there is an $1/\langle 24\rangle$ pole in the s-channel massless residue, which is spurious unless it can be interpreted as a t-channel pole $1/(t-m^2)$. Thus the massless residue for the amplitude tells us that there must be a two massive vector, one gluon matrix element that must be present to explain the apparent spurious singularity. The contribution of this matrix element for the s-channel is given by:



$$\rightarrow\left(x_{12}\frac{\langle\mathbf{1P}\rangle^2}{m}\times\langle\mathbf{P}3\rangle\langle\mathbf{P}4\rangle\frac{[34]^3}{m^4}\right)=\frac{[42][34]\langle\mathbf{1}|p_3p_4|\mathbf{1}\rangle}{\langle 32\rangle m^2}\,. \tag{2.136}$$

This suggests that we begin with the following piece which factorises correctly on the s and t-channel massive pole:

$$\frac{[42][34]\langle\mathbf{1}|p_3p_4|\mathbf{1}\rangle}{\langle 32\rangle(s-m^2)m^2}-\frac{[42][32]\langle\mathbf{1}|p_3p_2|\mathbf{1}\rangle}{\langle 34\rangle(t-m^2)m^2}\,. \tag{2.137}$$

Note that the above is symmetric in $(2\leftrightarrow 4)$ and contains $\langle 34\rangle$, $\langle 23\rangle$ poles as well. Taking $\langle 34\rangle\to 0$, only the second term in eq.(2.137) contributes to its residue:

$$Res\left[-\frac{[42][32]\langle\mathbf{1}|p_3p_2|\mathbf{1}\rangle}{\langle 34\rangle(t-m^2)m^2}\right]\Bigg|_{\langle 34\rangle=0}=\frac{[23]\langle\mathbf{1}|p_2p_3|\mathbf{1}\rangle}{\langle 24\rangle m^2}\,. \tag{2.138}$$

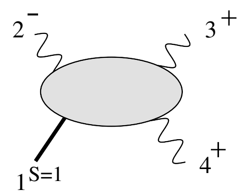
This is nothing but the spurious residue appearing in eq.(2.135)!

Putting the information built from the s- and t-channel massive, and s-channel massless residue together, leads to:

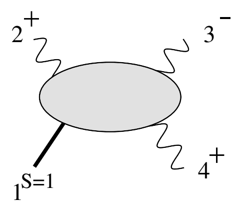
$$\begin{aligned}M(\mathbf{1}^{S=1}2^+3^+4^+) &= \frac{[42][23][34]}{m^2}\left\{\frac{1}{t}\left(\frac{\langle\mathbf{1}|p_3p_4|\mathbf{1}\rangle}{(s-m^2)}+\frac{2\langle\mathbf{1}|p_1p_4|\mathbf{1}\rangle}{m^2}\right)+\frac{\langle\mathbf{1}|p_4p_2|\mathbf{1}\rangle}{st}\right.\\ &+\left.\frac{1}{s}\left(\frac{\langle\mathbf{1}|p_3p_2|\mathbf{1}\rangle}{(t-m^2)}+\frac{2\langle\mathbf{1}|p_1p_2|\mathbf{1}\rangle}{m^2}\right)\right.\,.\end{aligned} \tag{2.139}$$

The matching to the massless t-channel is straight forward given that the above is symmetric in (2↔3). Note that unlike the uniqueness of the one massive two massless amplitude, a priori the coupling between the two massive and one massless vector does not have to match that of minimal coupling. It is the consistency between the massless and massive factorisation that fixes this choice. A quick recap: beginning with the massless residue, for which the three-point coupling involving the massive spin-1 is unique, the anti-symmetric property with respect to the massless legs tells us that the massive state must be in adjoint rep of the color group. Then the presence of an $1/\langle 24\rangle$ singularity becomes spurious unless it arises from the massive propagators evaluated on degenerate kinematics. Thus the massless residue in one channel encodes the massive residue in the other.

For the other helicity components, the derivation is simpler as one can construct the full amplitude from the residue of the massive channel, and we simply list the results:



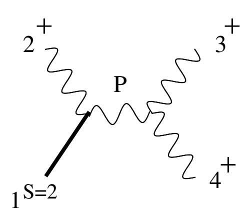
$$\rightarrow \frac{[3|p_1|2\rangle\langle 23\rangle[34][\mathbf{13}][\mathbf{14}]}{m^4 t(s-m^2)}$$



$$\rightarrow \frac{[\mathbf{1}2][\mathbf{1}4][24]^2\langle 32\rangle\langle 43\rangle}{stm^2}\,. \tag{2.140}$$

In the first line, we've listed the amplitude in the dotted frame for simplicity. One can check that the leading contribution for the H.E. limit of this amplitude yields the amplitude generated by the $tr(F^3)$ extension of Yang-Mills theory.

As a final example, let's consider a possible singlet massive spin-2 particle that interacts with gluons via a higher-dimensional operator RF^2. For the one massive three positive-helicity gluon amplitude, we expect that the final result is cyclic invariant in $(2,3,4)$. The massless s-channel residue can now be written as



$$\rightarrow \langle \mathbf{1}2\rangle^2\langle \mathbf{1}P\rangle^2\frac{[P2]^4}{m^7}\times\frac{[43]^3}{[4P][P3]}$$

$$= \frac{[43][\mathbf{1}|p_2|\mathbf{1}\rangle^2}{m^3}\left(\frac{[42]}{\langle 23\rangle}+\frac{[32]}{\langle 24\rangle}\right)\,, \tag{2.141}$$

where we've suppressed the symmetrized SL(2,C) indices, keeping in mind they are distributed amongst the $(p_i\cdot p_j)$s. Putting back the massless propagator, this suggest that we start with:

$$\frac{[\mathbf{1}|p_2|\mathbf{1}\rangle^2}{m^3\langle 34\rangle}\left(\frac{[42]}{\langle 23\rangle}+\frac{[32]}{\langle 24\rangle}\right). \tag{2.142}$$

The above result contains other additional poles, which under cyclic rotation $(2,3,4)$, will generate terms that will modify the original $\frac{1}{\langle 34\rangle}$ residue. Thus before summing over its cyclic image, we should augment eq.(2.142) with terms that kill the extra poles in $\langle 24\rangle$ and $\langle 23\rangle$. Putting everything together, we find:

$$\begin{aligned} M(\mathbf{1}^{S=2}2^+3^+4^+) \;&=\; \frac{[\mathbf{1}|p_2|\mathbf{1}\rangle^2}{m}\frac{[43][32][42]}{stu} \\ &-\; \frac{[2\mathbf{1}]^2}{m^3}\left(\frac{\langle 24\rangle[43][32]}{st}[4\mathbf{1}]^2+\frac{\langle 23\rangle[43][42]}{su}[3\mathbf{1}]^2\right) \\ &+\; cyclic(2,3,4)\,. \end{aligned} \tag{2.143}$$

We give further examples of massive amplitudes involving one massive higher spin and non-identical spin massless particles in appendix A.6.

The spinning polynomial basis

The fact our on-shell formalism provides a convenient basis to classify distinct three-point couplings lends itself to another important application: construction of a basis polynomial to expand the four-point amplitude. A well known example for such a polynomial is the Gegenbauer polynomial, or its four dimensional representation the Legendre polynomial, as a basis for the four-point scalar amplitude. The Gegenbauer polynomials arises from the exchange of a spin-S particle for a four scalar amplitude. Note that we have one polynomial for a given S because the three-point coupling between two scalars and a spin-S particle is fixed.

As we've seen in the previous discussion, the three-point amplitude for one massive, two massless particles is also unique. This implies that we can similarly construct "spinning" Gegenbauer polynomials for massless scattering amplitude, where each polynomial correspond to a different spin exchange. To see how this works let's consider the residue for a spin-S exchange in the s-channel for $M(1^{-h}2^{+h}3^{-h}4^{+h})$. We can write down the unique three-point amplitudes on both sides:

$$\frac{\lambda_1^{S+2h}\lambda_2^{S-2h}[12]^S}{m^{2S-1}},\quad \frac{\lambda_3^{S+2h}\lambda_4^{S-2h}[34]^S}{m^{2S-1}}.$$

Such coupling only exists for $S\geq 2h$. Now when we glue the two tensor structures together the indices on λ_1,λ_2 must be fully contracted with those on λ_3,λ_4. This can be done in many ways, each with its own pre-factor counting the number of equivalent contractions. The gluing procedure is thus a sum over all possible contractions with suitable combinatoric factors:

$$\frac{[12]^S[34]^S}{m^{4S-2}}\sum_a c_{2S,S+2h,a}\langle 13\rangle^a\langle 14\rangle^{S+2h-a}\langle 23\rangle^{S+2h-a}\langle 24\rangle^{a-4h} \tag{2.144}$$

where the summation a ranges from $4h$ to $S+2h$, and

$$c_{2S,S+2h,a}=\frac{(S+2h)!^2(S-2h)!^2}{a!(S+2h-a)!^2(a-4h)!}\,,. \tag{2.145}$$

It would be useful to convert this polynomial into a function of the scattering angle in the center of mass frame for particles 1 and 2. We write $\lambda_1=m^{\frac{1}{2}}\begin{pmatrix}1\\0\end{pmatrix}$, $\lambda_2=m^{\frac{1}{2}}\begin{pmatrix}0\\1\end{pmatrix}$, $\lambda_3=im^{\frac{1}{2}}\begin{pmatrix}\cos\frac{\theta}{2}\\ \sin\frac{\theta}{2}\end{pmatrix}$, $\lambda_4=im^{\frac{1}{2}}\begin{pmatrix}\sin\frac{\theta}{2}\\ -\cos\frac{\theta}{2}\end{pmatrix}$. The spinning Gegenbauer polynomial is then given as:

$$P_S^h(\cos\theta)=\frac{1}{(S!)^2}\sum_a\frac{(S+2h)!^2(S-2h)!^2}{a!(S+2h-a)!^2(a-4h)!}\left(\frac{\cos\theta-1}{2}\right)^{a-2h}\left(\frac{\cos\theta+1}{2}\right)^{S+2h-a}. \tag{2.146}$$

As a few example (with $x=\cos\theta$):

$$\begin{aligned}P_2^1(x)&=\frac{3}{2}(x-1)^2,\quad P_3^1(x)=\frac{5}{6}(x-1)^2(2+3x)\\ P_4^1(x)&=\frac{5}{8}(x-1)^2(1+7x(1+x))\,.\end{aligned} \tag{2.147}$$

The universal prefactor $(x-1)^2$ can be identified with $\langle 13\rangle^2[24]^2$ which takes care of the overall helicity weights of this amplitude. Taking $\ell=0$ we indeed recover the Legendre polynomials $P_S^0(x)=P_S(x)$.

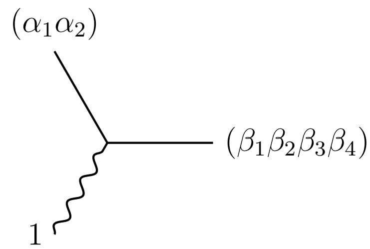
For completely general helicities h_1,h_2,h_3,h_4 of external massless particles, we have:

$$\begin{aligned}P_S^{h_i}(x)=&\frac{1}{(S!)^2}\sum_a\frac{(S+h_4-h_3)!(S+h_3-h_4)!(S+h_1-h_2)!(S+h_2-h_1)!}{a!(S+h_4-h_3-a)!(S+h_2-h_1-a)!(a+h_1+h_3-h_2-h_4)!}\\ &\times\left(\frac{x-1}{2}\right)^{a+\frac{h_1+h_3-h_2-h_4}{2}}\left(\frac{x+1}{2}\right)^{S-a-\frac{h_1+h_3-h_2-h_4}{2}}.\end{aligned}$$

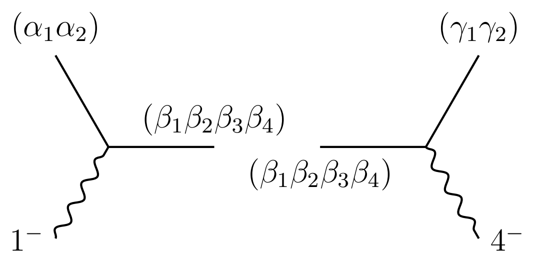
This reduces to equal spin polynomial if we take all $|h_i|$ to be equal.

Three-point couplings with more than one massive leg are no longer unique. This means that for a given spin-exchange, one instead has a symmetric matrix where the rows and the columns label the independent three-point vertices on both sides of the factorization channel. We illustrate this for the two massive spin-1 and two massless spin-1 amplitude.

Now the three-point coupling involved in the factorization involves a massive spin-1 spin-S and massless spin-1 amplitude. The number of such coupling is determined by the lowest spin massive particle, which in this case is 1 and there are three independent coupling. To give an explicit example, consider S=2



The building blocks of tensor structures will be $\{\lambda_1, P_2\tilde{\lambda}_1\} = \{v, u\}$. If the massless particle has $-$ helicity, we have three tensor structures listed in eq.(2.58). Now imagine gluing the two three-point amplitude:



The residue will be a polynomial of (u_L, v_L, u_R, v_R) with

$$u_L^\alpha = \epsilon^{\alpha\beta}(P_2)_{\beta\dot{\beta}}\tilde{\lambda}_1^{\dot{\beta}}, \quad v_L^\alpha = \lambda_1^\alpha$$
$$u_R^\alpha = \epsilon^{\alpha\beta}(P_3)_{\beta\dot{\beta}}\tilde{\lambda}_4^{\dot{\beta}}, \quad v_R^\alpha = \lambda_4^\alpha.$$

By gluing them we contract the internal indices in all possible ways, then sum them up with appropriate combinatoric factors. We can distribute indices carried by exchanged particle into a bunch of u's and v's:

$$\begin{aligned} \#(u_L) + \#(v_L) &= 2S \\ \#(u_R) + \#(v_R) &= 2S. \end{aligned}$$

where S is the spin of exchanged particle. For a contraction with $(u_L)^{k_1}$ and $(u_R)^{k_2}$ on exchanged leg, suppose u_L and u_R are contracted together k_3 times. Then we have

$$\langle u_L u_R\rangle^{k_3}\langle u_L v_R\rangle^{k_1-k_3}\langle v_L u_R\rangle^{k_2-k_3}\langle v_L v_R\rangle^{2N-k_1-k_2+k_3} \tag{2.148}$$

which means a factor of

$$\binom{k_2}{k_3}\binom{2N-k_2}{k_1-k_3}(k_1)!(2N-k_1)!. \tag{2.149}$$

The first two factors come from choosing which u_Ls and u_Rs are to be contracted together. Since we can always redefine coupling constants for interactions, the k_3-independent factors shall not concern us here. Summing this factor over k_3 one gets $(2N)!$, the total number of permutations on $2N$ indices.

Assigning a coupling constant g_i for each three-point vertex, the residue of the four-point amplitude can then be expanded as $g_i M_{ij} g_j$ where each element in M_{ij} is a polynomial given by the contraction of the corresponding three-point amplitudes. Since we have two external spin-1 particles, M_{ij} is a 3×3 symmetric matrix irrespective of the exchanged spin. For the case where one exchanges a spin-2, the matrix elements are given by:

$$\begin{aligned}
M_{11} &= 24\langle v_L v_R\rangle^4\langle u_L\mathbf{1}\rangle^2\langle u_R\mathbf{4}\rangle^2\\
M_{12} &= 24\langle v_L u_R\rangle\langle v_L v_R\rangle^3\langle u_L\mathbf{1}\rangle^2\langle u_R\mathbf{4}\rangle\langle v_R\mathbf{4}\rangle\\
M_{13} &= 24\langle v_L u_R\rangle^2\langle v_L v_R\rangle^2\langle u_L\mathbf{1}\rangle^2\langle v_R\mathbf{4}\rangle^2\\
M_{22} &= \left(18\langle u_L v_R\rangle\langle v_L u_R\rangle\langle v_L v_R\rangle^2 + 6\langle u_L u_R\rangle\langle v_L v_R\rangle^3\right)\langle u_L\mathbf{1}\rangle\langle v_L\mathbf{1}\rangle\langle u_R\mathbf{4}\rangle\langle v_R\mathbf{4}\rangle\\
M_{23} &= \left(12\langle u_L v_R\rangle\langle v_L u_R\rangle^2\langle v_L v_R\rangle + 12\langle u_L u_R\rangle\langle(\langle v_L u_R\rangle\langle v_L v_R\rangle^2\right)\langle u_L\mathbf{1}\rangle\langle v_L\mathbf{1}\rangle\langle v_R\mathbf{4}\rangle^2\\
M_{33} &= \left(4\langle u_L v_R\rangle^2\langle v_L u_R\rangle^2 + 16\langle u_L u_R\rangle\langle u_L v_R\rangle\langle v_L u_R\rangle\langle v_L v_R\rangle + 4\langle u_L u_R\rangle^2\langle v_L v_R\rangle^2\right)\\
&\quad\cdot\langle v_L\mathbf{1}\rangle^2\langle v_R\mathbf{4}\rangle^2\,,
\end{aligned} \tag{2.150}$$

where we've contracted each entry with the external λ_1^I, λ_4^Is.

For convenience, we will also give the representation in terms of scattering angle. We can parameterize the kinematics as

$$\begin{aligned}
p_1 &= (x,0,0,x)\\
p_2 &= (\sqrt{x^2+m_2^2},0,0,-x)\\
p_3 &= (-\sqrt{y^2+m_3^2},-y\sin\theta,0,-y\cos\theta)\\
p_4 &= (-y,y\sin\theta,0,y\cos\theta),
\end{aligned}$$

where

$$x=\sqrt{\frac{(m_2^2+t)^2}{-4t}},\quad y=\sqrt{\frac{(m_3^2+t)^2}{-4t}}.$$

One can explicitly check that $\sum_i p_i = 0$, $p_i^2 = m_i^2$. In this parametrization, the matrix elements then take the form, where we've stripped the external spinor dependent terms:

$$\begin{pmatrix} 6+12x+6x^2 & 12(1+x)\sqrt{1-x^2} & 24-24x^2 \\ 12(1+x)\sqrt{1-x^2} & 12-12x-24x^2 & -48x\sqrt{1-x^2} \\ 24-24x^2 & -48x\sqrt{1-x^2} & -32+96x^2 \end{pmatrix}. \tag{2.151}$$

2.6 (Super)Higgs Mechanism as IR Unification

Our exploration of consistent four-particle amplitudes has given us an almost complete understanding of the broad architecture of particle physics. Theories of massless particles are incredibly constrained, allowing only helicities (0,1/2,1,3/2,2), and limited to the (super)gravity coupled to (super)Yang-Mills theories. Massless higher spins are made impossible by the mere presence of gravity. We have also seen that the amplitudes for massive particles of sufficiently high spin have sick high-energy limits—as expected, since there is no consistent theory of massless high-spin particles they can match to at high-energies—so such particles cannot be "elementary".

The final case to consider is then that of massive particles of low spin $S \leq 2$. Here of course there *is* in principle a consistent high-energy theory to match to, but as we will see in this section, doing so puts non-trivial restrictions on the particle content and interactions of the theory. This investigation will lead to the on-shell discovery of the Higgs and Super-Higgs mechanisms.

Note that we will not simply be rephrasing well-known "bottom-up" facts, such as the high-energy growth of scattering amplitudes for longitudinal components of massive spin one particles, and the attendant need for the Higgs particle to tame this growth, in an on-shell language. It is of course perfectly possible to do this, and the on-shell methods do simplify the explicit computations, but the advantage is purely technical and does not add anything conceptually new to this standard textbook discussion.

We will instead take a different, "top-down" point of view, where as described above we insist that massive amplitudes manifestly match to consistent massless amplitudes in the high-energy limit. As we will see this gives us a satisfying understanding of the Higgs mechanism that is at least psychologically quite opposite to the usual picture of gauge symmetry "breaking". Indeed in textbook language, the gauge symmetry is "broken" or "hidden", and becomes more manifest only at high energies. By contrast in the on-shell picture, the massive "Higgsed" amplitudes do not "break" or "hide" the (non-existent in this formalism!) gauge redundancies. Instead, they unify the different helicity components of massive amplitudes, and thus the Higgs mechanism can be thought of as an *infrared unification* of massless amplitudes, and this unification is more disguised at high energies!

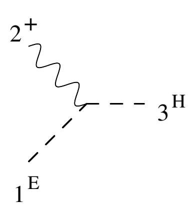
We will see this beginning already at the level of three-particle amplitudes. Here, the non-locality associated with the poles in massless three-particle amplitudes gets IR-deformed to $1/m$ poles. Such $1/m$ poles non-trivially disappear in the high-energy limit while the massive amplitudes unify different helicity components together. Matching the high-energy limit enforces all the usual consistency conditions associated with the Higgs mechanism. Moving on to four-particle amplitudes, we will obtain them both by gluing the three-particle amplitudes as usual, but also in a novel way, starting with the massless helicity amplitudes, simply adding them so they fit into massive multiplets, then shifting the poles and "**BOLD**"ing the spinor-helicity variables to make massive amplitudes! This will highlight the Higgs mechanism as an "IR unification" in an even more vivid way.

Rather than present a completely systematic analysis of all possible "Higgsings", in this section we will content ourselves with illustrating this physics in three standard examples: the Abelian Higgs model, the Super-Higgs mechanism in a simple model with $\mathcal{N} = 1$ SUSY, and the general structure of the non-Abelian Higgs mechanism for a model with enough scalars so that all the spin one particles are massive. As alluded to above we will also discuss why gravity cannot be Higgsed in this way.

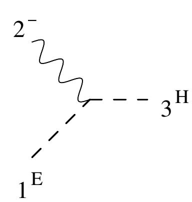
Abelian Higgs

Let us start with the simplest example - a theory with a massless photon and a charged scalar; we'll call the scalar's two real degrees of freedom "H" and "E".

The three-particle amplitudes are

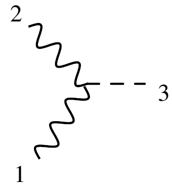


$$g\frac{[12][32]}{[13]},$$



$$g\frac{\langle 12\rangle\langle 32\rangle}{\langle 13\rangle}. \qquad (2.152)$$

We now want to see how to introduce masses as an "infrared deformation". The first step is a trivial kinematical one. We declare that $(+, -, E)$ are to become the 3 components of a massive spin 1 particle, leaving H as an additional scalar. Now, the two massive vector (with m_γ^2) and one massive scalar (with m_H^2) amplitude can only be,

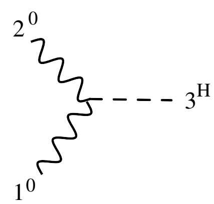


$$\frac{g}{m_\gamma}\langle\mathbf{12}\rangle[\mathbf{21}]\,. \tag{2.153}$$

The coefficient is fixed by the requirement that this 3 particle amplitude matches the massless amplitude in the high-energy limit. It is illuminating to see how this happens explicitly. Recall that to take the HE limit we put

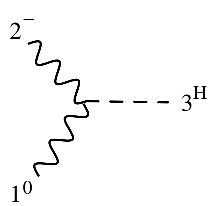
$$\begin{aligned}\lambda_\alpha^I &= \lambda_\alpha \xi_+^I + \eta_\alpha \xi_-^I \\ \tilde{\lambda}_{\dot\alpha}^I &= \tilde{\lambda}_{\dot\alpha}\xi_-^I + \tilde{\eta}_{\dot\alpha}\xi_+^I\,,\end{aligned} \tag{2.154}$$

where we scale each of $\eta, \tilde{\eta}$ as $\sim m$. We are looking for pieces that survive in the $m_\gamma, m_H \to 0$ limit. The leading piece in the numerator is that with zero $\eta, \tilde{\eta}$'s which is given as:



$$= \quad \frac{g}{m_\gamma}\langle 12\rangle[21] = \frac{g}{m_\gamma}(m_H^2)\,. \tag{2.155}$$

This indeed vanishes as $m_H \to 0$, as expected since we don't have an (EEH) coupling in the UV. For the order $\tilde{\eta}$ piece, we have



$$= \quad \frac{g}{m_\gamma}\langle 12\rangle[1\tilde{\eta}_2]\,. \tag{2.156}$$

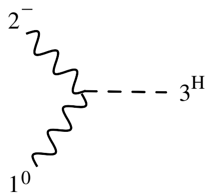
This term is more interesting. To compute it, note that in the UV we have our usual restrictions on 3 particle kinematics — either $\lambda_1 \propto \lambda_2 \propto \lambda_3$ or $\tilde{\lambda}_1 \propto \tilde{\lambda}_2 \propto \tilde{\lambda}_3$. This 3-particle amplitude vanishes in the first case. On the other hand, in the second case, we have by momentum conservation that

$$\tilde{\lambda}_1 = \langle 23\rangle\tilde{\xi}, \quad \tilde{\lambda}_2 = \langle 31\rangle\tilde{\xi}, \quad \tilde{\lambda}_3 = \langle 12\rangle\tilde{\xi}\,, \tag{2.157}$$

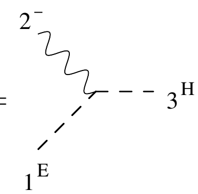
and so

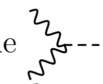
$$[1\tilde{\eta}_2] = [2\tilde{\eta}_2]\frac{\langle 23\rangle}{\langle 31\rangle} = m_\gamma\frac{\langle 23\rangle}{\langle 31\rangle}\,. \tag{2.158}$$

So this amplitude becomes



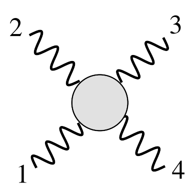
$$= \frac{g}{m_\gamma}\langle 12\rangle \times m_\gamma \frac{\langle 23\rangle}{\langle 31\rangle} = g\frac{\langle 12\rangle\langle 23\rangle}{\langle 31\rangle} \quad = \tag{2.159}$$



exactly as desired. Obviously we get the analogous result for 2^+. Thus quite beautifully the massive three-particle amplitude  reproduces the component helicity amplitudes and unifies them into a single object. Note also that despite appearances, there is no such singularity as $m_\gamma \to 0$.

Here we see an interesting counterpart to the purely massless 3pt amplitudes, which are not manifestly local due to the presence of poles. Healthy theories of massless particle (which we should reproduce in the UV) do not have such non-local poles at 4pts and higher. When we perform this "IR deformation", we have removed the non-local poles but are left with seeming factors of $\frac{1}{m_\gamma}$ in the amplitude. But as we have seen the 3pt amplitude is — by design — chosen to match the correct massless helicity amplitudes and thus be smooth as $m_\gamma \to 0$, and this will be inherited at higher points.

Indeed let us compute the 4-particle amplitude with all massive spin 1 particles consistent with factoring into the three-point amplitude in eq.(2.153). Since we have no "x" factors to worry about, we can proceed in the most naive possible way, simply gluing the 3-pt amplitudes in the s, t and u channels, and we find:



$$\frac{g^2}{m_\gamma^2}\left[\frac{\langle\mathbf{12}\rangle[\mathbf{12}]\langle\mathbf{34}\rangle[\mathbf{34}]}{s-M_h^2}+\frac{\langle\mathbf{23}\rangle[\mathbf{23}]\langle\mathbf{14}\rangle[\mathbf{14}]}{t-M_h^2}+\frac{\langle\mathbf{13}\rangle[\mathbf{13}]\langle\mathbf{24}\rangle[\mathbf{24}]}{u-M_h^2}\right]. \tag{2.160}$$

Since there are no three-point massive spin-1 amplitude, there is no poles involving m_γ. Note that all possible contact terms here can be eliminated since they give growing amplitudes for some of the helicity components in the UV, which we are assuming not to have. Now again, despite appearances this amplitude is guaranteed (by construction!) to be smooth in the high-energy (or $m_\gamma, m_H \to 0$) limit. Let us first show this directly for some of the helicity components. For instance, the all-longitudinal amplitude is

$$\frac{g^2}{m_\gamma^2}\left[\frac{(\hat{p}_1\cdot\hat{p}_2)(\hat{p}_3\cdot\hat{p}_4)}{s-M_h^2}+\frac{(\hat{p}_1\cdot\hat{p}_4)(\hat{p}_2\cdot\hat{p}_3)}{t-M_h^2}+\frac{(\hat{p}_1\cdot\hat{p}_3)(\hat{p}_2\cdot\hat{p}_4)}{u-M_h^2}\right], \tag{2.161}$$

where with $p = \lambda\tilde{\lambda} + \eta\tilde{\eta}$ we define $\hat{p} = \lambda\tilde{\lambda} - \eta\tilde{\eta}$. Just to take a first look at the HE limit, which naively goes as $\frac{g^2}{m\gamma^2}$, we drop the η's and find at $O(\frac{1}{m_\gamma^2})$

$$\frac{g^2}{m_\gamma^2} \times [s + t + u] = 0\,, \tag{2.162}$$

and so as expected there is no $(\frac{s,t,u}{m_\gamma^2})$ singularity as $m_\gamma \to 0$. In order to find the leading high-energy limit, let us define $q \equiv \eta\tilde{\eta}$. Note that $p \cdot q = \frac{m_\gamma^2}{2}$ so $q = O(m_\gamma^2)$, and we will work to first order in q. Using $2p_1 \cdot p_2 = s - 2m_\gamma^2$, and also $\hat{p} = p - 2q$, we find in the HE limit

$$\frac{4(\hat{p}_1 \cdot \hat{p}_2)(\hat{p}_3 \cdot \hat{p}_4)}{s - M_h^2} = s - 4m_\gamma^2 + M_h^2 - 4(q_1 \cdot p_2 + q_2 \cdot p_1 + q_3 \cdot p_4 + q_4 \cdot p_3) + \mathcal{O}(M_h^4, m_\gamma^4)\,. \tag{2.163}$$

So summing over channels gives

$$\begin{aligned} & s + t + u - 3 \times 4m_\gamma^2 + 3M_h^2 - 4(q_1 \cdot (p_2 + p_3 + p_4) + ...) \\ & = 4m_\gamma^2 - 3 \times 4m_\gamma^2 + 3M_h^2 + 2 \times 4m_\gamma^2 \\ & = 3M_h^2. \end{aligned} \tag{2.164}$$

Hence the all-longitudinal amplitude is fixed to be $\frac{3}{4}g^2\frac{M_h^2}{m_\gamma^2}$. This tells us we must have a quartic coupling in the UV, and by the $U(1)$ invariance it must be $\lambda(E^2 + H^2)^2$ with

$$\frac{\lambda}{M_H^2} = \frac{g^2}{m_\gamma^2}\,. \tag{2.165}$$

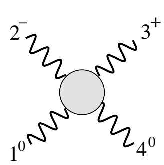
Let's see how some of the other component amplitudes work. Consider $(1^0 2^- 3^+ 4^0)$, which should match $(1^E 2^- e^+ 4^E)$ in the high-energy limit. This is

$$\frac{g^2}{m_\gamma^2}\left[\frac{\langle 12\rangle[1\tilde{\eta}_2]\langle 4\eta_3\rangle[43]}{s} + \frac{\langle 2\eta_3\rangle[\tilde{\eta}_2 3]\langle 14\rangle[14]}{t} + \frac{\langle 1\eta_3\rangle[13]\langle 24\rangle[\tilde{\eta}_2 4]}{u}\right]. \tag{2.166}$$

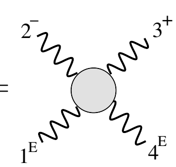
Note that since all that matters is $[2\tilde{\eta}_2] = m_\gamma$, $\langle 3\eta_3\rangle = m_\gamma$, $\tilde{\eta}_2, \eta_3$ are defined up to shifts such as $\tilde{\eta}_2 \to \tilde{\eta}_2 + \alpha\tilde{\lambda}_2$. Not surprisingly the above representation is independent of such shifts. The term in the brackets shifts as

$$\alpha\left[\langle 4\eta_3\rangle[43] + \langle 2\eta_3\rangle[23] + \langle 1\eta_3\rangle[13]\right] = \alpha\langle\eta_3|p_3|3] = 0\,. \tag{2.167}$$

Hence we are free to choose $\tilde{\eta}_2 = \frac{m_\gamma\tilde{\lambda}_3}{[23]}$, $\eta_3 = \frac{m_\gamma\lambda_2}{\langle 23\rangle}$, [13] so then only the $s + u$ channel terms contribute and we find



$$= g^2\langle 2|p_4|3]^2\left(\frac{1}{st} + \frac{1}{tu}\right) = -g^2\frac{\langle 2|(p_1 - p_4)|3]^2}{4su} = \tag{2.168}$$



[13] Using this representation for $\tilde{\eta}_2$, one can also show that the $\mathcal{O}(m_\gamma^{-1})$ term in the amplitude vanishes as well, with $\tilde{\lambda}_2^I \to \tilde{\eta}_2$, while all other massive spinors are set to their massless limit.

Thus we find the correct amplitude for minimally charged scalars in the UV. All other helicity amplitude components vanish as $m_\gamma \to 0$. We have thus verified that the 4pt massive amplitudes are an "infrared deformation" of the massive ones, reproducing and unifying the different helicities in the HE limit.

Higgsing as UV Unification → IR Deformation

Given that we see the massive amplitudes reproduce the massless ones at high energy, we are motivated to consider directly assembling the high-energy massless amplitudes in a way that one can readily "IR deform" the amplitude by simply putting in the mass for the propagator and "**BOLD**ing" the spinor brackets. We are then guaranteed to have a result that gives the correct high-energy behaviour, and what remains is simply to add in higher order corrections in mass that ensure the massive residue is matched.

Let's first consider all the different component amplitudes - Compton scattering for H, E, and the quartic interaction for E. We will first merely group these amplitudes together, ready to be "**BOLD**"ed + unified into a massive amplitude. The massive amplitude in the IR will be the four massive vector amplitude, and thus we will need a total of eight spinors to carry the SU(2) Little group indices, these are the objects that will be **BOLD**ed. Thus the name of the game is to write the massless amplitudes in a form which contains eight spinors, two for each legs, and every thing else can only be expressed as momenta. Note that because of this the E^4 quartic must be written in an interesting way. Naively it is just 3λ, but to put it in a form where by **BOLD**ing we can recognize it as a component of massive spin 1, we have to write it in the following way:

$$3\lambda = \lambda\frac{s^3+t^3+u^3}{stu} = \lambda\frac{(\langle 12\rangle[21][3|p_4|3\rangle[4|p_3|4\rangle+\langle 34\rangle[43][1|p_2|1\rangle[2|p_1|2\rangle) + \{u\}+\{t\}}{2stu}\,, \tag{2.169}$$

where $\lambda = \frac{g^2M_h^2}{m_\gamma^2}$ and $\{t\}$ $\{u\}$ represents its t, u image. This is the only way to represent the "constant" without introducing double poles. Similarly for the two photons two E amplitudes we write

$$-g^2\frac{\langle 2|p_1-p_4|3]^2}{4su} = g^2\frac{[14]\langle 14\rangle\langle 2|p_1-p_4|3]^2}{4stu}. \tag{2.170}$$

Collecting all the component amplitudes together, we are ready to IR deform: declaring the particles have mass m_γ by **BOLD**ing the spinors, and deforming $s \to s - M_h^2$ etc., giving an IR deformed object:

$$\frac{[\mathbf{12}]\langle\mathbf{12}\rangle(g^2\langle\mathbf{3}|p_1-p_2|\mathbf{4}]^2 + g^2\langle\mathbf{4}|p_1-p_2|\mathbf{3}]^2 + 2\lambda[\mathbf{3}|p_4|\mathbf{3}\rangle[\mathbf{4}|p_3|\mathbf{4}\rangle+(1,2\leftrightarrow 3,4)}{4(s-M_h^2)(t-M_h^2)(u-M_h^2)} + \{u\}+\{t\}\,. \tag{2.171}$$

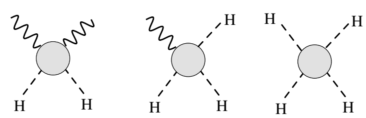
The above result by construction gives the correct answer in the High-energy limit, with mismatch at higher order in m_γ^2, M_h^2. Thus we have the identity

$$\begin{aligned}&\frac{g^2}{m_\gamma^2}\left(\frac{\langle\mathbf{12}\rangle[\mathbf{12}]\langle\mathbf{34}\rangle[\mathbf{34}]}{s-M_h^2}+\frac{\langle\mathbf{14}\rangle[\mathbf{14}]\langle\mathbf{32}\rangle[\mathbf{32}]}{t-M_h^2}+\frac{\langle\mathbf{13}\rangle[\mathbf{13}]\langle\mathbf{24}\rangle[\mathbf{24}]}{u-M_h^2}\right)\\ &=\frac{[\mathbf{12}]\langle\mathbf{12}\rangle(g^2\langle\mathbf{3}|p_1-p_2|\mathbf{4}]^2+g^2\langle\mathbf{4}|p_1-p_2|\mathbf{3}]^2+2\lambda[\mathbf{3}|p_4|\mathbf{3}\rangle[\mathbf{4}|p_3|\mathbf{4}\rangle+(1,2\leftrightarrow 3,4)}{4(s-M_h^2)(t-M_h^2)(u-M_h^2)}\\ &+\{u\}+\{t\}+\mathcal{O}(m_\gamma^2,M_h^2)\,.\end{aligned} \tag{2.172}$$

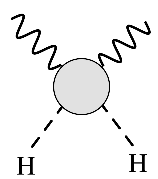
But now in this form, the challenge is to check the factorization channels, which will fix the $\mathcal{O}(m_\gamma^2, M_h^2)$ terms. For example in the limit where $m_\gamma^2 = M_h^2 \equiv m^2$, the remaining term is simply

$$\mathcal{O}(m^2)=\frac{m^2(\langle\mathbf{43}\rangle^2[\mathbf{12}]^2+\langle\mathbf{12}\rangle^2[\mathbf{34}]^2-\langle\mathbf{43}\rangle[\mathbf{43}]\langle\mathbf{12}\rangle[\mathbf{12}])+\{u\}+\{t\}}{(s-m^2)(t-m^2)(u-m^2)}\,. \tag{2.173}$$

We have thus seen the Higgs mechanism very explicitly as an IR deformation. Note that while it is pleasing to see everything work explicitly, the correct HE limit was guaranteed once we ensured the 3 particle amplitudes reproduced and unified the helicity amplitudes in the high-energy limit. Again, all the non-trivial physics was in the "unified packaging" of all the massless helicity amplitudes into the massive amplitudes—everything was guaranteed to work after that point.

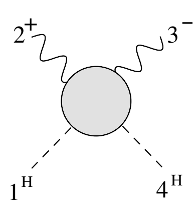
We could also consider 

and derive the rest of the physics. For example, from the fact that we know there is a coupling $\lambda(E^2+H^2)^2$ in the UV, tells us that we have an $(EEHH)=\lambda$ component that needs to be unified into



.

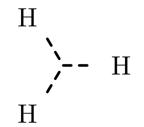
Naively, one would combine this with $(\gamma\gamma HH)$, however, the bolded version of this amplitude:



$$=\frac{\langle 3|p_1-p_4|2]^2}{4st}\ \rightarrow\ \frac{\langle\mathbf{3}|p_1-p_4|\mathbf{2}]^2}{4(s-m_\gamma^2)(t-m_\gamma^2)}\,, \tag{2.174}$$

will not contain such a high-energy scalar contact piece. This suggests that we should directly IR deform it:

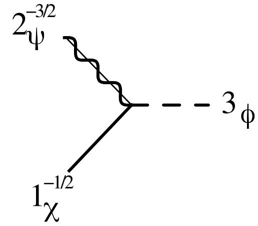
$$\lambda=\lambda\frac{\langle 23\rangle[23]}{t}\ \rightarrow\ \lambda\frac{\langle\mathbf{23}\rangle[\mathbf{23}]}{(t-m_H^2)}\,. \tag{2.175}$$

Thus we see that by IR deforming it, we are forced to have a Higgs propagator, whose residue reveals the presence of a Higgs cubic coupling 

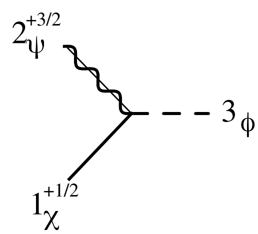
.

Super-Higgs

Let us now describe the Super-Higgs mechanism. Again, we will consider the simplest case, and $\mathcal{N} = 1$ SUGRA where we have a graviton, gravitino ψ as well as a chiral superfield: a fermion χ and a scalar ϕ. First, in the massless limit, in addition to the universal couplings to gravity we have

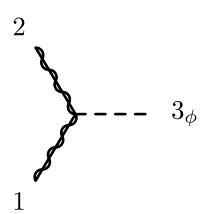


$$\frac{1}{M_{pl}}\frac{\langle 12\rangle^2\langle 23\rangle}{\langle 13\rangle}$$



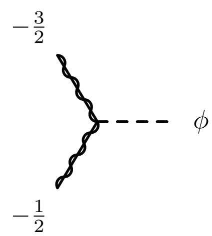
$$\frac{1}{M_{pl}}\frac{[12]^2[23]}{[13]} \tag{2.176}$$

Now, we wish to see whether the (ψ, χ) amplitudes can be unified into those of a single massive spin $\frac{3}{2}$ multiplet. The logic completely parallels to the Abelian Higgs mechanism we discussed above. Indeed, again we simply have the following massive amplitude for massive spin-$\frac{3}{2}$, spin-$\frac{3}{2}$ and scalar:



$$\frac{1}{M_{pl}}\frac{1}{m_{3/2}}\langle \mathbf{12}\rangle[\mathbf{12}]\left([\mathbf{12}] + \langle \mathbf{12}\rangle\right).$$

The correct HE limit emerges in exactly the same way. For instance the $(1^{-\frac{1}{2}}2^{-\frac{3}{2}}3^0)$

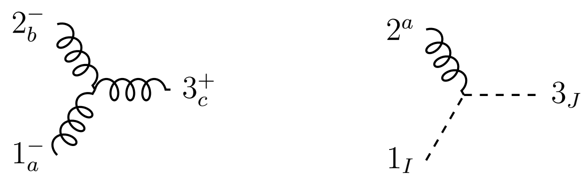


$$\frac{1}{M_{pl}}\frac{1}{m_{3/2}}\langle 21\rangle[1\tilde{\eta}_2]\left([\tilde{\eta}_1\tilde{\eta}_2] + \langle 12\rangle\right).$$

The first term vanishes as $m_{3/2} \to 0$, while the second term becomes $\frac{\langle 12\rangle^2[1\tilde{\eta}_2]}{M_{pl}m_{3/2}}$. Substituting $[1\tilde{\eta}_2] = m_{3/2}\frac{\langle 23\rangle}{\langle 13\rangle}$ yields the correct massless amplitude in the HE limit. After this point everything is guaranteed to work just as with the Abelian Higgs mechanism, and we omit the details. (We have described spontaneous SUSY breaking with the chiral superfield $X = \phi + \theta\chi + \theta^2 F_\phi$ and $W = \mu^2 X$.)

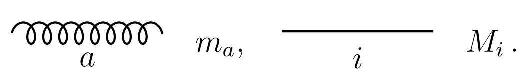
Non-Abelian Higgs

Let us now look at the most general case. In the UV we have gluons and scalars in some representation R:



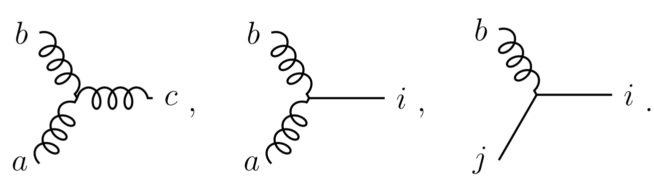
$$g f^{abc}\frac{\langle 12\rangle^3}{\langle 13\rangle\langle 23\rangle} \qquad\qquad g(T^a)_{IJ}\frac{\langle 12\rangle\langle 32\rangle}{\langle 13\rangle}\,.$$

Now, we want to take the $\pm$ component of index a, together with some linear combination of the scalars ($u^a_J\phi^J$), and make the part of a massive vector of mass m_a. Here, we are assuming that all the vectors are massive, in particular this means that the number of scalars N_ϕ is larger than or equal to the number of massless vectors. Then, what we are doing is considering a big $SO(N_\phi)$ matrix U_{IJ}, such that $U_{aJ}\phi_J$ will become the longitudinal component of the massive vector. The remaining scalars are "Higgses" $U_{iJ}\phi_J$. We can always diagonalise so these have mass M_i^2, i.e. $U_{aI}U_{bI} = \delta_{ab}, U_{aI}U_{iI} = 0, U_{iI}U_{jI} = \delta_{ij}$. So, we have



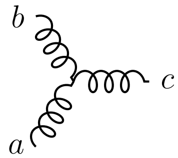
(2.177)

The relevant massive amplitudes in question includes

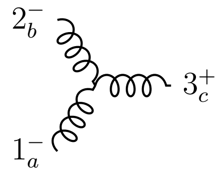


(2.178)

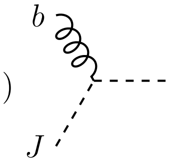
In particular in the high energy limit we must have, for example:

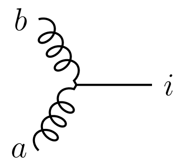


$$\xrightarrow{H.E.} \quad g f^{abc}$$

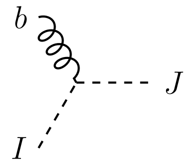


$$+ g U_I^a U_J^c (T_{IJ}^b) \tag{2.179}$$

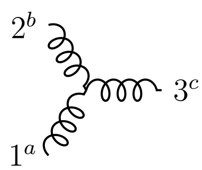




$$\xrightarrow{H.E.} \quad g U_I^a U_{Ji} (T_{IJ}^b) \tag{2.180}$$

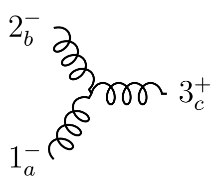


Being able to unify these into massive amplitudes will allow us some interesting interpretations of the U matrix. First, the only possibility for the first figure in (2.178) is[14]

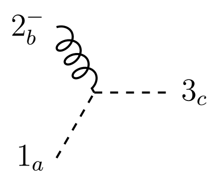


$$= \frac{g f^{abc}}{m_a m_b m_c} \left(\langle \mathbf{12} \rangle [\mathbf{12}] \langle \mathbf{3} | p_1 - p_2 | \mathbf{3}] + \text{cyc.} \right). \tag{2.181}$$

We can again compute the HE limit of the component amplitudes. The details of this limit is given in appendix A.5, and we simply summarise the result:



$$\longrightarrow g f^{abc} \frac{\langle 12 \rangle^3}{\langle 23 \rangle \langle 31 \rangle} \tag{2.182}$$



$$\longrightarrow \frac{g f^{abc}}{m_a m_c} \frac{\langle 12 \rangle \langle 23 \rangle}{\langle 31 \rangle} \left(m_b^2 - m_c^2 - m_a^2 \right). \tag{2.183}$$

From the above we see that in order for the massless amplitudes to be unified into a single massive amplitude, the matrix U_I^a must satisfy

$$U_I^a T_{IJ}^b U_J^c = f^{abc} \frac{m_b^2 - m_a^2 - m_c^2}{m_a m_c}. \tag{2.184}$$

[14] This can be verified by noting that $\epsilon_{\alpha\dot{\alpha}} = \frac{\lambda_\alpha^{\{I} \tilde{\lambda}_{\dot{\alpha}}^{J\}}}{m}$, and substitute into the usual Feynman rules.

Let's define $\tau_I^a = m_a U_I^a$, and then

$$\tau_I^a \tau_I^b = m_a^2 \delta^{ab}. \tag{2.185}$$

So, we can re-write the eq.(2.184) as

$$(\tau^a T^b \tau^c) = f^{abd}(\tau^b \tau^d - \tau^a \tau^d - \tau^c \tau^d) \tag{2.186}$$

where we have suppressed the contraction of indices I, J. The solution to the constraint for τ_I^a is simply that

$$\tau_I^a = T_{IJ}^a V_J \tag{2.187}$$

for some constant vector V_J (the "vev"). Indeed this is precisely what we get in the usual Higgs mechanism. The combination $T_{IJ}^a V_J \phi_I$ is "eaten", and diagonalising $(M^2)^{ab} = V^T T^a T^b V$.

One can check that after substituting for τ, eq.(2.186) becomes

$$V^T T^a T^b T^c V = -V^T T^c T^b T^a V = \frac{1}{2} V^T (T^a T^b T^c - T^c T^b T^a) V \tag{2.188}$$

(note we are always writing with real states so $T_{IJ}^a = -T_{JI}^a$). Now, if we assume that the "coupling tensor" f^{abc} is the structure constant for the Lie group associated with T^a, then we can repeatedly use $T^a T^b = f^{abd} T^d + T^b T^a$, and we find,

$$\begin{aligned} T^a T^b T^c &= f^{bcd} T^a T^d + T^a T^c T^b \\ &= f^{bcd} T^a T^d + f^{acd} T^d T^b + T^c T^a T^b \\ &= f^{bcd} T^a T^d + f^{acd} T^d T^b + f^{abd} T^c T^d + T^c T^b T^a. \end{aligned} \tag{2.189}$$

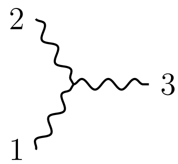
Using the fact that $V^T T^a T^b V$ is diagonalised, we find:

$$\begin{aligned} & V^T (T^a T^b T^c - T^c T^b T^a) V \\ &= f^{bca} m_a^2 + f^{acb} m_b^2 + f^{abc} m_c^2 \\ &= f^{abc} (m_a^2 + m_c^2 - m_b^2). \end{aligned} \tag{2.190}$$

Once eq.(2.184) is satisfied, the rest of the story is again the same as our previous examples. Note in particular that we *must* have Higgses! Even if we have $N_{\text{scalar}} = N_{\text{gluon}}$ precisely, the interactions are *not* the correct ones for the full UV theory due to the standard polynomial growth of the longitudinal piece scattering, which is not present for the UV theory. But with the "uneaten Higgses" included, is simply chosen to match the high energy limit, and we manifestly match to a healthy UV theory.

Obstruction for Spin 2

We now consider massive spin-2 particles, which in the HE limit should yield a graviton, a massless vector and scalar. We would like to see if the massless interactions can be consistently unified into an IR massive amplitude. The three-point massive spin-2 amplitude can be easily written down as:



$$= \frac{1}{M_{pl}m^6}\left[\langle\mathbf{12}\rangle[\mathbf{12}]\langle\mathbf{3}|p_1 - p_2|\mathbf{3}] + \text{cyc.}\right]^2 , \tag{2.191}$$

where m is the mass of the massive graviton. Let us look at the HE limit. We can directly import what was done for non-abelian Higgs, and one finds:

$$\frac{1}{M_{pl}m^6}\left[\langle\mathbf{12}\rangle[\mathbf{12}]\langle\mathbf{3}|p_1 - p_2|\mathbf{3}] + \text{cyc.}\right]^2 \quad \xrightarrow{\ HE\ } \quad \begin{cases} (-2,-2,+2): & \frac{1}{M_{pl}}\frac{\langle 12\rangle^6}{\langle 13\rangle^2\langle 23\rangle^2} \\ (0,-2,0): & \frac{3}{M_{pl}}\frac{\langle 12\rangle^2\langle 23\rangle^2}{\langle 13\rangle^2} \end{cases} . \tag{2.192}$$

Notice the extra factor of 3 associated with the minimally coupled scalars. This extra factor is due to the three different combinations $(+,-,-)\times(-,-,+)$, $(-,-,+)\times(+,-,-)$, and $(0,-,0)\times(0,-,0)$. Thus the scalar coupling at high energy is three times what it should be. This is unacceptable since gravitational coupling is universal, and the coupling strength M_{pl} has already been set by the self-interaction. Note that similar difficulties arise for the HE limit that yields the one-graviton two-minimally-coupled-vector, where one obtains $-2\langle 12\rangle^4/M_{pl}\langle 13\rangle^2$. Again the factor of 2 is inconsistent with graviton self coupling. Thus we see that there is a fundamental obstruction in organising the massless degrees of freedom into a massive spin-2 particle, in a way such that the massive interactions have HE limit that morphs into a consistent UV theory.

2.7 Loop Amplitudes

In this section we briefly touch on constructing loop amplitudes by an on-shell gluing of the tree amplitudes we have found in previous sections. We will follow the philosophy of "generalized unitarity" [30–32, 34, 35, 39, 78], where the integrand for loop amplitudes is determined by a knowledge of its (generalized) cuts, putting internal propagators on-shell. As is well-known, at one-loop this gives a systematic way of determining the integrand from gluing together on-shell tree amplitudes.[15] While we are not adding anything new to this conceptual framework, the technical advantages offered by our formalism for massive

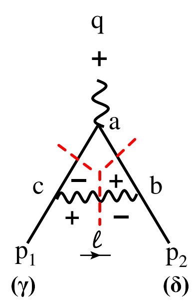
[15]There is an obvious subtlety in this on-shell approach to loop amplitudes, regarding "wavefunction renormalization". In the unitarity approach where one glues tree amplitude on both sides of the cut, there will be diagrams which correspond to a bubble insertion on the external leg, and hence give rise

particles with spin are significant in many cases, including the dispensation of complicated gamma matrix algebra, the clear separation of electric and magnetic moments for charged particles, the extraction of UV divergent properties without the contamination from IR divergences (by virtue of using massive external and internal states), and finally directly obtaining the (internal) mass depending pieces in the small mass expansion relevant for obtaining rational terms for massless one-loop amplitudes. In all of these processes, as they do not have tree counterparts, bubbles on external legs do not contribute. It is pleasing to continue seeing directly the way in which Poincare symmetry and Unitarity fully determines the physics, not just at tree-level but also in incorporating the leading quantum loop corrections as well.

$g{-}2$ for spin-$\frac{1}{2}$ and 1

As seen in previous discussions the simplicity of minimal coupling allows us to straightforwardly separate the magnetic moment pieces. The same simplicity translates to a straightforward computation for the loop level magnetic moment.

Let's consider the $e^+, e^- \to \gamma$ at one loop. The diagram we want to build is:



$$\sim e^3 m^3 x_a \varepsilon_{\alpha\beta}\left[\varepsilon^{\beta\gamma}\frac{x_b}{x_c}\left(\varepsilon + x_c\frac{\lambda_\ell\lambda_\ell}{m}\right)^{\alpha\delta} + \varepsilon^{\alpha\delta}\frac{x_c}{x_b}\left(\varepsilon - x_b\frac{\lambda_\ell\lambda_\ell}{m}\right)^{\beta\gamma}\right]. \quad (2.193)$$

where we've glued the three-point vertices according to the two possible helicity configurations in the internal photon lines. Notice that here, we are using the three point amplitude in the SL(2,C) undotted basis. This is motivated by eq.(2.74), which yields a clear separation of $(g{-}2)$ factors in this basis. One can also understand this from the fact that anomalous moments should arise only if the particle carries spin. By expanding the integrand in eq.(2.193), one notices that the λ independent terms will be present for charged scalars as well, and thus the piece of the integrand that can contain the magnetic

to an 1/0 from the on-shell propagator. In the Feynman diagram approach, these are wave function diagrams that are to be amputated, replaced by counter terms. This procedure breaks gauge invariance in the intermediate steps. For massless internal states, these can be side stepped since there will be UV-IR cancellation for these diagrams. For massive internal particles this is no-longer the case, and we refer the reader to [19, 41, 68] for unitarity based treatments of this issue. This subtlety will not affect any of the examples we discuss in this section: for (g-2) and rational terms, the 1-loop corrections are leading, while for the beta function the external massive particles are merely probes.

moment is:

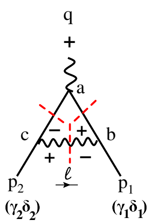
$$e^2m^2x_a(x_b-x_c)\lambda_\ell^\delta\lambda_\ell^\gamma = -mx_a q^{\delta}{}_{\dot\alpha}\ell^{\dot\alpha\beta}\,. \tag{2.194}$$

This gives us the following integrand:

$$-\,mx_a\int\frac{d^4\ell}{(2\pi)^4}\frac{q^{\delta}{}_{\dot\alpha}\ell^{\dot\alpha\beta}}{\ell^2((\ell-p_2)^2-m^2)((\ell+p_1)^2-m^2)} = \frac{e^2}{(4\pi)^2}2x_a\frac{q^{\delta}{}_{\dot\alpha}p_1^{\dot\alpha\beta}}{m} = \frac{\alpha}{2\pi}x_a^2\lambda_q^\gamma\lambda_q^\delta\,. \tag{2.195}$$

This gives the $(g{-}2)=\frac{\alpha}{2\pi}$ by comparing with eq.(2.74).

Just to give us a little bit more challenge, let's now consider the $W^+,W^-\to\gamma$ at one loop involving only photon coupling. The integrand is again built from:



$$\sim \quad e^3m^3x_a\varepsilon_{\{\alpha_1\alpha_2}\varepsilon_{\beta_1\}\beta_2}\left[\varepsilon^{\beta_1\{\gamma_1}\varepsilon^{\alpha_1\delta_1\}}\frac{x_b}{x_c}\left(\varepsilon+x_c\frac{\lambda_\ell\lambda_\ell}{m}\right)^{\alpha_2\{\delta_2}\left(\varepsilon+x_c\frac{\lambda_\ell\lambda_\ell}{m}\right)^{\beta_2\gamma_2\}}\right.$$
$$\left.+\varepsilon^{\beta_2\{\gamma_2}\varepsilon^{\alpha_2\delta_2\}}\frac{x_c}{x_b}\left(\varepsilon-x_b\frac{\lambda_\ell\lambda_\ell}{m}\right)^{\alpha_1\{\delta_1}\left(\varepsilon-x_b\frac{\lambda_\ell\lambda_\ell}{m}\right)^{\beta_1\gamma_1\}}\right]\,. \tag{2.196}$$

Leaving behind the electric coupling, we now have two structures for the numerator of the integrand:

$$\begin{aligned}
& e^2x_a(x_b-x_c)m^2\left[4\left(\varepsilon^{\delta_1\{\delta_2}\lambda_\ell^{\gamma_1}\lambda_\ell^{\gamma_2\}}+\varepsilon^{\gamma_1\{\delta_2}\lambda_\ell^{\delta_1}\lambda_\ell^{\gamma_2\}}\right)\right]+16e^2x_ax_bx_cm\lambda_\ell^{\delta_1}\lambda_\ell^{\delta_2}\lambda_\ell^{\gamma_1}\lambda_\ell^{\gamma_2}\\
= & \underbrace{-4e^2x_am\left[\varepsilon^{\delta_1\{\delta_2}q^{\gamma_1}{}_{\dot\alpha}\ell^{\dot\alpha\gamma_2\}}+\varepsilon^{\gamma_1\{\delta_2}q^{\delta_1}{}_{\dot\alpha}\ell^{\dot\alpha\gamma_2\}}\right]}_{f_1(q)}\\
+ & \underbrace{\frac{2e^2x_a}{3m}(p_{1\dot\alpha}{}^{\{\delta_1}\ell^{\dot\alpha\gamma_1\}})(p_{2\dot\alpha}{}^{\{\delta_2}\ell^{\dot\alpha\gamma_2\}})}_{f_2(q)}.
\end{aligned} \tag{2.197}$$

Here $f_1(q)$ is the same as the electron moment, and leads to:

$$\begin{aligned}
F_1(q) \quad &= \int\tfrac{d^4\ell}{(2\pi)^4}\tfrac{f_1(q)}{\ell^2((\ell-p_2)^2-m^2)((\ell+p_1)^2-m^2)}\\
&= 4\tfrac{\alpha}{2\pi}x^a\left(\varepsilon^{\delta_1\{\delta_2}\lambda_q^{\gamma_1}\lambda_q^{\gamma_2\}}+\varepsilon^{\gamma_1\{\delta_2}\lambda_q^{\delta_1}\lambda_q^{\gamma_2\}}\right)\,.
\end{aligned} \tag{2.198}$$

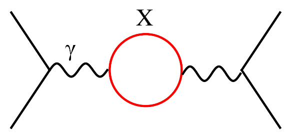
For the second tensor structure, one has:

$$F_2(q)=\int\frac{d^4\ell}{(2\pi)^4}\frac{f_2(q)}{\ell^2((\ell-p_2)^2-m^2)((\ell+p_1)^2-m^2)}=\frac{\alpha}{(4\pi)9m^3}\mathcal{O}_{1,2}^{\{\delta_1\gamma_1\}}\mathcal{O}_{1,2}^{\{\delta_2\gamma_2\}}\,, \tag{2.199}$$

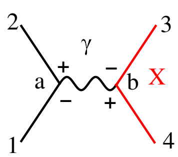
where we've defined $\mathcal{O}_{i,j}^{\alpha\beta}\equiv p_{i\dot\alpha}{}^{\alpha}p_j^{\dot\alpha\beta}$.

The beta function

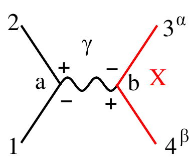
Let's now turn to the extraction of beta function. For massless amplitudes, these can be obtained by extracting the coefficient for the bubble integrals in the scalar integral basis [7, 78]. However, extra care needs to be taken for the subtraction of infrared divergence. Here we will instead consider two massive scalar probes of a photon propagator, and consider the correction to the propagator due to an internal massive scalar, fermion and vector (denoted by X):



The UV divergence of this amplitude contains *the* contribution of a scalar to the beta function, without the IR-contamination. The loop amplitude will be constructed by gluing the 2→2 amplitude involving the scalar probe particle exchanging a photon with X. This will allow us to obtain the beta function for different spins. From the massive vector, we will also be able to extract the contribution for a massless vector by simply subtracting a scalar. Assuming that the mass of X is identical with that of the scalar probe, the relevant tree amplitudes can be easily constructed by generalizing the examples in subsection 2.5:

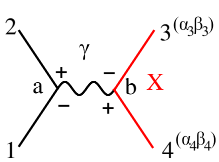


$$: \quad X \in scalar \quad \frac{m^2}{s}\left(\frac{x_a}{x_b}+\frac{x_b}{x_a}\right) = \frac{(p_1-p_2)\cdot p_3}{s}$$



$$: \quad X \in fermion \quad \frac{m}{s}\left(\frac{x_a}{x_b}[\mathbf{34}]+\frac{x_b}{x_a}\langle\mathbf{34}\rangle\right)$$

$$= \frac{1}{2ms}\left(2(p_1-p_2)\cdot p_3\langle\mathbf{34}\rangle - \langle\mathbf{3}|p_1P-Pp_1|\mathbf{4}\rangle\right)$$

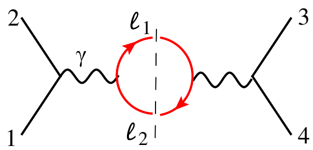


$$: \quad X \in vector \quad \frac{1}{s}\left(\frac{x_a}{x_b}[\mathbf{34}]^2+\frac{x_b}{x_a}\langle\mathbf{34}\rangle^2\right)$$

$$= \frac{1}{m^2 s}\Big((p_1-p_2)\cdot p_3\langle\mathbf{34}\rangle^2 - \langle\mathbf{34}\rangle\langle\mathbf{3}|p_1P-Pp_1|\mathbf{4}\rangle - \frac{\langle\mathbf{3}|p_1P-Pp_1|\mathbf{4}\rangle\langle\mathbf{3}|P|\mathbf{4}]}{2m}\Big) \tag{2.200}$$

where we've again summed over the two possible photon helicity configuration and $P = p_3 + p_4$. The second equality for each amplitude gives the manifest local form, which can be checked against the H.E. limit where one should find a finite result as $m \to 0$. Note that each term contains a piece which is identical to the scalar contribution.

We can now glue the tree amplitudes into the one-loop integrand. The beta function can be readily read off by picking out the divergent piece which is proportional to the tree amplitude. For further simplification, we can take the $s \to 0$ limit, and we will be looking for the term that is proportional to $\frac{2(p_1 \cdot p_3)}{s}$. Let us use the scalar correction as an example. The one-loop amplitude is now



$$= A_4^{scalar}(p_1, \ell_1) A_4^{scalar}(\ell_2, p_3)\big|_{s\to 0} = \frac{4(p_1 \cdot \ell_1)(p_3 \cdot \ell_2)}{s^2}\,. \qquad (2.201)$$

The one-loop integrand is then simply:

$$\frac{4}{s^2}\int \frac{d^{4-2\epsilon}\ell}{(2\pi)^4}\frac{(p_1 \cdot \ell_1)(p_3 \cdot \ell_2)}{(\ell^2 - m^2)((\ell - P)^2 - m^2)} = -\frac{1}{(4\pi)^2\epsilon}\frac{1}{6}\frac{(2p_1 \cdot p_3)}{s} + \cdots\,. \qquad (2.202)$$

where $\cdots$ represent terms terms that are purely functions of s, or finite. For fermions, there are now two pieces that are relevant: the square of the scalar piece, and the square of the $p_i P$ piece. All other contributions cannot generate the $p_1 \cdot p_3$ tensor structure. We find:

$$A_4^{fermion}(p_1, \ell_1) A_4^{fermion}(\ell_2, p_3) = \frac{8(p_1 \cdot \ell_1)(p_3 \cdot \ell_2)}{s^2} - 2\frac{(p_1 \cdot p_3)}{s} + \cdots\,. \qquad (2.203)$$

The relevant part of the one-loop integrand is then:

$$\frac{1}{s}\int \frac{d^{4-2\epsilon}\ell}{(2\pi)^4}\frac{8(p_1 \cdot \ell_1)(p_3 \cdot \ell_2)/s - 2(p_1 \cdot p_3)}{(\ell^2 - m^2)((\ell - P)^2 - m^2)} = -\frac{1}{(4\pi)^2\epsilon}\frac{4}{3}\frac{(2p_1 \cdot p_3)}{s} + \cdots\,. \qquad (2.204)$$

Finally, similar analysis for vectors yields:

$$A_4^{vector}(p_1, \ell_1) A_4^{vector}(\ell_2, p_3) = \frac{12(p_1 \cdot \ell_1)(p_3 \cdot \ell_2)}{s^2} + 8\frac{(p_1 \cdot p_3)}{s} \qquad (2.205)$$

which leads to

$$\frac{1}{s}\int \frac{d^{4-2\epsilon}\ell}{(2\pi)^4}\frac{12(p_1 \cdot \ell_1)(p_3 \cdot \ell_2)/s + 8(p_1 \cdot p_3)}{(\ell^2 - m^2)((\ell - P)^2 - m^2)} = \frac{1}{(4\pi)^2\epsilon}\frac{7}{2}\frac{(2p_1 \cdot p_3)}{s} + \cdots\,. \qquad (2.206)$$

Thus we've found that the beta function for a scalar is $\frac{1}{6}$ a Dirac fermion $\frac{4}{3}$ and a massless vector being $-\frac{7}{2} + \frac{1}{6} = -\frac{11}{3}$, where we've subtracted the scalar "eaten" by the massive vector.

Rational terms

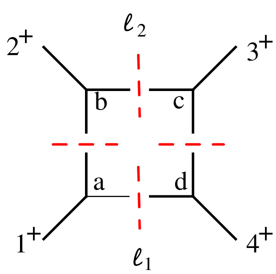
Another application of massive amplitudes is to derive rational terms for massless amplitudes that are not constructible via four-dimensional cuts. These terms appear due to the fact that the integrals are regulated and one can encounter $\epsilon/\epsilon \sim \mathcal{O}(1)$ effects. These terms can be obtained by considering the states in the internal loops to be massive [16, 140], where the mass m^2 is identified with the extra -2ϵ dimension piece of ℓ^2, denoted as μ^2.[16] For QCD, one considers the contribution of a massive adjoint scalar state that is minimally coupled to the external gluons. These "μ" terms are computed using the tree-level amplitudes in D-dimensions [29, 33] and consider the extra dimension momenta as four-dimensional mass.

Here we will directly use the four-dimensional massive amplitudes to obtain the integral coefficients for $I_4[\mu^{2k}]$, the four-point scalar box integral with μ^{2k} as its numerator. For the box-integral coefficient one considers the quadruple cut, where the two solutions for the cut loop momentum are:

$$\ell_1 = \frac{1}{2}\left(c^{\pm}\tilde{\lambda}_1\lambda_4 - \frac{m^2}{tc^{\pm}}\lambda_1\tilde{\lambda}_4\right), \quad c^{\pm} = \frac{\langle 12\rangle}{2\langle 42\rangle}\left(1 \pm \sqrt{1+\frac{4m^2u}{st}}\right). \tag{2.207}$$

The box-coefficient is then obtained by gluing the four tree-amplitudes substituted with the cut loop momenta.

First consider the four-point all-plus amplitude, where the cut is given by:

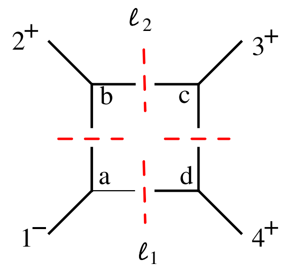


$$\begin{aligned} \sim m^4 x_a x_b x_c x_d \quad &= \quad m^4 \frac{[41]m}{\langle 4|\ell_1|1]}\frac{\langle 4|\ell_1|1]}{\langle 41\rangle m}\frac{[23]m}{\langle 2|\ell_2|3]}\frac{\langle 2|\ell_2|3]}{\langle 23\rangle m} \\ &= \quad m^4\frac{[41][23]}{\langle 41\rangle\langle 23\rangle} \end{aligned} \tag{2.208}$$

This directly gives the all plus integrand, $\frac{[12][34]}{\langle 12\rangle\langle 34\rangle}I_4[\mu^4]$. For the single minus amplitude,

[16] See [76] for some recent applications.

one instead has:



$$\sim m^4 \frac{x_b x_c x_d}{x_a} = m^4 \frac{\langle 4|\ell_1|1]^2}{\langle 41\rangle[14]m^2} \frac{[23]m}{\langle 2|\ell_2|3]} \frac{\langle 2|\ell_2|3]}{\langle 23\rangle m}$$

$$= m^2 \frac{\langle 4|\ell_1|1]^2}{t} \frac{[23]}{\langle 23\rangle} \,. \tag{2.209}$$

Substituting the two solutions for the cut in eq.(2.207) and summing the results, one obtains

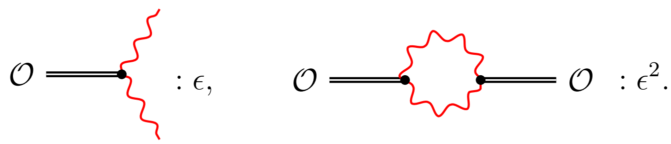
$$\frac{[23][42]\langle 12\rangle}{4\langle 23\rangle\langle 42\rangle[12]} \left(\frac{st}{2u} I_4[\mu^2] + I_4[\mu^4] \right) \,. \tag{2.210}$$

The above rational terms are in agreement with [29].

2.8 Form Factors and Correlation Functions

The ability to discuss scattering amplitudes for general mass and spin largely removes the distinction between amplitudes and "off-shell" objects such as correlation functions and form-factors. Consider correlation functions for the stress tensor for some theory. The computations are precisely the same as what we would carry out if we were computing the scattering amplitude for a massive spin two particle, (arbitrarily) weakly coupled to the theory. The scattering amplitude for these massive particles gives us the correlation function in momentum-space, which corresponds closely to the experiments that are actually done to measure correlation functions. Strictly speaking we are coupling a continuum of particles of different masses, and we are getting the correlator in momentum space for the external legs p_a in the timelike Lorentzian region where $p_a^2 > 0$. But we can then define the correlators for null and spacelike momenta by analytic continuation. At least in perturbation theory—which is what we will largely concern ourselves with here in this subsection—there is no ambiguity for what this means in practice.

It is important to imagine that the massive particle $\mathcal{O}$ corresponding to the operator is simply an external probe and does not participate in the dynamics. In other words, we should not have any "internal propagators" associated with cuts that put $\mathcal{O}$ on-shell. In practice, this means that we should be able to make the coupling of $\mathcal{O}$ to our system proportional to a parameter ϵ that we can make as small as we wish. To take an example, consider a 3-point coupling of $\mathcal{O}$ to a pair of massless particles for the system of interest; making this proportional to ϵ means that the leading amplitudes will never involve internal $\mathcal{O}$ particles:



(2.211)

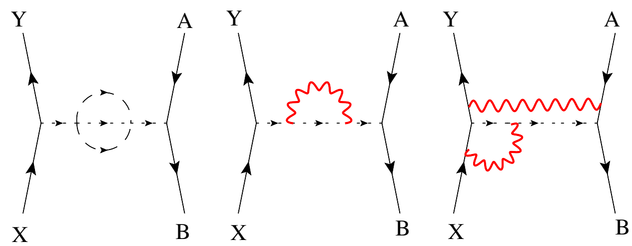
In general, the leading amplitude involving N $\mathcal{O}$'s will be proportional to ϵ^N and will never involve internal $\mathcal{O}$ particles.

Observables in Gauge Theories and Gravity

Before moving on to illustrating how this interpretation is useful in concrete calculations, let us pause to interpret some standard and elementary facts about observables in gauge theories and gravity from this on-shell perspective.

In particular, let us understand the reason for the absence of charged local operators in gauge theory, or any local operators whatsoever in gravity. Consider a charged operator Φ. We know that consistency enforces universal coupling of Φ to photons/gluons, with strength set by the gauge coupling g, and so we *can't* arbitrarily weakly couple Φ to the system. Thus we can't speak of charged local operator. Similarly with gravity, the coupling of any particle to gravity is universal given by $\sqrt{G_N}$, so in the presence of gravity we can't meaningfully talk about any local operators at all. In a conventional Lagrangian description of the physics, this is associated with the impossibility of making local charged operators gauge invariant. Of course we can always fix a gauge and compute correlators for operators in that gauge, but then these are not quite local. If we start with correlators of local operators in the limit as $g^2 \to 0$ or $G_N \to 0$, the weak gauging attaches Wilson lines to the operators in some way. Of course this also has an obvious on-shell meaning, again corresponding closely to physical experiments that measure these Wilson-line dressed correlators.

Consider again a charged scalar Φ of charge $+1$ in an abelian gauge theory, and let's consider the correlator $\langle \Phi^*(x)\Phi(y)\rangle$ first in the limit where we turn off the gauge coupling. We may have $U(1)$ invariant self-interactions for Φ of the form $(\Phi^*\Phi)^2$ for example, and we can also turn on the gauge-interactions. But we also couple Φ to some heavy external probe particles $X^{(q)}$, $Y^{(q+1)}$ and $A^{(Q)}, B^{(Q+1)}$ via the couplings $\epsilon X^{(q)} Y^{(q+1)*}\Phi$, $\epsilon' A^{(-Q)} B^{(-Q-1)*}\Phi^*$. Let's now look at the (XY^*B^*A) scattering amplitude. Since this breaks the global particle number symmetries acting separately on X, Y, A, B as $\epsilon, \epsilon' \to 0$, this amplitude is proportional to the product $\epsilon\epsilon'$; some of the diagrams contributing to the amplitude are shown below:



$$. \qquad (2.212)$$

As $\epsilon, \epsilon' \to 0$, stripping off this product from the amplitude yields the correlator where $\langle \Phi^*(x)\Phi(y) \rangle$ is dressed with Wilson lines in the p_X, p_Y, p_A, p_B directions:

$$M(p_X, p_Y, p_A, p_B) \to \epsilon\epsilon' \int_{x,y} e^{i(p_X+p_Y)x} e^{i(p_A+p_B)y} \langle \left(W^q_{p_X} \Phi W^{*(q+1)}_{p_Y} \right)(x) \left(W^{-Q}_{p_A} \Phi W^{*(-Q-1)}_{p_B} \right)(y) \rangle. \qquad (2.213)$$

The fact that inequivalent "dressings" of the local operator with Wilson lines are possible simply reflects the many different ways we can couple Φ to external probes; since the probes themselves are charged and emit long-range gauge fields, the amplitudes (and hence the extracted correlator) do depend on the choices that are made. Thus, while correlation functions for local charged operators don't exist, dressed version of these correlators exist, for both gauge theory and gravity, to all orders in g and $\sqrt{G_N}$.

There is a deeper difficulty with gravity, which makes even these quasi-local "Wilson-line dressed" correlators ambiguous at a non-perturbatively tiny level, of $O(exp(-M_{Pl}^2/s))$. As we saw in our example above, in order to be able to identify the piece of the amplitude for the heavy probes that is unambiguously associated with the coupling to the operator Φ, it was important that the coupling to the probe broke some global symmetry of the problem. But we expect that gravity breaks all global symmetries, and in particular, we can't say that the XY^*A^*B amplitude, for example, is arbitrarily small; there is some (perhaps virtual black-hole mediated) rate for this process of $O(exp(-M_{Pl}^2/s))$ that pollutes any attempt to associate this amplitude with *the* (Wilson-line dressed) correlator of interest, making it impossible to pick out a piece proportional to $\epsilon\epsilon'$ as $\epsilon, \epsilon' \to 0$.

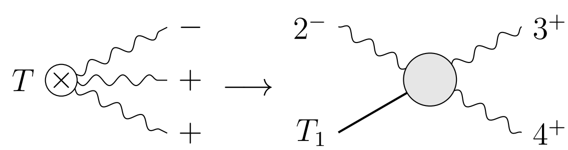
Summarizing more informally, in both gauge theories and gravity we don't have meaningful correlators of local charged operators, for the (relatively trivial) reason that we can't ignore the long-range gauge and gravitational fields. This can already be seen perturbatively in g^2, G_N, but to all order in these couplings, there are dressed versions of local operators that take care of the long-range fields at infinity, smoothly deforming the local correlators we have when $g^2, G_N = 0$. But in gravity, due to exponentially small effects of $O(exp(-Area/G_N))$, associated with black-hole physics, even these dressed versions of local operators don't make precise sense. This is a concrete sense in which any notion of spacetime becomes ambiguous in quantum gravity, for example highlighting that the breakdown of locality in the context of the black-hole information paradox is an effect of $O(exp(-S_{BH}))$, and is otherwise invisible to every order in G_N.

Weinberg-Witten

The interpretation of correlators in terms of massive amplitudes allows us to re-interpret some familiar facts about massive amplitudes we have already encountered to other well-known facts about QFT's. Consider the Weinberg-Witten theorem [172], which in this way of thinking is essentially identical to Yang's theorem. Recall the discussion of consistent couplings of a massive spin S particle to massless particles. Note that since conserved currents and stress tensors measure the charge and the momentum on single particle states respectively, we will be interested in the interaction of the massive state with two opposite helicity massless-particles $h_1 = -h_2$.[17] Our analysis showed that $S + h_2 - h_1$ and $S + h_1 - h_2$ must always be greater or equal to 0, this tells us that for $S = 1$, $|h_1| = |h_2| \leq \frac{1}{2}$, i.e. massless particles with spin $> \frac{1}{2}$ cannot couple to a Lorentz covariant conserved current. Similarly for $S = 2$, $|h_1| = |h_2| \leq 1$, and massless particles with spin > 1 cannot couple to a conserved stress-tensor. This is precisely the Weinberg-Witten theorem.

Form Factors Example: Stress Tensor/Gluons

From Weinberg-Witten theorem we know that the stress tensor can only couple to massless particles of spin ≤ 1, thus we will consider form factors of a stress tensor and three gluons. Identifying the stress tensor as a massive spin-2 state, we will map this to a four-point amplitude involving one massive and three massless states:



. (2.214)

Let us consider the t-channel massless residue. Since the gluon is "charged" under the stress tensor, for the one massive two massless coupling, one should consider opposite helicity gluons. The t-channel residue can then be written as:

$$(\lambda_P)^4 \frac{[p4]^2}{m^3} \frac{[3P]^3}{[P2][23]} = \frac{(\lambda_2)^4 m[23]}{\langle 43\rangle\langle 24\rangle}, \tag{2.215}$$

where again, the equality holds for $\langle 23\rangle = 0$. This leads us to the following simple expression for the form factor:

$$\langle \tilde{T}(1)|2^- 3^+ 4^+\rangle = \frac{(\lambda_2)^4 m}{\langle 43\rangle\langle 32\rangle\langle 24\rangle}. \tag{2.216}$$

[17] Recall that all momenta are out going, so for p_1 and p_2 to represent the same particle, $h_1 = -h_2$.

It is straight forward to check that the above result matches all three factorisation channels, as expected from its cyclic invariant form, up to the over all factor of $(\lambda_2)^4$ that takes care of the excess helicity weight and the stress tensor's SL(2,C) indices. We can straightforwardly extend to two stress tensors coupled to two gluons:

$$\langle \tilde{T}(1)\tilde{T}(2)|3^-4^+\rangle = (\lambda_3)^4\left(\frac{([4|p_1)^2([3|p_2)^2}{t}+\frac{([3|p_1)^2([4|p_2)^2}{u}\right). \tag{2.217}$$

There is an elephant in the room that we have not yet addressed. So far we have been considering conserved operators as massive spinning states. But conserved operators are a tiny subset of an infinite number tensor operators, for which all must have well defined form factors (and in the next section momentum space correlation functions). Furthermore, we should be able to see there must be a kinematic distinction between conserved operators and non-conserved operators, such that higher-spin conserved currents for an interacting theory can be ruled out, à la Coleman–Mandula theorem [53].

As an exercise let's consider a theory with two scalars $(\phi,\bar{\phi})$ and the operators $\mathcal{O}_{1\mu} = \phi\overleftrightarrow{\partial}_\mu\bar{\phi}$ and $\mathcal{O}_{2\mu}=\phi\partial_\mu\bar{\phi}$. The first is a conserved current while the second is not. Let us now consider the three-point form factor for

$$\langle\tilde{\mathcal{O}}_{1\alpha\dot{\alpha},\beta\dot{\beta}}|p_1p_2\rangle\sim(p_1-p_2)^{\alpha\dot{\alpha}},\quad \langle\tilde{\mathcal{O}}_{2\alpha\dot{\alpha},\beta\dot{\beta}}|p_1p_2\rangle\sim p_1^{\alpha\dot{\alpha}}. \tag{2.218}$$

Converting the above result into pure undotted SL(2,C) indices by contracting with (p_1+p_2), one finds:

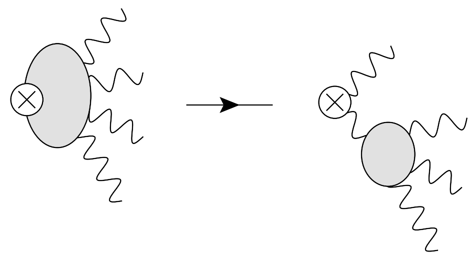
$$\langle\tilde{\mathcal{O}}_1|p_1p_2\rangle\sim[12]\lambda_1^{\{\alpha_1}\lambda_2^{\alpha_2\}},\quad \langle\tilde{\mathcal{O}}_2|p_1p_2\rangle\sim[12]\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}=\frac{1}{2}[12]\left(\lambda_1^{\{\alpha_1}\lambda_2^{\alpha_2\}}+\langle12\rangle\varepsilon^{\alpha_2\alpha_1}\right). \tag{2.219}$$

Not surprisingly the form factor for $\mathcal{O}_2$ can be further decomposed into a combination of $S=2,1,$ and 0 states. Thus we see that a general operator simply corresponds to a linear combination of lower spin states. In position space this is a statement that a general current, for example, can

$$\mathcal{O}^\mu=(\eta^{\mu\nu}-\frac{\partial^\mu\partial^\nu}{\Box})\mathcal{O}_\nu+\frac{\partial^\mu\partial^\nu}{\Box}\mathcal{O}_\nu\equiv\hat{\mathcal{O}}^\mu+\frac{\partial^\mu\partial^\nu}{\Box}\mathcal{O}_\nu. \tag{2.220}$$

where $\hat{\mathcal{O}}^\mu$ is the conserved piece. Note that while there is a conserved piece in a general operator, the projection introduces non-locality and is thus distinct from a genuine conserved operator. This non-locality is present in all the lower spin components in the projection.

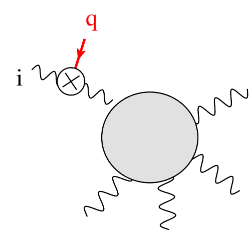
Let us look at this distinction more closely in the context of general form factors. For an interacting theory, the form factor will in general have poles whose residue reveals the existence of a non-trivial S-matrix:



$$+ \ldots . \tag{2.221}$$

Let us consider the particles to be massless, and take the momenta of the operator to be soft. Then just like the usual Weinberg's soft theorems for S-matrix, the form factor will be dominated by diagrams where one has the operator attached to the external leg

$$\sum_i$$



$$= \sum_i \frac{n(p_i, q)}{p_i \cdot q} M_n. \tag{2.222}$$

where q is the soft momenta of the operator, $n(p_i, q)$ is the numerator function. If the operator is a tensor, then $n(p_i, q)$ should carry the corresponding Lorentz indices. Conserved tensor is reflected in that the form factor must vanish when contract with q^μ. If we have a conserved current, then we can have $n(p_i, q)^\mu = e_i p_i^\mu$, where e_i is the charge of each external state. The requirement of conservation then simply corresponds to the requirement of charge conservation. Similarly for conserved stress tensor we have $n(p_i, q)^{\mu\nu} = \kappa p_i^\mu p_i^\nu$, and the conservation condition simply stems from momentum conservation if the coupling κ is universal. Note that for higher spins, $S > 2$, there are no local solutions for $n(p_i, q)^{\mu_1 \cdots \mu_S}$ such that the conserved quantity is respected. This is the Coleman-Mandula theorem! The assumptions that went into this argument is the existent of a non-trivial S-matrix, the analyticity of the form factor which can be interpreted as a massive S-matrix, and Lorentz invariance. The fact that the argument is closely related to Weinberg's soft theorems for gauge bosons is not a surprise in view of our usual intuition that if a conserved tensor exists in an interacting theory, then we can always weakly gauge it and have non-trivial S-matrix involving the gauge boson.

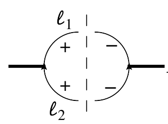
Note that while one can always project out a conserved piece for non-conserved tensors, the corresponding form factor will include non-local pieces. Indeed in this case we can have, for example, $n(p_i, q)^\mu = \frac{q^\mu \tilde{n}(p_i,q)}{q^2} = \frac{q^\mu \tilde{n}(p_i,q)}{m^2}$. This non-locality is again reflected in the singularity of the $m^2 \to 0$ limit. This of course is an artifact of our projection, since there will be lower spin contributions coming along that will contain the same singularity and conspire to cancel, producing a smooth $m^2 \to 0$ limit.

Current and Stress-Tensor Correlators

Let's consider the two and three-point correlation functions for stress-tensors in a conformal theory. In momentum space, the tree-level correlator are computed by gluing tree-level amplitudes with one massive leg and two massless legs. For conformal theories, the available tensor structures are constrained by conformal symmetry. In momentum space, this constraint is simply a reflection of the uniqueness of the three-point amplitude, which is fixed by the spin of the massive state and the helicities of the massless legs.

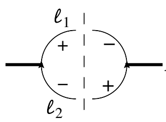
For example, the two point function receives contribution from:

$$\langle T_{\alpha_1\alpha_2\alpha_3\alpha_4} T_{\beta_1\beta_2\beta_3\beta_4}\rangle =$$

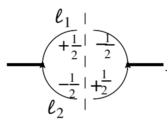


$$I_2\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_1\ell_2}_{\alpha_3\alpha_4} S^{\ell_1\ell_2}_{\beta_1\beta_2} S^{\ell_1\ell_2}_{\beta_3\beta_4}\right]$$

$$+$$



$$I_2\left[\prod_{i=1}^{4} S^{\ell_1\ell_2}_{\beta_i\alpha_i}\right] +$$



$$I_2\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_1\ell_2}_{\beta_1\beta_2} S^{\ell_1\ell_2}_{\alpha_3\beta_3} S^{\ell_1\ell_2}_{\alpha_4\beta_4}\right], \tag{2.223}$$

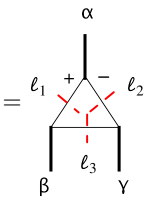
where we've listed the contributions from different internal helicity configuration and $I_2[X]$ is defined as:

$$I_2[X] \equiv \int d^4\ell \frac{X}{\ell^2(\ell-k)^2}, \tag{2.224}$$

where k is the momenta of the stress tensor. The operator $S^{\ell_1\ell_2}_{\alpha_1\alpha_2}$ is a shorthand notation for $\ell_{1\alpha_1\dot{\beta}}\ell_{2\alpha_2}{}^{\dot{\beta}}$. Note that it is understood that the expression must be symmetrized over $\{\alpha_i\}$ and $\{\beta_i\}$ separately, as well as over exchanging $\alpha_i \leftrightarrow \beta_i$, which takes into account the conjugate helicity configurations. For the scalar and and equal helicity fermion contributions, their tensor structure are identical to that of equal helicity gauge field.

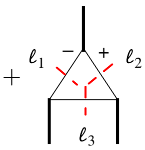
For the three-point function one has:

$$\langle T_{\alpha_1\alpha_2\alpha_3\alpha_4} T_{\beta_1\beta_2\beta_3\beta_4} T_{\gamma_1\gamma_2\gamma_3\gamma_4}\rangle =$$



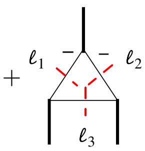
$$\Big\{ I_3\left[S^{\ell_1\ell_2}_{\beta_1\alpha_1} S^{\ell_1\ell_2}_{\beta_2\alpha_2} S^{\ell_1\ell_3}_{\beta_3\gamma_1} S^{\ell_1\ell_3}_{\beta_4\gamma_2} S^{\ell_3\ell_2}_{\gamma_3\alpha_3} S^{\ell_3\ell_2}_{\gamma_4\alpha_4}\right]$$

$$+ \quad I_3\left[S^{\ell_1\ell_2}_{\beta_1\alpha_1} S^{\ell_1\ell_2}_{\beta_2\alpha_2} S^{\ell_2\ell_3}_{\alpha_3\beta_3} S^{\ell_2\ell_3}_{\alpha_4\beta_4} S^{\ell_2\ell_3}_{\gamma_1\gamma_2} S^{\ell_2\ell_3}_{\gamma_3\gamma_4}\right]\Big\} +$$



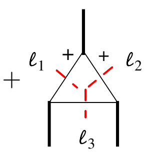
$$\Big\{ I_3\left[S^{\ell_1\ell_2}_{\alpha_1\gamma_1} S^{\ell_1\ell_2}_{\alpha_2\gamma_2} S^{\ell_1\ell_3}_{\alpha_3\beta_1} S^{\ell_1\ell_3}_{\alpha_4\beta_2} S^{\ell_3\ell_2}_{\beta_3\gamma_3} S^{\ell_3\ell_2}_{\beta_4\gamma_4}\right]$$

$$+ \quad I_3\left[S^{\ell_1\ell_2}_{\alpha_1\gamma_1} S^{\ell_1\ell_2}_{\alpha_2\gamma_2} S^{\ell_1\ell_3}_{\alpha_3\gamma_3} S^{\ell_1\ell_3}_{\alpha_4\gamma_4} S^{\ell_1\ell_3}_{\beta_1\beta_2} S^{\ell_1\ell_3}_{\beta_3\beta_4}\right]\Big\} +$$



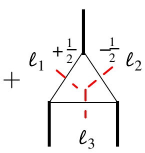
$$\Big\{ I_3\left[S^{\ell_1\ell_3}_{\alpha_1\beta_1} S^{\ell_1\ell_3}_{\alpha_2\beta_2} S^{\ell_2\ell_3}_{\alpha_3\beta_3} S^{\ell_2\ell_3}_{\alpha_4\beta_4} S^{\ell_2\ell_3}_{\gamma_1\gamma_2} S^{\ell_2\ell_3}_{\gamma_3\gamma_4}\right]$$

$$+ \quad I_3\left[S^{\ell_1\ell_3}_{\beta_1\beta_2} S^{\ell_1\ell_3}_{\beta_3\beta_4} S^{\ell_1\ell_3}_{\alpha_1\gamma_1} S^{\ell_1\ell_3}_{\alpha_2\gamma_2} S^{\ell_2\ell_3}_{\alpha_3\gamma_3} S^{\ell_2\ell_3}_{\alpha_4\gamma_4}\right]\Big\} +$$



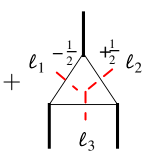
$$\Big\{ I_3\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_1\ell_2}_{\alpha_3\alpha_4} S^{\ell_1\ell_2}_{\beta_1\gamma_1} S^{\ell_1\ell_2}_{\beta_2\gamma_2} S^{\ell_2\ell_3}_{\gamma_3\beta_3} S^{\ell_2\ell_3}_{\gamma_4\beta_4}\right]$$

$$+ \quad I_3\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_1\ell_2}_{\alpha_3\alpha_4} S^{\ell_1\ell_2}_{\beta_1\gamma_1} S^{\ell_1\ell_2}_{\beta_2\gamma_2} S^{\ell_1\ell_3}_{\beta_3\gamma_3} S^{\ell_1\ell_3}_{\beta_4\gamma_4}\right]\Big\} +$$



$$\Big\{ I_3\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_1\ell_2}_{\beta_3\alpha_3} S^{\ell_1\ell_3}_{\beta_1\beta_2} S^{\ell_2\ell_3}_{\gamma_1\gamma_2} S^{\ell_2\ell_3}_{\gamma_3\gamma_4} S^{\ell_2\ell_3}_{\alpha_4\beta_4}\right]$$

$$+ \quad I_3\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_1\ell_2}_{\beta_3\alpha_3} S^{\ell_1\ell_3}_{\beta_1\beta_2} S^{\ell_1\ell_3}_{\beta_4\gamma_3} S^{\ell_2\ell_3}_{\gamma_1\gamma_2} S^{\ell_2\ell_3}_{\alpha_4\gamma_4}\right]\Big\} +$$



$$\Big\{ I_3\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_2\ell_1}_{\gamma_3\alpha_3} S^{\ell_2\ell_3}_{\gamma_1\gamma_2} S^{\ell_1\ell_3}_{\beta_1\beta_2} S^{\ell_1\ell_3}_{\beta_3\beta_4} S^{\ell_1\ell_3}_{\alpha_4\gamma_4}\right]$$

$$+ \quad I_3\left[S^{\ell_1\ell_2}_{\alpha_1\alpha_2} S^{\ell_2\ell_1}_{\gamma_3\alpha_3} S^{\ell_2\ell_3}_{\gamma_1\gamma_2} S^{\ell_2\ell_3}_{\gamma_4\beta_3} S^{\ell_1\ell_3}_{\beta_1\beta_2} S^{\ell_1\ell_3}_{\alpha_4\beta_4}\right]\Big\} \tag{2.225}$$

and $I_3[X]$ is defined as:

$$I_3[X] \equiv \int d^4\ell \frac{X}{\ell_1^2(\ell_1 - k_2)^2(\ell_1 + k_1)^2} \tag{2.226}$$

where k_1, k_2 are the momenta carried by the α_i and β_i indexed stress-tensor respectively. Again symmetrisation interns of $\{\alpha_i\}$, $\{\beta_i\}$, and $\{\gamma_i\}$ are implied and the equal helicity fermion on one of the vertices as well as internal scalars does not produce new tensor structures.

2.9 Outlook

Relativistic quantum mechanics governs the laws of nature at low enough energies so that physics can be described in flat space, with a finite number of interacting particles. "Quantum field theory" is the standard textbook approach to this physics, where, as useful theoretical constructs, "local quantum fields" are introduced, along with the attendant baggage of field redefinition and gauge redundancies, in order to allow a description of the physics in a way compatible with relativistic locality and unitarity. But the on-shell approach to scattering amplitudes suggests that this may not be the only way—that we might instead be able to describe relativistic quantum mechanics without local quantum fields, directly in terms of the physical particles. [18]

In this paper we have taken the first steps to extending the ideas of this on-shell approach to cover particles of all masses and spins in four dimensions. The purely kinematical part of our discussion has been fundamentally trivial—but trivializing the kinematics allows for understanding the structure of the physics as following seamlessly from the foundational principles of Poincare Invariance, Locality, and Unitarity in a satisfying way.

We have seen many aspects of this understanding throughout this paper. The structure of three particle amplitudes, for any mass and spin, is fixed by Poincare invariance. For massless particles, there is a peculiarity for high enough spin—the three particle amplitudes are superficially "non-local" in the sense of having poles; while this doesn't show up in $(3,1)$ signature Minkowski space where these amplitudes vanish, it does mean that consistent factorization at four points is non-trivial, and indeed, all but the usual massless theories we know and love, of interacting spin $(0,1/2,1,3/2,2)$, are ruled out by these considerations. We learn that we can only have a single massless spin two particle, with universal couplings, that the massless spin one particles must have the structure of Yang-Mills theories, and spin $3/2$ requires supersymmetry. Furthermore the mere existence of a consistent amplitude coupling to gravitons rules out all higher spin massless particles.

Similarly there is still a superficial "non-locality" associated with the coupling of a single massive particle to massless particles with spin—the "$x-$ factor"—which again makes factorization non-trivial. Unlike the case for massless particles, we *can* (non-trivially) find consistently factorizing four-particle amplitudes for any choice of three-particle couplings, (with the usual restrictions on consistent couplings to massless spin one and spin 2 particles). But for massive particles of high enough spin, these consistently factorizing amplitudes are badly behaved at high energies—growing with powers of $(p_i \cdot p_j/m^2)$, so that the massless limit cannot be taken smoothly. This tells us that even massive particles of high enough spin cannot be separated by a parametrically large gap from

[18]It is amusing that the on-shell program is often contrasted with the standard approach using Feynman diagrams, since Feynman's primary physical motivation for introducing his diagrams to begin with was to get rid of quantum fields—and he was famously disappointed to learn, via Dyson's proof, that his diagrams were so closely related to field theory after all!

other particles—massive particles with high spin cannot be "elementary". Finally, three particle amplitudes involving all massive particles are local, but naturally have powers of $1/m$. Thus, theories of massive particles can only smoothly interpolate to massless amplitudes at high energies for special choices of spectra and couplings; conversely, starting from massless helicity amplitudes at high energies, we can "unify" subsets of these amplitudes into massive ones in some cases. This can be done for spin 1 and spin $3/2$ particles, representing the on-shell avatars of the Higgs and super-Higgs mechanism, but we can see that gravity can't be "Higgsed" in this way.

In the context of this summary it is perhaps also worth briefly describing the on-shell understanding of the most famous general consequences of relativistic quantum mechanics: the existence of antiparticles and the spin-statistics connection.

The existence of antiparticles is essentially hardwired into the on-shell formalism, since by fiat we are considering analytic functions of Lorentz-invariant kinematical variables, with consistent factorization on all possible channels. To be a little more explicit on these ancient points, we can ask how causality is encoded in the S-matrix in any theory, with or without Lorentz invariance. At tree-level, causality tells us that the amplitude can only have simple poles as a function of energy variables. If the particles have a dispersion relation of the form $E = \omega(\vec{p})$, the poles can be either be of the form $1/(E + \omega(\vec{p}))$, or also $1/(E - \omega(\vec{p}))$ if the interaction Hamiltonian allows particle production. But in a Lorentz invariant theory, neither $(E + \omega(\vec{p}))$ nor $(E - \omega(\vec{p}))$ are individually invariant, so Lorentz invariance and causality forces us to have poles of the form $\frac{1}{(E^2 - \omega(\vec{p})^2)} = \frac{1}{p^2 - m^2}$. This is how we see that causality demands this familiar pole structure at tree-level, which as a byproduct also forces the existence of non-zero amplitudes for the production of degenerate particles and antiparticles.

The on-shell understanding of the connection between spin and statistics is slightly more interesting, and makes use of the universality of coupling to gravity. Indeed we saw vividly that the structure of the four-particle amplitude for gravi-compton scattering off particles of general mass and spin is completely fixed, and in particular forces the correct spin-statistics connection. This deeply relies on the non-triviality of how residues in different channels are consistent with each other, forcing the "s" and "u" channels—related by particle interchange—to have fixed relative signs. It is not surprising that an on-shell understanding of a classic fact related to locality and unitarity should be related to coupling to gravity—after all it is precisely the ability to "weakly gauge" gravity that gives a physical probe (via the existence of an energy momentum tensor) of the locality of quantum field theory. We also described how other famous general results in field theory, such as the Weinberg-Witten and Coleman-Mandula theorems, are interpreted in directly on-shell terms.

Moving beyond tree-scattering, we also took some first steps for computing amplitudes at one-loop, where the on-shell picture is especially powerful, as seen in the speed and transparency of the computation for electron $(g-2)$ and the QCD beta function. While not discussed in this paper, chiral anomalies, together with the possibility of cancelling them via the Green-Schwartz mechanism, also have a beautiful on-shell understanding, arising from the necessity to interpret poles in one-loop amplitudes fixed by generalized unitarity [100].

But of course, much more importantly than providing a conceptually transparent and technically straightforward understanding of standard results, we hope that the formalism introduced in this paper removes the trivial barriers to exploring the new frontier of massive scattering amplitudes, which is filled with fascinating physical questions. We close by listing just a small number of these.

We have focused almost entirely on the computation of tree-level three- and four-particle amplitudes, so one completely obvious question is, for example, how you would extend the BCFW recursion relations to any number of external particles, especially for Higgsed Yang-Mills theories. Of course for massless particles the BCFW shift must be performed for massless particles of appropriate helicity in order to ensure the absence of poles at infinity, so the obvious challenge is that the massive amplitudes unify both the "good" and "bad" helicity combinations into a single object.

Another clear goal is the systematic computation of all the massive amplitudes in the Standard Model, starting at tree level but moving to multi-loop level. It is worth mentioning at least one exciting motivation for this undertaking. Future Higgs factories—like the CEPC or TLEP—can also run on the Z-pole, producing between $10^9 - 10^1 1$ Z particles. Making full use of this data will require a computation of Z-couplings at three to four loop accuracy. And unlike QCD calculations of backgrounds at the LHC, for which the perturbative computations must ultimately be convolved with non-perturbative information such PDF's and hadron fragmentation functions to connect with experiment, these precision electroweak calculations are unaffected by hadronic uncertainties at the needed level of precision, so any theoretical predictions can be unambiguously connected to exquisitely precise experimental measurements!

It is also clearly of interest to investigate massive amplitudes in supersymmetric theories, as this should of course be especially interesting in the context of the $\mathcal{N} = 4$ SYM on the Coulomb branch. Now even our first look at the on-shell avatar of the Higgs and Super-Higgs mechanisms showed that the Higgsed amplitudes are *more* unified than their massless counterparts. Thus we should expect that all the natural objects encountered for massless amplitudes—such as tree amplitudes, leading singularities, and on-shell diagrams, which are separated into different "k" sectors—are somehow unified into more interesting objects. Amongst other things the extension of BCFW to the Higgsed theories might be most natural in the massive $\mathcal{N} = 4$ on-shell diagram formulation. And of course it would be fascinating to see if the Grassmannian/Amplituhedron structures underlying the theory and the origin of the moduli space is somehow extended/deformed away from the origin.

All of the physics we have discussed in this paper has revolved around the consistency of long-distance physics: the on-shell focus on factorization and cuts at tree and loop level is meant to ensure that infrared singularities needed by locality and unitarity are correctly accounted for, and this fixes the structure of the amplitudes. For theories with growing amplitudes in the ultraviolet, needing a UV completion, it is very natural to ask the same questions: can the physics of UV completion also be determined from the consistency conditions of locality and unitarity? If the UV completion has a weak coupling, the question becomes perfectly sharply posed, and in the context of unitarizing the Fermi interaction or WW scattering, searching for a tree-level UV completion correctly led to the prediction of massive W particles and Higgses as the completion of the weak interactions. Turning to the even more famous problem of UV completion for gravity scattering amplitudes, we encounter a well-known novelty. As will be discussed at greater length in [9], any weakly coupled UV completion for gravity amplitudes, (or for that matter, also Yang-Mills or ϕ^3 theory, any theory with non-trivial three-particle amplitudes), must involve an infinite tower of particles with infinitely increasing spins, as of course familiar from string theory. It is a tantalizing prospect to try and "derive string theory" in this way, as giving the only possible consistent tree scattering amplitudes for gravitons coupled to the infinite tower of massive higher spin particles necessary for UV completion. But consideration of amplitudes involving massive higher spin particles is necessary for any possible uniqueness, since as shown in [9], deformations of the string scattering amplitudes with only gravitons as external particles, compatible with all the standard rules, have been identified. This is not at all surprising. Since we know the presence of gravity makes massless higher spin particles impossible, the coexistence of gravity unified with an infinite tower of massive higher spin particles must involve the strongest consistency conditions imaginable. Again, the massive amplitude formalism we have discussed in this paper trivializes kinematical issues so that important physics points can be studied with an unobstructed view, and with this in hand we will return to string theory and the challenge of UV completion in [9].

Acknowledgements

We are grateful to Zvi Bern, Rutger Boels, Nathaniel Craig, Lance Dixon, and David Kosower for interesting discussions and detailed comments on the draft. We also thank Ming-Zhi Chung, Neil Christensen, and Alexander Ochirov for comments for the revised version of the draft. We would especially like to thank the referee's tremendous effort in carefully working through the manuscript, giving us valuable feedbacks. The work of NAH is supported by the DOE under grant DOE DE-SC0009988. Y-t Huang is supported by MoST grant 106-2628-M-002 -012 -MY3. Y-t Huang would like to thank the Institute for Advanced Study for its hospitality during the completion of this work. T.-C. Huang is supported by the Orr Fellowship.

Chapter 3

THE EFT-HEDRON

[1]

3.1 Introduction

There is a long-appreciated, close connection between vacuum stability, causality, and unitarity, and analyticity/positivity properties of scattering amplitudes, going back to the 1960's S-matrix program. In this standard story, there are three fundamental origins of positivity: the positivity of energies (vacuum stability), the sharp localization of signals inside the lightcone (causality, and the positivity of probabilities (unitarity). These basic positivities, together with analyticity properties of scattering amplitudes meant to reflect causality, allow the derivation of more non-trivial positivity constraints on coefficients of higher-dimension operators in low-energy effective field theories (as in [4, 5, 143]). In recent years, a sort of opposite of the S-matrix program has emerged in a number of theories, where notions of positivity take a central role, determining certain "positive geometries" in the the kinematic space of particle scattering with a fundamentally combinatorial definition, from which the amplitudes are naturally extracted. In this picture, locality and unitarity are not taken as fundamental principles, but instead arise, joined at the hip, from the the study of the boundary structure of the positive geometries. These examples suggest that there is vastly more "hidden positivity" in scattering amplitudes than meets the eye, with locality and unitarity as *derived from*, rather than *the origin of*, positivity properties.

Motivated by these discoveries, in this paper we will revisit the positivity properties of $2 \to 2$ scattering amplitudes, and re-examine the *usual* positivity properties dictated by analyticity, causality, and unitarity. We will find that there are *infinitely many* constraints on the coefficients of higher-dimension operators, and that these constraints involve very similar mathematical structures as have already been seen in the story of positivity geometries and amplituhedra.

To illustrate the nature of the constraints, consider for simplicity the scattering amplitudes for two massless scalars $ab \to ab$, and suppose we are working in an approximation

[1]This chapter is adapted from Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-Tin Huang. "The EFT-Hedron". In: *JHEP* 05 (2021), p. 259. DOI: `10.1007/JHEP05(2021)259`. arXiv: `2012.15849 [hep-th]`.

where we have integrated out massive states but not yet accounted for massless loops in the low-energy theory. Then, the low-energy amplitude has a power-series expansion in the Mandelstam variables s, t:

$$\mathcal{A}(s,t) = \sum_{\Delta,q} a_{\Delta,q} s^{\Delta-q} t^q \tag{3.1}$$

and all the information in the low-energy effective field theory is captured in the coefficients $a_{\Delta,q}$ which we can organize into a table:

$$\begin{array}{cccccc} & q{=}0 & 1 & 2 & 3 & \cdots \\ \Delta{=}1 & a_{1,0} & a_{1,1} & & & \\ \Delta{=}2 & a_{2,0} & a_{2,1} & a_{2,2} & & \\ \Delta{=}3 & a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}, \tag{3.2}$$

There are infinitely many constraints on the $a_{\Delta,q}$, forcing this infinite table of coefficients to lie inside "the EFT-hedron".

These constraints quantify certain intuitions about "garden variety" higher dimension operators contributing to $ab \to ab$ scattering, into sharp bounds. For instance we shouldn't expect operators of the same mass dimension Δ to have vastly different coefficients; these correspond to the coefficients in the same row in our table. But we might also think that this is a consequence of "naturalness", and that by fine-adjustments of the parameters in the high-energy theory, we can engineer any possible relative sizes between these operators we like. The EFT-hedron shows that this is not the case: not everything goes, and indeed the coefficients $a_{\Delta,q}$ for a fixed Δ must satisfy linear inequalities, which force them to lie inside a certain polytope. We would also expect all operators to be suppressed by a similar scale, i.e. not to have dimension 6 operators suppressed by the TeV scale while dimension 8 operators are suppressed by the Planck scale, though again one might think this can be done with suitable fine-tuning. Again, the EFT-hedron shows this is impossible, and imposes non-linear inequalities between different $a_{\Delta,q}$, which in the simplest case constrain the relative sizes of coefficients at fixed q, in a fixed column of the table. We will initiate a systematic study of the EFT-hedron in this paper. But before diving in, let us give a high-level overview of the physical and mathematical engines at work.

The physical starting point is a dispersive representation of $2 \to 2$ scattering amplitudes, as a function of s working at fixed t. To begin with we will assume, as mentioned above, that we integrate out massive states of some typical mass M, which generates higher-dimension operators in the low-energy theory, and for the purpose of these introductory comments let us ignore the further running of these higher dimension operators by massless loops in the low-energy theory (we will revisit this point in the body of the paper). Working at fixed t with $|t| \ll M^2$, it can be argued that the amplitudes only have singularities on the real s axis, with discontinuities reflecting particle production in the s and u channels. The discontinuity across these cuts has a partial wave expansion, as a sum over spins with positive coefficients. Furthermore, causality is reflected in a bound on the amplitude at large s for fixed $|t|$. In a theory with a mass gap, we have the Froissart bound telling us the amplitude is bounded by $\mathcal{A} < s \log^2 s$. In quantum gravity, we expect that for any UV completion with a weak coupling (like in string theory), the high-energy amplitude in the physical region, with fixed negative t, is bounded by $\mathcal{A} < s^p$ with $p < 2$. Thus at fixed t, for any theory, we have a dispersive representation for the amplitude at fixed t, of the form

$$\mathcal{A}(s,t) = \mathcal{A}_0(t)+\mathcal{A}_1(t)s+\int dM^2 \sum_l p_l(M^2)G_l(1+\frac{2t}{M^2})\left(\frac{1}{s-M^2}+\frac{1}{u-M^2}\right), \tag{3.3}$$

where $G_l(x)$ are Gegenbauer polynomials.

Now, this dispersive representation has the two basic and crucial long-appreciated positivities we have alluded to: the positivity of energies is reflected in $M^2 > 0$, and the positivity of probabilities in $p_l(M^2) > 0$. The new surprise we will explore in this paper, are further hidden positive structures associated with the propagator $1/(s-M^2)$, and with the Gegenbauer polynomials $G_l(x)$. It is these new positivities that are responsible for the non-trivial geometry of the EFT-hedron and the associated infinite number of new constraints on the $a_{\Delta,q}$. Here we content ourselves here with summarizing the basic mathematical facts of these hidden positivities, whose consequences we will explore in detail in the body of the paper.

Let's begin with the positivity associated with propagators, which can be illustrated in a simplified setting, where we imagine a dispersive representation for a function $F(s)$ of the form

$$F(s) = \int dM^2 \frac{p(M^2)}{M^2 - s}. \tag{3.4}$$

This has a power-series expansion at small s, $F(s) = \sum_n f_n s^n$, where

$$f_n = \int dM^2 \frac{p(M^2)}{M^2}(\frac{1}{M^2})^n. \tag{3.5}$$

This can be interpreted geometrically as saying that the vector $\mathbf{f} = (f_0, f_1, f_2, \cdots)$ lies in the convex hull of the continuous moment curve $(1, x, x^2, \cdots)$, where here $x = 1/M^2$, so we also impose that $x > 0$. Thus we have a well-posed mathematical question: what is the region in $\mathbf{f}$ space that is carved out by the convex hull of the half-moment curve with $x > 0$? This question has a beautifully simple answer. To begin with, we associate a "Hankel matrix" $\mathbf{F}$ with the vector $\mathbf{f}$ via $\mathbf{F}_{ij} = \mathbf{f}_{i+j}$:

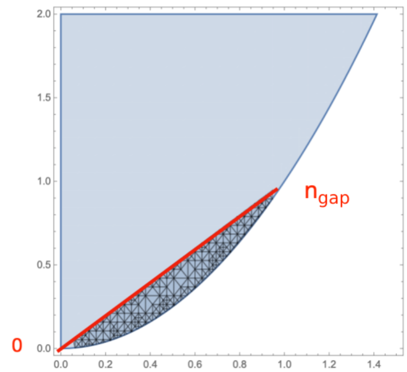
$$\mathbf{F} = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots \\ f_1 & f_2 & f_3 & \cdots \\ f_2 & f_3 & f_4 & \cdots \\ f_3 & f_4 & f_5 & \cdots \end{pmatrix}. \tag{3.6}$$

Then the allowed region in $\mathbf{f}$ space is completely specified by demanding that *all* of the square $k \times k$ minors of the Hankel matrix $\mathbf{F}$ are positive! This is abbreviated by saying the $\mathbf{F}$ is a "totally positive" matrix. For $k = 1$, this just tells us that all the f_n are positive, which is essentially the amplitude positivity found in the early works of [4]. But there are also infinitely many non-linear positivity conditions. It is striking to see "all minors of a matrix positive" conditions–earlier seen in the context of the positive grassmannian [13] and the amplituhedron [11] for $\mathcal{N} = 4$ SYM, show up again in a different setting, and in such a basic way, for completely general theories.

Note that all these conditions are homogeneous in the mass dimension of the operators, as they should be, since we have not input any further knowledge of the UV mass scales. But suppose we were also given the gap M_{gap} to the first massive states. In this case, the vector $\mathbf{f}$ would lie in the convex hull of the moment curve, starting at $x = 0$ and cut-off at $x = x_{gap} = 1/M_{gap}^2$. Working in units where $M_{gap} = 1$, the region in $\mathbf{f}$ space is carved out by looking not only at f, but also at its discrete derivatives,

$$\begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix}, \begin{pmatrix} f_1 - f_2 \\ f_2 - f_3 \\ f_3 - f_4 \\ \vdots \end{pmatrix}, \begin{pmatrix} (f_2 - f_3) - (f_3 - f_4) \\ (f_3 - f_4) - (f_4 - f_5) \\ (f_4 - f_5) - (f_5 - f_6) \\ \vdots \end{pmatrix}, \cdots \tag{3.7}$$

and demanding that the Hankel matrices associated with all of these vectors are totally positive. A simple illustration of the region in $(f_1/f_0, f_2/f_0)$ space carved out with (patterned region) and without knowledge of the gap is shown in the following plot:



.

Now to illustrate Gegenbauer positivity, let us again focus on simplest example illustrating the non-trivial point. Consider a dispersive representation for some function $F(s,t)$ only containing s-channel (and no u-channel) poles:

$$F(s,t) = \int dM^2 \sum_l \frac{p_l(M^2) G_l(1{+}2t/M^2)}{M^2{-}s} \tag{3.8}$$

and consider the low-energy expansion in powers of $s, (2t)$, as

$$F(s,t) = \sum_{\Delta,q} f_{\Delta,q} s^{\Delta-q} (2t)^q,$$

yielding

$$\begin{pmatrix} f_{\Delta,0} \\ f_{\Delta,1} \\ \vdots \\ f_{\Delta,\Delta} \end{pmatrix} = \sum_l P_l \begin{pmatrix} G_l^{(0)}(x{=}1) \\ G_l^{(1)}(x{=}1) \\ \vdots \\ G_l^{(D)}(x{=}1) \end{pmatrix} \text{ where } P_l = \int dM^2 \frac{p_l(M^2)}{(M^2)^{\Delta+1}} > 0\,. \tag{3.9}$$

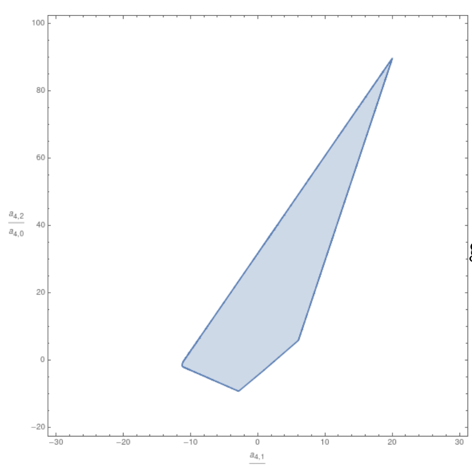
Here $G_l^{(q)}(x=1)$ are the q'th derivatives of the Gegenbauer polynomials, evaluated at the "forward limit" where $x = 1$. The above expression tells us that the projective vector $\mathbf{f}_\Delta = (f_{\Delta,0}, \cdots, f_{\Delta,\Delta})$ lies in the convex hull of all the "Gegenbauer derivative" vectors. Finding the space of all consistent $\mathbf{f}_\Delta$ is then a standard polytope problem: we are given a collection of vectors (an infinite number in this case) whose convex hull specifies some polytope, and we'd like to determine how to characterize the polytope instead by the inequalities that cut out its facets. As we will review in the body of the paper, the facet structure of a Δ-dimensional polytope, in turn, is fully captured by the knowledge of the signs of the all the determinants made from any $(\Delta{+}1)$ vectors of the vertices. In our context, then, we should look at the infinite "Gegenbauer matrix" $G_{l,q} = G_l^{(q)}(x=1)$, and consider the top $\Delta{+}1$ rows of this matrix and look at all the corresponding $(\Delta{+}1) \times (\Delta{+}1)$ minors. Remarkably, it turn out that all these minors of the Gegenbauer matrix are positive! This is another appearance of the "matrix with all positive minors" phenomenon, and it immediately allows us to fully determine the inequalities cutting out the corresponding polytope in $\mathbf{f}$ space, which are the famous "cyclic polytopes". Cyclic polytopes have already made a prominent appearance in the story of $\mathcal{N}=4$ SYM amplitudes, as the simplest example of "amplituhedra" for the case of next-to-MHV tree scattering amplitudes. Indeed tree amplituhedra can be thought of as grassmannian generalizations of the notion of cyclic polytopes. It is again interesting to see the same objects show up in the totally different, very general setting of the EFT-hedron. A morally similar geometry was seen in the conformal bootstrap [8].

We close our introductory remarks with two comments. First, we stress that these constraints on effective field theory are non-trivial statements about any theory, and in particular non-trivial constraints on quantum gravity in the real wold. Of course we don't usually care about the relative sizes of very high dimension "garden variety" operators, for phenomenological purposes, but we nonetheless find it fascinating that the structure of low-energy dynamics is vastly more constrained than previously appreciated. As a sampling of our results, let's look at some of the constraints for photon and graviton scattering. For the $(-,-,+,+)$ helicity configuration, where the helicities are identical in the s-channel, the amplitude for the D^8F^4 and D^8R^4 operator takes the form:

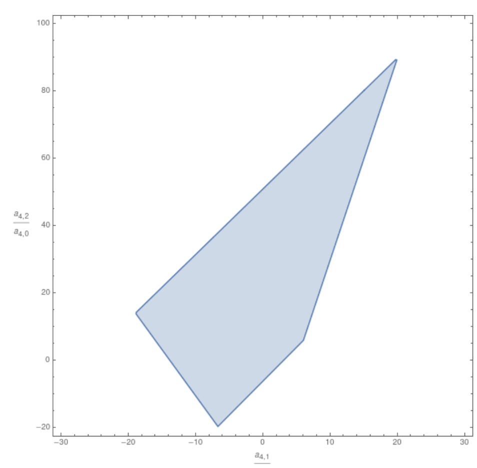
$$\langle 12\rangle^{2h}[34]^{2h}(a_{4,0}s^4{+}a_{4,1}s^3t{+}a_{4,2}s^2t^2\cdots)\,, \tag{3.10}$$

where $h=1,2$ for photon and graviton respectively. The allowed region for $\frac{a_{4,1}}{a_{4,0}}$, $\frac{a_{4,2}}{a_{4,0}}$ is given as:

$photon$



$graviton$



Note that the allowed region is bounded.

It is also important to note that, while the EFT-hedron places extreme constraints on the effective field theory expansion, sensible effective field theories do not appear to populate the entire region allowed by the EFT-hedron, but cluster close to its boundaries. The reason is likely that the physical constraints we have imposed, while clearly necessary, are still not enough to capture consistency with fully healthy UV theories. In particular, our dispersive representation at fixed t, does not make it easy to impose the softness of high-energy, fixed-angle amplitudes where both s, t are large with t/s fixed. It would be fascinating to find a way to incorporate this extra information about UV softness into the constraints.

Having given this high-level overview of the physical and mathematical basis for our results, we proceed to a more systematic discussion. Through sections 3.2, 3.3, 3.4, 3.5 we will present an elementary introduction of EFT amplitudes with explicit examples, the analytic definition of $a_{D,q}$ through dispersion relations and their potential obstructions, and finally the theory space that emerges from the dispersive representation. Next in sec.3.6, we take a brief sojourn in the positive geometries relevant to our analysis, giving a pedagogical discussion of convex hulls of moment curves and cyclic polytopes. These geometries will be immediately utilized to define the s-channel EFT-hedron in sec.3.7, where we focus on the theory space for scalar EFTs that allow for preferred ordering and hence the absence of u-channel thresholds. This will be generalized to include u-channel thresholds in sec.3.8, as well as photon and gravitons in sec. 3.9. We will study explicit examples of EFTs and their "positions" in the EFT-hedron in sec. 3.10. Finally IR logarithms generated by the massless loops will be incorporated in sec.3.11.

Many of the results of this paper have been presented in conferences and schools over the past few years [2]. As we were preparing our manuscript, a number of independent works appeared on the arxiv overlapping with some of this work. In particular, new positivity constraints involving scale dependent "arc moments" were introduced in [24], and are intimately related to the geometry of the gap discussed in subsection 3.7. These constraints arise from the knowledge of the precise UV cut off, and hence the reach of validity for the EFT description. Bounds involving the combination of positivity away from the forward limit and full permutation invariance was discussed in [164] and [46],

[2]See for example, N. Arkani-Hamed and Y.-T. Huang, *talk at Strings 2018; Lectures at the cern winter school on supergravity, strings and gauge theory (2019); talk at UV Meets the IR: Effective Field Theory Bounds from QFT to String Theory KITP 2020.*

which have some overlap with the s-u polytope discussion in subsection 3.8. Other related works can be found in [22, 146, 147, 155].

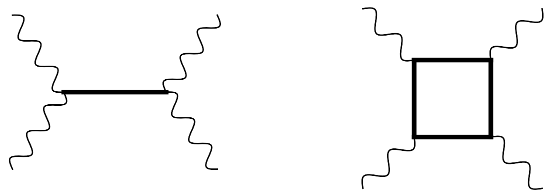


Figure 3.1: Different origins for the EFT: (I) Integrating away massive states in tree exchanges, for example the Higgs for the Sigma model and the infinite tower of higher spin states in string amplitudes, (II) or massive states in the loop, for example the φX^2 coupling.

3.2 EFT from the UV

Let's begin by considering a few concrete examples of EFTs emerging from their UV parent amplitudes. We will give a broad stroke description of what types of high energy theories/amplitudes they can arise from, the features that we will be focusing on, and their relations to local operators, leaving the detailed analysis for the the remainder of the paper.

Explicit EFT amplitudes

The amplitude for the low energy degrees of freedom may originate from a UV amplitude where they interact through a tree-level exchange of massive particles. A simple example is the case of the linear sigma model in the broken phase:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 - \frac{m_h^2}{2}h^2 + \left(1+\frac{h}{v}\right)^2 \frac{1}{2}(\partial\pi\cdot\partial\pi) + V(h). \tag{3.11}$$

where $v = m_h\sqrt{\frac{2}{\lambda}}$, λ is the quartic coupling for the potential in the unbroken phase. As the massless Goldstone boson π couples to the massive Higgs via cubic coupling $\pi^2 h$, the following four π amplitude in the UV is given by (see fig3.1):

$$M(s,t) = -\frac{\lambda}{8m_h^2}\left(\frac{s^2}{s-m_h^2}+\frac{t^2}{t-m_h^2}+\frac{u^2}{u-m_h^2}\right), \tag{3.12}$$

where $s = (p_1+p_2)^2$, $t = (p_1+p_4)^2$, and $u = (p_1+p_3)^2$, and as the pions are massless $s+t+u = 0$. In the center of mass frame, we have $s = E_{CM}^2$ as the center of mass energy and $t = -\frac{s}{2}(1-\cos\theta)$, where θ is the scattering angle. At low energies, all Mandelstam variables are small compared to the UV scale m_h, and thus the low energy EFT amplitude is obtained by expanding in $\frac{p^2}{m_h^2} \ll 1$,

$$M^{\mathrm{IR}}(s,t) = \frac{\lambda}{8m_h^2}\left(\frac{s^2+t^2+u^2}{m_h^2}+\frac{s^3+t^3+u^3}{m_h^4}+\cdots\right) = \frac{\lambda}{8}\sum_{n=2}^{\infty}\frac{\sigma_n}{m_h^{2n}}, \tag{3.13}$$

where $\sigma_n = s^n+t^n+u^n$. We see that the IR description is given by an infinite series of polynomial terms, reflecting the presence of an infinite number of higher dimensional operators from integrating out the massive Higgs.

Note that the residues of the poles for the UV amplitude eq.(3.12), say in the s-channel, are constants. This reflects the fact that the exchanged particle is spinless. In general a spin-J exchange in the s-channel will lead to a residue that is polynomial in t up to degree J. For example, consider the four-gluon amplitude of type-I open string theory, given by

$$M(1^-2^-3^+4^+) = -g_s\alpha'^2\langle 12\rangle^2[34]^2\frac{\Gamma[-\alpha's]\Gamma[-\alpha't]}{\Gamma[1-\alpha's-\alpha't]}\,, \tag{3.14}$$

where we have put the gauge bosons in a four-dimensional subspace and thus the helicity dependence is carried by the spinor brackets. The definition of these brackets as well as their relation to the local operators will be introduced shortly. Here g_s is the string coupling and in this paper we will set the string scale $\alpha' = 1$. The gamma functions in the numerator have poles at $s, t \in \mathbb{N}^+$, reflecting an infinite number of massive states. The residue at $s = n$ is given by

$$g_s\langle 12\rangle^2[34]^2\frac{(-)^n}{n!}\prod_{i=1}^{n-1}(t+i)\,, \tag{3.15}$$

where the non-trivial dependence in t reflects the spinning nature of the exchanged particle. Since $\alpha' = 1$ low energy is simply $p^2 \ll 1$, and the low energy amplitude is given as:

$$M^{\rm IR}(1^+2^+3^-4^-) = g_s\langle 12\rangle^2[34]^2\left(-\frac{1}{st}+\zeta_2+\zeta_3(s+t)+\cdots\right)\,, \tag{3.16}$$

where the leading term contains massless poles corresponding to the field theory Yang-Mills piece. The coefficients for the polynomials are now zeta values $\zeta_n \equiv \sum_{\ell=1}^{\infty}\frac{1}{\ell^n}$, reflecting the fact that each term in the polynomial expansion receives contribution from the infinite number of UV states at integer values of m^2. The same feature can be found for the four-graviton amplitude of type-II closed string theory:

$$M(1^{-2}2^{-2}3^{+2}4^{+2}) = g_s^2\langle 12\rangle^4[34]^4\frac{\Gamma[-s]\Gamma[-t]\Gamma[-u]}{\Gamma[1+s]\Gamma[1+u]\Gamma[1+t]}\,, \tag{3.17}$$

where the low energy expansion gives:

$$M^{\rm IR}(1^{-2}2^{-2}3^{+2}4^{+2}) = M(s,t)|_{\alpha'\to 0} = G_N\langle 12\rangle^4[34]^4\left(\frac{-1}{stu}+2\zeta_3+\zeta_5\sigma_2+2\zeta_3^2stu\cdots\right)\,. \tag{3.18}$$

The leading piece with the massless poles $\frac{1}{stu}$ correspond to the contribution from the Einstein-Hilbert term and we've identified $G_N = g_s^2$.

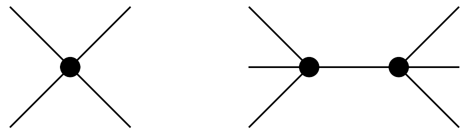


Figure 3.2: An operator of four fields will contribute to the four-point amplitude as a polynomial, and the six-point amplitude as a rational term.

Instead of tree-level exchanges, the massive UV states can also contribute via loop process. For example, consider a massless scalar φ coupled to massive X via $\lambda\varphi X^2$. In the UV four φs can interact through a massive X loop, and the amplitude is simply the scalar box-integral (see fig3.1):

$$M(s,t)=\lambda^4 \int\frac{d^4\ell}{(2\pi)^4}\frac{1}{[\ell^2-m_X^2][(\ell-p_1)^2-m_X^2][(\ell-p_1-p_2)^2-m_X^2][(\ell+p_4)^2-m_X^2]} +perm(2,3,4)\,. \tag{3.19}$$

The analytic result of the box integral is given as [59]:

$$I_4[s,t]=\frac{1}{(4\pi)^2}\frac{uv}{8\beta_{uv}}\Bigg\{2\log^2\left(\frac{\beta_{uv}+\beta_u}{\beta_{uv}+\beta_v}\right)+\log\left(\frac{\beta_{uv}-\beta_u}{\beta_{uv}+\beta_u}\right)\log\left(\frac{\beta_{uv}-\beta_v}{\beta_{uv}+\beta_v}\right)-\frac{\pi^2}{2}$$
$$+\sum_{i=u,v}\left[2\mathrm{L}i_2\left(\frac{\beta_i-1}{\beta_{uv}+\beta_i}\right)-2\mathrm{L}i_2\left(-\frac{\beta_{uv}-\beta_i}{\beta_i+1}\right)-\log^2\left(\frac{\beta_i+1}{\beta_{uv}+\beta_i}\right)\right]\Bigg\}\,, \tag{3.20}$$

where $u=-\frac{4m_X^2}{s}$ and $v=-\frac{4m_X^2}{t}$, and

$$\beta_u=\sqrt{1+u},\quad \beta_v=\sqrt{1+v},\quad \beta_{uv}=\sqrt{1+u+v}\,. \tag{3.21}$$

This gives the following low energy expansion:

$$M^{\mathrm{IR}}(s,t) \;=\; \frac{g^4}{2m_X^4}\left(1+\frac{1}{5!}\frac{\sigma_2}{m_X^4}+\frac{20}{7!3}\frac{\sigma_3}{m_X^6}+\frac{2}{7!3}\frac{\sigma_2^2}{m_X^8}+\frac{1}{6!33}\frac{\sigma_3\sigma_2}{m_X^{10}}+\cdots\right)\,. \tag{3.22}$$

Note that in general for identical scalars, the polynomial part of the four-point amplitude can be expanded on the basis of two permutation invariant polynomials σ_2 and σ_3.

From local amplitudes to local operators

In this paper we are interested in theories whose IR description admits an expansion in terms of local operators, i.e. $\mathcal{L}=\mathcal{L}_{kin}+\mathcal{L}_I[\phi,\partial\phi]$, with $\mathcal{L}_I[x]$ being polynomial functions. A local operator that contains n fields, for example $(\partial\phi\cdot\partial\phi)\phi^{n-2}$, will contribute to the

n-point scattering amplitude as a polynomial of Mandelstam invariants $s_{i,j}$. At higher points, it appears in factorization channels, contributing to the residue of rational terms, as illustrated in fig. 3.2. This translates to the low energy four-point amplitude taking the form:

$$M^{\text{IR}}(s,t) \equiv M(s,t)|_{s,t\to 0} = \{\text{m}assless\ \ poles\} + \{\text{p}olynomials\}\,, \tag{3.23}$$

where $\{\text{m}assless\ \ poles\}$ reflect the presence of cubic operators, and $\{\text{p}olynomials\}$ quartic ones. The coefficients of the cubic operators appear in the residue for the $\{\text{m}assless\ \ poles\}$, while that of quartic operators are linearly mapped in to the Taylor coefficients in $\{\text{p}olynomials\}$. Here we have ignored the logarithms arisings massless loops. These effects are of course intimately tied with what we mean by EFT coefficients, as they inevitably run. However, for the sake of simplicity in our presentation, we will focus on tree-level EFT amplitudes for now, and assign section 3.11 to discuss how these results extend to the situation where massless loops are present.

Let's begin with operators involving only scalars. First, since the momentum inner products vanish for three-point kinematics,

$$p_3^2 = (p_1{+}p_2)^2 = 2p_1\cdot p_2 = 0\,, \tag{3.24}$$

the only non-trivial three-point amplitude is a constant. In terms of cubic operators, this is a reflection of the fact that any three-scalar operator with derivatives much vanish via equations of motion:

$$(\partial\phi\cdot\partial\phi)\phi \sim \phi^2\Box\phi = 0\,, \tag{3.25}$$

i.e. it can be removed by a field redefinition. At four-points the amplitude can be expressed as:

$$M^{\text{IR}}(s,t) = \{\text{m}assless\ \ poles\} \quad + \sum_{k,q} a_{k,q}\, s^{k-q}t^q\,. \tag{3.26}$$

Here k labels the total degree in Mandelstam variables, q the degree in t,. This labeling will be convenient for considering the expansion near the forward limit, i.e. $t=0$. For fixed k these correspond to dimension $2k{+}4$ operators in four-dimensions. For example, $(\partial\phi\cdot\partial\phi)^2$, $(\partial\phi\cdot\partial\phi)(\partial^2\phi\cdot\partial^2\phi)$, translate to

$$(\partial\phi\cdot\partial\phi)^2 \to (2s^2{+}2t^2{+}2st), \quad (\partial\phi\cdot\partial\phi)(\partial^2\phi\cdot\partial^2\phi) \to -st^2 - s^2t\,. \tag{3.27}$$

Thus the coefficients of the EFT operators are translated into the coefficients of the polynomials $s^{k-q}t^q$. Note that we do not have an $k=1$ operator $(\partial\phi\cdot\partial\phi)\phi^2$, since on-shell it vanishes by momentum conservation $s{+}t{+}u=0$. Once again, as with the three-point example, this illustrates the important advantage of such "on-shell basis": it is free from field redefinition or integration by parts ambiguities.

Generally, it is unnatural for scalars to be massless unless they're Goldstone bosons for some broken symmetry. Thus the degrees of freedom in low energy effective field theories are more naturally associated with photons and gravitons, and the local operators are built out of field strengths and Riemann tensors (Ricci tensor and scalars vanish under Einstein equations). Their imprint on the amplitudes can be more conveniently captured by the spinor-helicity variables, where one express the momenta as:

$$p_{i\mu} \to p_{i\alpha\dot{\alpha}} = p_{i\mu}(\sigma^{\mu})_{\alpha\dot{\alpha}} = \lambda_{i\alpha}\tilde{\lambda}_{i\dot{\alpha}}\,. \tag{3.28}$$

Under the massless U(1) little group, these transforms as $\lambda_{i\alpha} \to e^{-i\frac{\theta_i}{2}}$ and $\tilde{\lambda}_{i\dot{\alpha}} \to e^{i\frac{\theta_i}{2}}\tilde{\lambda}_{i\dot{\alpha}}$. The polarization vectors are then expressed as

$$\varepsilon^{+}_{i\alpha\dot{\alpha}} = \frac{1}{\sqrt{2}}\frac{\eta_{\alpha}\tilde{\lambda}_{i\dot{\alpha}}}{\langle i\eta\rangle}, \quad \varepsilon^{-}_{i\alpha\dot{\alpha}} = \frac{1}{\sqrt{2}}\frac{\tilde{\lambda}_{i\alpha}\eta_{\dot{\alpha}}}{[i\eta]} \tag{3.29}$$

where $\langle ij\rangle = \lambda_i^{\alpha}\lambda_{j\alpha} = \epsilon^{\alpha\beta}\lambda_{i\beta}\lambda_{j\alpha}$, and $[ij] = \tilde{\lambda}_{i\dot{\alpha}}\tilde{\lambda}_j^{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\lambda_i^{\dot{\beta}}\lambda_j^{\dot{\alpha}}$. Here η are the reference spinors parameterizing the gauge redundancy associated with the polarization vectors, and drops out for any gauge invariant quantity. Polarization tensors are just the square of these vectors. It is straightforward to see, in terms of these on-shell variables, the field strength and the linear part of Riemann tensor are expressed as:

$$\begin{aligned}
F_{\mu\nu} &\to F_{\alpha\dot{\alpha},\beta\dot{\beta}} = F^{+}_{\dot{\alpha}\dot{\beta}}\,\epsilon_{\alpha\beta} + F^{-}_{\alpha\beta}\,\epsilon_{\dot{\alpha}\dot{\beta}}, \quad F^{+}_{\dot{\alpha}\dot{\beta}} = \sqrt{2}\tilde{\lambda}_{\dot{\alpha}}\tilde{\lambda}_{\dot{\alpha}}, \quad F^{-}_{\dot{\alpha}\dot{\beta}} = \sqrt{2}\lambda_{\alpha}\lambda_{\alpha}, \\
R_{\mu\nu\rho\sigma} &\to R_{\alpha\dot{\alpha}\beta\dot{\beta}\gamma\dot{\gamma}\delta\dot{\delta}} = \epsilon_{\alpha\beta}\epsilon_{\gamma\delta}\,R^{+}_{\dot{\alpha}\dot{\beta}\dot{\gamma}\dot{\delta}} + \epsilon_{\dot{\alpha}\dot{\beta}}\epsilon_{\dot{\gamma}\dot{\delta}}\,R^{-}_{\alpha\beta\gamma\delta} \\
&\quad R^{+}_{\dot{\alpha}\dot{\beta}\dot{\gamma}\dot{\delta}} = \sqrt{2}\tilde{\lambda}_{\dot{\alpha}}\tilde{\lambda}_{\dot{\beta}}\tilde{\lambda}_{\dot{\gamma}}\tilde{\lambda}_{\dot{\delta}}, \quad R^{-}_{\alpha\beta\gamma\delta} = \sqrt{2}\lambda_{\alpha}\lambda_{\beta}\lambda_{\gamma}\lambda_{\delta}\,,
\end{aligned} \tag{3.30}$$

where the $\pm$ superscript indicates the $\pm h$ helicity of the polarization (tensors) vector. Indeed up to an overall constant, the above form is uniquely fixed by the little group scaling and dimension analysis.

Thus polynomials of spinor brackets can be straightforwardly translated to local operators of field strengths and Riemann tensors. For example, for the three-point amplitude, possible polynomial representation for self interacting spin-1 and 2 particles can be immediately translated into F^3 and R^3 operators:

$$\begin{aligned}
M_3(1^{-}2^{-}3^{-}) &\to 2\sqrt{2}\langle 12\rangle\langle 23\rangle\langle 31\rangle = (F_1^{-})_{\alpha}{}^{\beta}(F_2^{-})_{\beta}{}^{\gamma}(F_3^{-})_{\gamma}{}^{\alpha} \\
M_3(1^{+}2^{+}3^{+}) &\to 2\sqrt{2}[12][23][31] = (F_1^{+})_{\dot{\alpha}}{}^{\dot{\beta}}(F_2^{+})_{\dot{\beta}}{}^{\dot{\gamma}}(F_3^{+})_{\dot{\gamma}}{}^{\dot{\alpha}} \\
M_3(1^{-2}2^{-2}3^{-2}) &\to 2\sqrt{2}\langle 12\rangle^2\langle 23\rangle^2\langle 31\rangle^2 = (R_1^{-})_{\alpha_1\alpha_2}{}^{\beta_1\beta_2}(R_2^{-})_{\beta_1\beta_2}{}^{\gamma_1\gamma_2}(R_3^{-})_{\gamma_1\gamma_2}{}^{\alpha_1\alpha_2} \\
M_3(1^{+2}2^{+2}3^{+2}) &\to 2\sqrt{2}[12]^2[23]^2[31]^2 = (R_1^{+})_{\dot{\alpha}_1\dot{\alpha}_2}{}^{\dot{\beta}_1\dot{\beta}_2}(R_2^{+})_{\dot{\beta}_1\dot{\beta}_2}{}^{\dot{\gamma}_1\dot{\gamma}_2}(R_3^{+})_{\dot{\gamma}_1\dot{\gamma}_2}{}^{\dot{\alpha}_1\dot{\alpha}_2}
\end{aligned} \tag{3.31}$$

Note that there are no amplitudes associated with R^2, reflecting the fact that the Gauss-Bonnet term is a total derivative in four dimensions. Higher dimensional R^2 upon dimensional reduction will reduce to ϕR^2 in four-dimensions, and generate the amplitude for a dilaton coupled to two gravitons:

$$\begin{aligned} M_3(1^0 2^{+2} 3^{+2}) &\rightarrow 2[23]^4 = (R_1^+)_{\dot\alpha_1\dot\alpha_2}{}^{\dot\beta_1\dot\beta_2}(R_2^+)_{\dot\beta_1\dot\beta_2}{}^{\dot\alpha_1\dot\alpha_2}, \\ M_3(1^0 2^{-2} 3^{-2}) &\rightarrow 2\langle 23\rangle^4 = (R_1^-)_{\alpha_1\alpha_2}{}^{\beta_1\beta_2}(R_2^-)_{\beta_1\beta_2}{}^{\alpha_1\alpha_2}, \end{aligned} \tag{3.32}$$

and similar amplitudes for ϕF^2.

Extending to four-points we find that there are three possible helicity structures that admit polynomial representations. For spin-1 we have for the lowest mass-dimensions:

$$\begin{aligned} M_4(1^+2^+3^+4^+) &\rightarrow 4\left([12]^2[34]^2{+}[13]^2[24]^2{+}[14]^2[23]^2\right) \\ &= (F_1^+\cdot F_2^+)(F_3^+\cdot F_4^+) + (F_1^+\cdot F_3^+)(F_4^+\cdot F_2^+) + (F_1^+\cdot F_4^+)(F_2^+\cdot F_3^+) \\ M_4(1^+2^+3^-4^-) &\rightarrow 4[12]^2\langle 34\rangle^2 = (F_1^+\cdot F_2^+)(F_3^-\cdot F_4^-)\ , \end{aligned} \tag{3.33}$$

where $(F_i^+\cdot F_j^+) \equiv (F_i^+)_\alpha{}^\beta (F_j^+)_\beta{}^\alpha$ and similar definition for $(F_i^-\cdot F_j^-)$. We also have $M_4(1^-2^-3^-4^-)$ which is simply changing the square brackets of $M_4(1^+2^+3^+4^+)$ to angles. It is straightforward to translate this back to vector representations, for which the independent F^4 contractions are given by:

$$(F^2)^2 \equiv (F_{\mu\nu}F^{\mu\nu})^2, \quad (F^2)(F\tilde F) \equiv (F_{\mu\nu}F^{\mu\nu})(\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}), \quad (F\tilde F)^2\,. \tag{3.34}$$

The linear map between them are given as:

$$\begin{aligned} M_4(1^+2^+3^+4^+) =& 8\left((F^2)^2 - 4(F\tilde F)^2 + 2(F^2)(F\tilde F)\right) \\ M_4(1^+2^-3^+4^-) =& 8(F^2)^2 + 32(F\tilde F)^2 \\ M_4(1^-2^-3^-4^-) =& 8\left((F^2)^2 - 4(F\tilde F)^2 - 2(F^2)(F\tilde F)\right)\,. \end{aligned} \tag{3.35}$$

From the above we immediately see that the combination $(F^2)^2{+}\frac{1}{4}(F\widetilde{F})^2$, which is the square of the Maxwell stress-tensor, only generates the MHV helicity configuration. Similar identification applies to spin-2, where we also have three distinct tensor structures for R^4 mapping to the three helicity structures. For higher derivative operators such as $D^{2n}F^4$ or $D^{2n}R^4$, we simply have extra Mandelstam variables multiplying the spinor brackets. For example

$$\sigma_2\langle 12\rangle^4[34]^4 \rightarrow D^4R^4\,. \tag{3.36}$$

Thus the EFT amplitude for massless spinning particles can in general be written in such a way that the spinor brackets are prefactors:

$$\begin{aligned}
M_4^{\rm IR}(1^+2^+3^+4^+) =& \frac{[12][34]}{\langle 12\rangle\langle 34\rangle} \times \left(\sum_{k,q} a_{k,q}^{all\,+}\, s^{k-q}t^q\right)\\
M_4^{\rm IR}(1^+2^+3^+4^-) =& \frac{[12][23]\langle 24\rangle}{\langle 12\rangle\langle 23\rangle[24]} \times \left(\sum_{k,q} a_{k,q}^{single\,-}\, s^{k-q}t^q\right)\\
M_4^{\rm IR}(1^-2^-3^+4^+) =& \frac{\langle 12\rangle^2[34]^2}{stu} \times \left(\sum_{k,q} a_{k,q}^{\rm MHV}\, s^{k-q}t^q\right)\,.
\end{aligned} \tag{3.37}$$

where the spinor prefactors are written in such a way that all possible massless poles are contained and are invariant under the permutation of the same helicity legs. The superscript for the Taylor coefficients $a_{k,q}^{\cdots}$ label the helicity configuration.

Let's consider explicit examples. The low energy expansion for Type-I and II superstring in eq.(3.16) and eq.(3.18) gives prime examples of gauge and gravitational EFT amplitudes. However due to being supersymmetric, only MHV configurations are present. For a more general set up, lets consider the open bosonic string amplitude, which contains all three sectors:

$$M^{\rm Bos}(s,t) = \frac{\Gamma[-s]\Gamma[-t]}{\Gamma[1+u]} f_{\{I\}},\qquad \begin{aligned}
f_{++++} &= \frac{[12][34]}{\langle 12\rangle\langle 34\rangle} stu\left(1-\frac{1}{s+1}-\frac{1}{u+1}-\frac{1}{t+1}\right)\\
f_{+++-} &= stu\frac{[12][23]\langle 24\rangle}{\langle 12\rangle\langle 23\rangle[24]}\\
f_{++--} &= -[12]^2\langle 34\rangle^2\left(1-\frac{tu}{s+1}\right)\,.
\end{aligned} \tag{3.38}$$

The low energy EFT is then given as:

$$\begin{aligned}
M^{IR}(1^+2^+3^+4^+) &= 2u\frac{[12][23][34][41]}{st}+2[13]^2[24]^2-[12][23][34][41](\frac{\pi^2}{3}-2)+\cdots\\
M^{IR}(1^+2^+3^+4^-) &= [12]^2[23]^2\langle 24\rangle^2\left(-\frac{1}{st}+\frac{\pi^2}{6}-u\zeta_3+\frac{\pi^4}{360}(4s^2+st+4t^2)+\cdots\right)\\
M^{IR}(1^+2^+3^-4^-) &= [12]^2\langle 34\rangle^2\left(-\frac{1}{st}+\frac{u}{s}+\frac{\pi^2}{6}-u(1+\zeta_3)+\cdots\right)\,,
\end{aligned} \tag{3.39}$$

where we've rewritten the spinor brackets in a form that exposes the massless poles. It is instructive to identify local operators in each helicity sector. For the all plus helicity the leading term correspond to the gluon exchange between the Yang-Mills vertex and F^3, followed by two types of contractions for $(F^+)^4$. For the single minus sector, we have massless poles associated with the exchange of a vector between $(F^+)^3$ and a Yang-Mills vertex, while the leading four-point local operator correspond to $D^2(F^+)^3F^-$. For

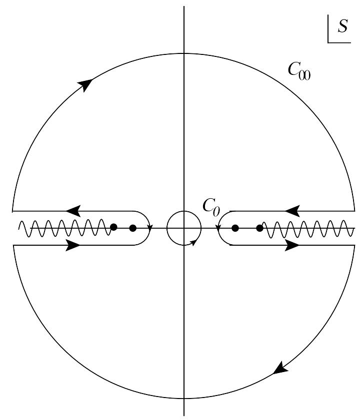


Figure 3.3: We define the low energy couplings through a contour integral on the complex s-plane, where the contour $\mathcal{C}_0$ encircles the origin. On the complex plane, if the amplitude only has singularities on the real-s axes on either poles or branch points, then we can deform to contour $\mathcal{C}_\infty$.

the MHV sector, we have two sets of massless poles, the leading corresponding to the exchange between the Yang-Mills vertex, while the subleading is between $(F^+)^3$ and $(F^-)^3$. The leading four-point local operator is $(F^+)^2(F^-)^2$.

3.3 Dispersive representation for EFT coefficients

In the previous section, we've seen that given the UV theory, the low energy EFT can be obtained by expanding the UV amplitude in Mandelstam variables, leading to an IR amplitude of the form

$$M(s,t)|_{s,t\ll m^2} = M^{\rm IR}(s,t) = \{\text{m}assless\ \ poles\} \quad + \sum_{k,q} a_{k,q}\, s^{k-q} t^q\,. \tag{3.40}$$

Mapping to on-shell local operators is then a straightforward task. However, it has been long appreciated that general principles of unitarity and Lorentz invariance imposes non-trivial constraint on the IR description. These constraints arise through the analyticity of the scattering amplitude, where the poles and branch cuts on the complex Mandelstam variable plane are associated with threshold productions. For the four-point amplitude, such analytic property allows us to equate the low energy couplings $a_{k,q}$ to the discontinuities of the branch cuts (or residues of poles), giving a dispersive representation for the couplings.

Let's begin by holding $t = t^* \ll m^2$ fixed, where m^2 is the characteristic mass associated with the UV completion, and consider four-point amplitude $M(s,t^*)$ as a function of s. We will imagine that we are only integrating out the massive states, which generate contact terms in the low-energy effective theory. Of course there will also be calculable massless loops in the low-energy effective theory, which induce logarithmic variation in these coefficients. We will return to discussing this point later in section 3.11. Note, however, that the very notion of "higher dimension operators" is only well-defined when there is a weak coupling in the UV theory, so that the contact operators induced by integrating out the massive states dominate over the ones generated by massless loops in the low-energy theory, so that this first-pass analysis captures the most interesting UV physics. In practice, we are assuming that, for small fixed $t \ll m^2$, the amplitude is analytic in the s plane for small s, and develops its first singularity (be it a pole at tree-level, or more generically a branch cut associated with UV particle production) at $s{=}m^2$.

It is important that when t is $\ll m^2$, the *only* singularities of the amplitude are on the real s axis, and correspond to particle production thresholds. This is not true when t is comparable to m^2, where new sorts of singularities, simplest amongst them the infamous "anomalous thresholds", with no Lorentzian particle production interpretation, also appear. But for our purposes of controlling EFT coefficients, we only need $t \ll m^2$ and never have to worry about anomalous thresholds. See appendix B.2 for a more detailed discussion of these issues.

As is standard from the study of dispersion relations, we consider the contour integral

$$\frac{i}{2\pi}\int_{\mathcal{C}_0}\frac{ds}{s^{n+1}}M(s,t^*)\,, \tag{3.41}$$

where $\mathcal{C}_0$ represents the contour that encircles the origin. Since at the origin both $s, t^* \ll m^2$ we know that amplitude takes its low energy form in eq.(3.40), and the residue for the measure $\frac{1}{s^{n+1}}$ will be given by terms in eq.(3.40) proportional to s^n. In the absence of t-channel massless pole, this residue will be a polynomial function of t, giving a well defined Taylor expansion around $t = 0$. Thus we find that $a_{k,q}$ can be identified as:

$$a_{k,q} = \frac{1}{q!}\left[\frac{\partial^q}{\partial t^q}\frac{i}{2\pi}\int_{\mathcal{C}_0}\frac{ds}{s^{k-q+1}}M(s,t)\right]\Bigg|_{t=0}\,. \tag{3.42}$$

In other words, the low energy couplings can be analytically defined through the on-shell amplitude. Note that taking the residue is equivalent to taking derivatives, and the result of this action is often referred to as the subtracted amplitude. Now instead of $\mathcal{C}_0$ we deform to the contour encircling infinity $\mathcal{C}_\infty$. If the non-analyticities are associated with particle production, they occur on the real axes where depending their origin as s or u-channel threshold, they will lie on the positive or negative real s-axes respectively. Thus the contour $\mathcal{C}_\infty$ takes the form shown in fig.3.3, where one picks up the discontinuity on the real axes as well as boundary contributions. At large s, if the amplitude falls of faster than s^{k-q} then the latter simply yields zero, and we would have an identity between $a_{k,q}$ and the residues or discontinuities.

Let us consider the linear sigma model as an explicit example. Once again the UV tree-amplitude is given as:

$$M(s,t) = -\frac{\lambda}{8m_h^2}\left(\frac{s^2}{s-m_h^2}+\frac{t^2}{t-m_h^2}+\frac{u^2}{u-m_h^2}\right). \tag{3.43}$$

As $s \to \infty$ the amplitude grows linearly in s, the contour deformation of eq.(3.42) will have no boundary contributions when $k-q \geq 2$. Focusing on the couplings with $q=0$, i.e. those that survive in the forward scattering limit $t=0$, we find eq.(3.42) implies:

$$a_{k,0} = -\frac{1}{(m_h^2)^{k+1}}\left(Res_s M(s,0)+(-)^k Res_u M(s,0)\right). \tag{3.44}$$

That is, the coupling $a_{k,0}$ is given by the residue of the Higgs pole in the s and u channel. Plugging in $Res_{s=m_h^2} M(s,0) = -\frac{\lambda m_h^2}{8}$ and $Res_{s=-m_h^2} M(s,0) = -\frac{\lambda m_h^2}{8}$, we have

$$a_{k,0} = \frac{\lambda}{4(m_h^2)^k}, \quad k \in even, \tag{3.45}$$

and 0 for $k \in$ *odd*. Indeed this reproduces the low energy couplings in eq.(3.13), for $k \geq 2$.

In general for theories whose four-point amplitude admits a convergent partial wave expansion, causality and unitarity dictate that the four-particle amplitude at $t=0$ is bounded by $s\log^{D-2} s$, i.e. the Froissart bound [82, 126]. When massless particles are present, such as in gravity, the t-channel singularity obstructs a convergent polynomial expansion in t and the Froissart analysis no longer holds. However, assuming a weakly coupled UV completion for gravity, causality consideration requires the presence of an infinite tower of massive higher spin states, leading to the forward amplitude behaving as s^p for $p<2$ at large s for fixed negative t [45]. From now on we will assume that for $|t| \ll m^2$ the amplitude is bounded by s^2 at large s. For a more detailed discussion, see Appendix B.1.

For general tree-level UV completions it is obvious that all poles lies on the real s-axes. More generally, the amplitude admits a dispersive representation

$$M(s,t)|_{t\ll m^2} = M^{\mathrm{Sub}} + \int_{M_s^2}^{\infty} dM^2\, \frac{\rho_s(M^2)}{s-M^2} + \int_{M_u^2}^{\infty} dM^2\, \frac{\rho_u(M^2)}{u-M^2} \tag{3.46}$$

where $M^{\mathrm{S}ub}$ represents the appropriate subtraction terms, representing the contributions from infinity in the dispersion relation. Note again the importance of keeping $t \ll m^2$

here. In general, we don't have good control on the analytic structure even of 4pt amplitudes in general theories. But we do have good control on the analytic structure of 2-pt functions as restricted by causality and unitarity. Intuitively, by keeping $t \ll m^2$, our 4-pt amplitude is close to forward scattering and hence a 2-pt function. A standard justification that the only singularities for $t \ll m^2$ are associated with usual particle production is given by studying Landau equations. In appendix B.2 we give a different, more direct derivation following directly from Feynman/Schwinger parametrization of loop integrals. Putting everything together, we conclude that for $k{-}q \geq 2$:

$$a_{k,q} = -\frac{1}{q!}\frac{\partial^q}{\partial t^q}\left(\sum_a \quad \frac{Res_{s=m_a^2}M(s,t)}{(m_a^2)^{k-q+1}} + \int_{4m_a^2} \frac{ds'}{s'^{k-q+1}} DisM(s,t)\right)\Bigg|_{t=0} +\{u\}\,, \quad (3.47)$$

where a labels all the massive states and $\{u\}$ represents the u-channel contributions.

Let us study the above identity with two explicit examples: the infinite resonance of a string theory tree level exchange and the one-loop massive bubble in three-dimensions.

Tree-level dispersive representation: Let's begin with the type-I string amplitude introduced in eq.(3.14), where the s-channel residue is given as:

$$Res_{s=n}\left[-\frac{\Gamma[-s]\Gamma[-t]}{\Gamma[1+u]}\right] = -\frac{(t+1)(t+2)\cdots(t{+}n{-}1)}{n!}. \quad (3.48)$$

Now using eq.(3.47) we have,

$$a_{k,q} = \frac{1}{q!}\frac{\partial^q}{\partial t^q}\left(\sum_{n=1}^{\infty}\frac{1}{n!}\ \frac{(t+1)(t+2)\cdots(t{+}n{-}1)}{n^{k-q+1}}\right). \quad (3.49)$$

First consider the coefficients relevant to the strict forward limit, $a_{k,0}$, which corresponds to setting $t = 0$ in the above, and we find:

$$a_{k,0} = \sum_{n=1}^{\infty}\frac{1}{n^{k+2}} = \zeta_{k+2}\,. \quad (3.50)$$

Indeed this is the reproduces the ζ_2 and ζ_3 for the constant and the coefficient for s in eq.(3.16) respectively. Now let's move away from the strict forward limit and consider coefficients of t to the first power. From eq.(3.49) we have,

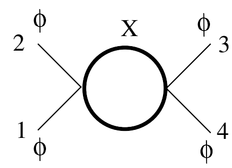
$$a_{k,1} = \sum_{n=2}^{\infty}\frac{1}{n^{k+1}}\left(1{+}\frac{1}{2}{+}\frac{1}{3}{+}\cdots{+}\frac{1}{n{-}1}\right), \quad . \quad (3.51)$$

Explicitly expanding eq.(3.14) to the fifth power in Mandelstam variables one finds,

$$a_{5,1} = -\frac{1}{90}(\pi^4\zeta_3{+}15\pi^2\zeta_5{-}270\zeta_7)\,, \quad (3.52)$$

which once again agrees with eq.(3.51).

Loop-level dispersive representation: Consider a three-dimensional theory with a massless scalar ϕ and a massive one X, interacting via the quartic coupling $\lambda\phi^2 X^2$. At low energies we have an effective action for ϕ, generated by integrating away the massive X loops. For example at leading order in λ, operators of the form $\partial^{2n}\phi^4$ are obtained by integrating out X from the one-loop bubble diagrams:



. (3.53)

This yields the following UV amplitude

$$\begin{aligned} M(s,t) &= \lambda^2\left[\mathcal{I}^3_{bubble}(s)+\mathcal{I}^3_{bubble}(t)+\mathcal{I}^3_{bubble}(u)\right]\,, \\ \mathcal{I}^3_{bubble}(s) &= \int\frac{d\ell^3}{(2\pi)^3}\frac{1}{(\ell^2-m^2)((\ell+p_{12})^2-m^2)} = \frac{1}{8\pi\sqrt{s}}\log\left(\frac{2m+\sqrt{s}}{2m-\sqrt{s}}\right)\,. \end{aligned} \tag{3.54}$$

The low energy expansion yields,

$$M^{\text{IR}}(s,t) = \frac{\lambda^2}{8\pi m}\left(3+\frac{\sigma_2}{80m^4}+\frac{\sigma_3}{448m^6}+\frac{\sigma_4}{2304m^8}\right)+\mathcal{O}\left(\frac{1}{m^{11}}\right)\,. \tag{3.55}$$

Now since the UV amplitude eq.(3.54) behaves as $\sim s^0$ as $s\to\infty$, we expect that through eq.(3.47) we can recover all low energy coefficients in eq.(3.55) with degree 1 and higher in s from the discontinuity of the bubble integrals. For fixed t, only the s- and u-channel bubble integrals contain branch cuts. The $\mathcal{I}^3_{bubble}(s)$ has a branch cut starting from $4m^2$ to ∞, with the discontinuity given by $\frac{i}{4\sqrt{s}}$, while the branch cut for $\mathcal{I}^3_{bubble}(u)$ is on the negative real s-axes from $-4m^2-t$ to $-\infty$, with discontinuity $\frac{i}{4\sqrt{-t-4m^2}}$. Thus from eq.(3.47), we find

$$a_{n+q,q} = \frac{1}{q!}\frac{\partial^q}{\partial t^q}\left[\frac{1}{2\pi i}\left(\int_{4m^2}^{\infty}\frac{1}{s^{n+1}}\frac{i}{4\sqrt{s}}+\int_{-t-4m^2}^{-\infty}\frac{1}{s^{n+1}}\frac{i}{4\sqrt{-t-4m^2}}\right)\right]\Bigg|_{t=0}\,. \tag{3.56}$$

For example to reproduce the coefficients of s^2t^q, we take $n=2$ in the square bracket above, yielding:

$$\begin{aligned} &\frac{1}{640m^5\pi}+\frac{1}{64\pi t^{5/2}}\left(3\pi-6\tan^{-1}\left(\frac{2m}{\sqrt{t}}\right)-\frac{4m\sqrt{t}(12m^2+5t)}{(4m^2+t)^2}\right) \\ =\;&\frac{1}{320m^5\pi}-\frac{3t}{3584m^7\pi}+\frac{t^2}{3072m^9\pi}+\mathcal{O}(t^3)\,. \end{aligned} \tag{3.57}$$

Indeed the first three terms in the t expansion match with the coefficients of s^2, s^2t, and s^2t^2 in eq.(3.55) respectively.

Before closing this section, we comment on two potential obstructions in utilizing the dispersive representation:

- The the residue at $s = 0$ contains t-channel singularity.
- The presence of massless cuts, which leads to branch point singularity at the origin.

A $1/t$ pole in the residue at $s = 0$ renders the Taylor expansion in eq.(3.42) ill defined. More precisely since by Cauchy theorem the t-channel pole must be reproduced by the sum over residues and branch cuts, the singularity in the $t \to 0$ limit indicates that the sum is not convergent. The graviton pole mentioned previously is a famous example of such obstruction. We will discuss this in great detail in the following section.

At loop-level there are two forms of non-analyticity at the origin for massless theories, IR singularities and massless cuts. For those with massless three-point interactions, such as gravity, loop-corrections are accompanied by collinear divergences. However, if we assume that the UV completion occurs while the self-coupling of the massless states are still perturbative, these divergences can be suppressed or computed order by order. The presence of massless cuts imply that one can no longer define the EFT couplings via the contour at $\mathcal{C}_0$. As previously mentioned this is reflecting the subtlety in what we mean by EFT couplings when log runnings are present. As we will see in sec. 3.11, the choice of "scale" against which the couplings run, are naturally introduced by moving the contour off the origin. After introducing such "generalized coupling" the remaining analysis are almost identical of the tree amplitude.

3.4 Obstructions from the massless poles

The presence of massless poles in the four-point amplitude, can potentially forbid a near forward limit dispersion representation. Take for an example an IR amplitude that behaves as

$$M^{\text{IR}}(s,t)|_{s,t\to 0} \sim \frac{s^n}{t} + a_{n,0}s^n + \mathcal{O}(t)\,. \tag{3.58}$$

Applying the dispersive representation for $a_{n,0}$ in eq.(3.47), we find:

$$\frac{1}{t} + a_{n,0} + \sum_a \left.\frac{Res_{s=m_a^2}\, M(s,t)}{(m_a^2)^{n+1}}\right|_{t=0} = 0\,. \tag{3.59}$$

Now since the above equality holds in the limit where $t \to 0$, the divergent behaviour of the $\frac{1}{t}$ pole tells us that the remaining summation cannot be convergent. For a concrete example, let's consider the four gluon amplitude in type-I super string. Stripping off the spinor factors, the following contour integral yields,

$$\frac{i}{2\pi}\int_{\mathcal{C}_0} ds \frac{M^{\mathrm{Type}I}(1^+2^+3^-4^-)}{[12]^2\langle 34\rangle^2} = \frac{1}{t}\,. \tag{3.60}$$

This isolates the field theory contribution $\frac{1}{st}$ in the low energy amplitude. Now at large s and small t, the amplitude scales as

$$\left.\frac{M^{\mathrm{Type}I}(1^+2^+3^-4^-)}{[12]^2\langle 34\rangle^2}\right|_{s\to\infty} < s^{-1}\,. \tag{3.61}$$

Thus if we deform the contour to $\mathcal{C}_\infty$, there are no boundary contributions and one only picks up poles on the real axes, whose residue is given by eq.(3.48). Thus we have

$$\frac{1}{t} - \sum_{n=1}^{\infty}\frac{(t{+}1)(t{+}2)\cdots(t{+}n{-}1)}{n!} = 0\,, \tag{3.62}$$

and setting $t = 0$ we indeed find that the summation is non convergent, $\sum_{n=1}^{\infty}\frac{1}{n} = \infty$!

In this paper, we will focus on $a, b \to a, b$ scattering where a, b may or may not be of the same type. When embedded in a gravitational theory one inevitably encounters the t-channel graviton exchange. For example consider the four-dilaton amplitude of type-II string theory

$$M^{\mathrm{Type}-II}(1^0 2^0 3^0 4^0) = g_s^2(st+tu+su)^2\frac{\Gamma[-s]\Gamma[-u]\Gamma[-t]}{\Gamma[1{+}s]\Gamma[1{+}u]\Gamma[1{+}t]}\,. \tag{3.63}$$

At low energies, beyond the tree-level graviton exchange the leading local amplitude is associated with $D^8\phi^4$,

$$M^{\mathrm{IR}}(1^0 2^0 3^0 4^0) = G_N\left(-\frac{st}{u}-\frac{tu}{s}-\frac{su}{t}+2\zeta_3(st+tu+su)^2+\cdots\right)\,. \tag{3.64}$$

Note that there are no four derivative couplings $D^4\phi^4$, which appears to violate the positivity bound $a_{2,0} > 0$ introduced long ago [4]. The resolution precisely lies in the presence of the t graviton pole! Let us see how this plays out in detail. First, as the amplitude enjoy $s \leftrightarrow u$ symmetry, we manifest this symmetry by switching to

$$z = s{+}\frac{t}{2}\,, \tag{3.65}$$

whereby $s \leftrightarrow u$ translates to $z \leftrightarrow -z$. We take the contour integral in z-plane, and defining the low energy coupling via its degree in z, t. Now let's compare the dispersive representation for the coupling of the four- and eight-derivative couplings, $a_{4,0}$ $a_{2,0}$. The integrals of interest are then:

$$\begin{aligned}
\frac{i}{2\pi}\int_{\mathcal{C}_0}\frac{dz}{z^3}\,M^{\text{Type}-II}(1^0 2^0 3^0 4^0) &= -\frac{1}{t}+\sum_{q=0}^{\infty} a_{q+2,q}t^q,\\
\frac{i}{2\pi}\int_{\mathcal{C}_0}\frac{dz}{z^5}\,M^{\text{Type}-II}(1^0 2^0 3^0 4^0) &= \sum_{q=0}^{\infty} a_{q+4,q}t^q.
\end{aligned} \tag{3.66}$$

Note that the contour $\mathcal{C}_0$ picked up residues at $z = 0, \pm t/2$, since $t \to 0$. Comparing the two integrals we see that the dispersive representation should be convergent for $a_{q+4,q}$ (including $a_{4,0}$), but not for $a_{q+2,q}$ (including $a_{2,0}$). As the representation is not convergent for $a_{2,0}$, positivity based on such dispersive arguments is no longer applicable.

However, the presence of massless t-poles in the field theory amplitude *does not* necessarily imply an obstruction. Consider a gravitational EFT whose low energy limit is given by the Einstein-Hilbert action and no modification to the graviton cubic couplings (i.e. no R^3). The low energy amplitude for $M(1^{+2}2^{+2}3^{-2}4^{-2})$ is given by

$$M^{\text{IR}}(1^{+2}2^{+2}3^{-2}4^{-2}) = [12]^4\langle 34\rangle^4\left(\frac{1}{stu}+\sum_{k,q} a_{k,q}s^{k-q}t^q\right). \tag{3.67}$$

Even though the low energy amplitude contains massless t poles, the $\mathcal{C}_0$ contour actually picks up multiple $1/t$ that cancels

$$\int_{\mathcal{C}_0}\frac{ds}{s^n}\frac{M^{\text{IR}}(1^{+2}2^{+2}3^{-2}4^{-2})}{[12]^4\langle 34\rangle^4} = -\frac{1}{t^{n+2}}+\frac{1}{t^{n+2}}+\sum_q a_{q+n-1,q}t^q\,. \tag{3.68}$$

This can be tied to the massless poles coming in the combination $\frac{1}{stu}$. This result is deeply tied to the fact that the amplitude for minimally coupled self-interacting massless particles are "3-particle constructible", i.e. consistent factorization in one channel automatically enforces consistency in all other channels.

Thus in summary, while graviton exchanges can introduce t-channel singularity, *if the four-point amplitude is 3-particle constructible, then the combined contributions cancel each other and we are free of t-channel obstruction.* Examples include four-graviton amplitude of pure Einstein-Hilbert gravity, as well as the gravitational Compton amplitude for minimally coupled particles. If we have extra symmetry which relates the amplitude to a 3-particle constructible partner, or that it suppresses the t-channel exchange, one can similarly avoid the t-channel obstruction. Let us go through explicit examples for spin-0, 1, and 2 amplitudes with graviton exchange.

Scalars We have discussed identical scalars in eq.(3.66). For distinct scalars, we can arrange the scalars such that there are no t-channel exchanges. For example, a pair of complex scalars with U(1) symmetry, the graviton exchange is given by:

$$M^{IR}(\phi_1\overline{\phi}_2\overline{\phi}_3\phi_4) = \frac{tu}{s} + \frac{st}{u}\,, \tag{3.69}$$

where there would be no t-channel poles and free from obstructions.

Photons The graviton poles and its residues are dictated by their minimal coupling, $F^2\phi$ and RF^2 operators. Let's start by choosing the same helicity to be in the t-channel, where one has:

$$M^{\text{IR}}(1^-2^+3^+4^-) = [23]^2\langle 14\rangle^2\left(\frac{1}{s}+\frac{1}{u}+\alpha_1\frac{1}{t}+\alpha_2\frac{su}{t}+\cdots\right)\,, \tag{3.70}$$

where α_1 and α_2 represent contribution from ϕF^2 and RF^2 respectively. Note that due to the helicity arrangements, the contribution from the latter only appears in t-channel. Factoring out the universal helicity factor and taking the contour integral near the origin we find,

$$\int\frac{dz}{z^{n+1}}\left(\frac{4t}{4z^2-t^2}+\frac{\alpha_1}{t}+\alpha_2\left(\frac{t}{4}-\frac{z^2}{t}\right)\right) = \begin{cases} \frac{\alpha_1}{t}+\frac{\alpha_2 t}{4} & \text{for } n=0 \\ -\frac{\alpha_2}{t} & \text{for } n=2 \end{cases} \tag{3.71}$$

while the integral vanishes for other n. Thus we see that minimal coupling does not introduce t-channel poles, while the presence of ϕF^2 and RF^2 leads to t-channel obstruction for the four and eight derivative terms respectively. Following our scalar example, let's arrange the helicity such that contributions from these higher-derivative operators only appear in the s-channel, as:

$$M^{\text{IR}}(1^-2^-3^+4^+) = [34]^2\langle 12\rangle^2\left(\frac{1}{t}+\frac{1}{u}+\alpha_1\frac{1}{s}+\alpha_2\frac{tu}{s}+\cdots\right)\,, \tag{3.72}$$

This time we find,

$$\int\frac{du}{u^{n+1}}\left(\frac{1}{t}+\frac{1}{u}-\alpha_1\frac{1}{u+t}-\alpha_2\frac{tu}{u+t}\right) = \frac{1}{t}-\alpha_2 t \quad \text{for } n=0 \tag{3.73}$$

and zero otherwise. Since we've factored out the spinor brackets, we see that t-channel singularities from minimal coupling obstruct the convergence of four derivative operators.

Let's consider the case where we wish to apply dispersive representation to the coefficient of F^4 operators, relevant for the analysis of weak gravity conjecture. After factoring out the spinor brackets, the coefficient of F^4 is mapped to $a_{0,0}$. For helicity $(1^-2^-3^+4^+)$ the spinor brackets are s^2 and thus we can bound $a_{0,0}$. However due to eq.(3.73) we see that $a_{0,0}$ suffers the t-pole obstruction. One might attempt to use the configuration $(1^-2^+3^+4^-)$, where there are no t-pole obstruction for the four-derivative term. However in this case the spinor prefactor is simply t^2 up to a phase, thus the coefficient for F^4 is mapped to the coefficient of s^0 for which the dispersive representation is not applicable due to boundary contributions.

Gravitons

For external gravitons, the analysis is parallel to the photon case except that the relevant couplings are now the Einstein-Hilbert term, ϕR^2 and R^3. For the MHV amplitude, with equal helicity in the s-channel we have

$$M^{IR}(1^{-2}2^{-2}3^{+2}4^{+2}) = [12]^4\langle 34\rangle^4\left(-\frac{1}{stu}+\alpha_1\frac{1}{s}+\alpha_2\frac{tu}{s}\right)\,, \tag{3.74}$$

where now α_1 and α_2 represents ϕR^2 and R^3 respectively. Since as previously discussed summing over the massless residues cancels for the Einstein-Hilbert term, there are no potential t-channel singularities. If we were to choose the other two channels, then from t-channel exchanges between ϕR^2 or R^3, we would have encounter the similar obstruction as the photon case for the eight and twelve derivative terms respectively.

The t-channel pole and Reggie behaviour In cases where the t-channel singularity implies non-convergence of the dispersive representation, it is instructive to see how the singularity is analytically reproduced. Let's reexamine the summation eq.(3.62) in the $t\to 0$ limit. In such case it can be approximated as

$$\sum_{n=1}^{\infty}\frac{(t{+}1)(t{+}2)\cdots(t{+}n{-}1)}{n!}\sim\sum_{n=1}^{\infty}\frac{1+t+\frac{t}{2}+\cdots\frac{t}{n-1}}{n}\sim\sum_{n=1}^{\infty}\frac{1+t\log n}{n}\sim\sum_{n=1}^{\infty}\frac{e^{t\log n}}{n}. \tag{3.75}$$

Finally, the last line simply becomes $\sum_{n=1}^{\infty}n^{t-1}$ which after approximating the sum as an integral yields $\frac{1}{t}$. Recall that the summation is over the residues of the amplitude at $s=n$, which is the dominant contribution for the amplitude as s nears threshold. The fact that at small t the residue is approximated by n^t implies that the amplitude behaves as s^t in the near forward limit. This is nothing but the linear Regge behaviour of string theory, except that it holds true for large but finite values of s. Of course this is not surprising given that in order for equation eq.(3.62) to hold, the amplitude is required to die off at $s\to\infty$, which is true precisely due to such Regge behaviour.

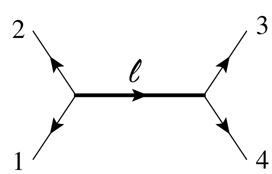
3.5 Theory space as a convex hull

As we have reviewed, there is a simple expression for the coefficients of low-energy effective field theory coefficients in terms of the spectrum and discontinuities of the high-energy amplitude:

$$a_{k,q} = -\frac{1}{q!}\frac{\partial^q}{\partial t^q}\left(\sum_a \frac{Res_{s=m_a^2}M(s,t)}{(m_a^2)^{k-q+1}} + \int_{4m_a^2}\frac{ds'}{s'^{k-q+1}}DisM(s,t)\right)\Bigg|_{t=0} + \{u\}\,. \quad (3.76)$$

Since optical theorem tells us that the sum of residue and discontinuity of the forward amplitude is proportional to the total cross-section $\sigma(s)$, Im $M(s,0) = -s\sigma(s)$, one immediately concludes that $a_{k,0} > 0$.

However, this is not the whole story since the optical theorem is really a "coarse grained" description of the residues and discontinuity. Lorentz invariance and factorization tells us vastly more than just the positivity in the forward limit. In particular when combined with unitarity, Lorentz invariance tells us that the discontinuities are positively expandable on a preferred polynomial basis! To see this, consider the $2 \to 2$ scattering of scalar particles $M(1^a, 2^b, 3^b, 4^a)$, where a, b labels the distinct species. Let's consider the general form of the residue from a tree-level spin-ℓ exchange:



(3.77)

The residue is given by the product of three-point amplitudes for two scalars a, b coupled to the spin-ℓ state. The amplitude is fixed by Lorentz invariance to be:

$$M_3(1^a, 2^b, \epsilon_I) = ic_\ell(p_1 - p_2)^{\mu_1}\cdots(p_1 - p_2)^{\mu_\ell}\epsilon_{I\mu_1\cdots\mu_\ell}\,, \quad (3.78)$$

where c_ℓ is the coupling constant, $\epsilon_{I\mu_1\cdots\mu_\ell}$ is the polarization tensor, and I labels the components of the spin-ℓ representation of the SO(D-1) massive Little group. The residue is then:

$$\sum_I M_3(1^a, 2^b, \epsilon_I)M_3(3^b, 4^a, \epsilon_I)\,. \quad (3.79)$$

Denoting $(p_1{-}p_2)$ and $(p_3{-}p_4)$ as (X, Y), in the center of mass (c.o.m) frame these are $(D{-}1)$-dimensional vectors. The sum over the I converts the product of polarization tensors into a polynomial of $\eta_{\mu\nu}$s, which is symmetric and traceless in the Lorentz indices on both sides of the factorization pole. This suggests that eq.(3.79) is simply a polynomial function of $(X^2, Y^2, X\cdot Y)$ that is of degree ℓ in X and Y respectively, and vanishes under the Laplacian ∇_X^2 and ∇_Y^2. The last constraint is a reflection of the traceless condition. In other words, one can read off the polynomial from the $D{-}1$ dimension solution to the Laplace equation:

$$\frac{1}{(X^2 - 2X\cdot Y + Y^2)^{\frac{D-3}{2}}}\,. \quad (3.80)$$

Without loss of generality, we can scale $|X| = 1$, $|Y| = r$, and $X \cdot Y = r\cos\theta$, where θ is the scattering angle. Then the polynomial can be identified through

$$\frac{1}{(1-2r\cos\theta+r^2)^{\frac{D-3}{2}}} = \sum_{\ell=0}^{\infty} r^\ell G_\ell^{(\mathrm{D})}(\cos\theta)\,. \tag{3.81}$$

which is the generating function for the Gegenbauer polynomials. For $D = 4$ this reduces to Legendre polynomial, while the three-dimensional counter part is the Chebyshev polynomials. From now on we will suppress the superscript (D) unless needed.

We've seen that the residue is simply a sum of Gegenbauer polynomials. Now due to our specific choice of external states, $M(1^a, 2^b, 3^b, 4^a)$, the three-point couplings on both sides of the (u) s-channel exchange are identical, i.e. the coupling constants squared c_ℓ^2. Thus we see that the residue is a function that is *positively expandable* on the Gegenbauer basis:

$$\boxed{Res_{s=m^2}M(s,t) = -\sum_\ell \mathsf{p}_\ell\, G_\ell(\cos\theta), \quad \mathsf{p}_\ell \geq 0}, \tag{3.82}$$

where $\cos\theta = 1+\frac{2t}{m^2}$. Functions that have such property are referred to as positive functions, and they enjoy the feature that such positivity is preserved under multiplication and differentiation. Note that since Gegenbauer polynomials are positive when $\theta = 0$, the optical theorem is simply a corollary of eq.(3.82). Gegenbauer polynomials are a particular example of orthogonal polynomials that are orthogonal to each other under prescribed integration measure. Gegenbauer polynomials are orthogonal with respect to SO$(D-1)$ invariant measure $(\sin\theta)^{D-4}d\cos\theta$. Since SO$(D-1)$ symmetry is simply a reflection of our kinematic setup, it is applicable for discontinuities as well. Indeed as we will demonstrate in appendix B.3, when combined with unitarity, the discontinuity in the near forward limit is again given by a positive sum of Gegenbauer polynomials:

$$\boxed{Dis_{s\geq 4m^2}M(s,t) = -\sum_\ell \mathsf{p}_\ell(s)\, G_\ell(\cos\theta)\,, \quad \mathsf{p}_\ell(s) \geq 0\,.} \tag{3.83}$$

Here, $\mathsf{p}_\ell(s)$ is the positive "spinning" spectral function. Note that at weak couplings, $\mathsf{p}_\ell > 0$ is all we can say. The full non-linear constraint implied by unitarity, $Im[\mathbf{a}_\ell(s)] \geq |\mathbf{a}_\ell(s)|^2$ where $\mathbf{a}_\ell$s are the partial wave coefficients, is only relevant for theories where the amplitudes becomes genuinely large/the theory is genuinely strongly coupled in the UV.

While the discussion so far is applicable the scattering amplitude of scalars, and hence scalar EFT, one can easily generalize when ever the three-point couplings of two massless one massive state are kinematically unique. This is the case in four-dimensions with external helicity states [10], where the corresponding orthogonal polynomials are Jacobi polynomials. We will review and discuss its property in great detail in sec.3.6.

Now that we see the residue/discontinuity of the four-point amplitude is given by a special class of functions, positive functions, we would like to extract the image of this property on the space of low energy couplings. Naturally this can be done through eq.(3.47). In other words, we would like to explore the full implication of:

$$a_{k,q} = \frac{1}{q!}\frac{d^q}{dt^q}\left(\sum_a \frac{\mathsf{p}_a G_{\ell_a}(1+2\frac{t}{m_a^2})}{(m_a^2)^{k-q+1}} + \sum_b \int ds' \mathsf{p}_{b,\ell}(s') \frac{G_\ell(1+2\frac{t}{s'})}{(s')^{k-q+1}} + \{u\}\right)\Bigg|_{t=0}, \quad (3.84)$$

where the equality is understood to hold as a Taylor series in t. i.e. $|t| \ll m^2$. More precisely, *coefficients of the higher dimensional operators as an expansion away from the forward limit must be given as a positive sum of the Taylor expansion of Gegenbauer polynomials.* Note that since the difference between contributions from residues and discontinuities is simply whether the spectrum of mass is discrete or continuous, by not assuming discreteness we will cover both. In this context, the previous forward limit positivity constraint at $q = 0$ is really the "tip" of the iceberg. It is coarse grained because it did not fully exploit the fact that the residue and discontinuity are positive functions.

Collecting the low energy couplings, eq.(3.84) is equivalent to:

$$\sum_{k,q} a_{k,q} s^{k-q} t^q = -\sum_a \mathsf{p}_a G_{\ell_a}\left(1+\frac{2t}{m_a^2}\right)\left(\frac{1}{s-m_a^2} - \frac{1}{s+t+m_a^2}\right), \quad (3.85)$$

where again the equality is understood in the sense of Taylor expansion in t, s. In other words, the near forward limit low energy expansion is captured by the s and u-channel factorizations alone. Now eq.(3.85) gives us a relation between $a_{k,q}$ and the Taylor coefficients of the Gegenbauer polynomials expanded around 1,

$$G_\ell(1+2\delta) = \sum_{q=0} v_{\ell,q}\, \delta^q\,. \quad (3.86)$$

If we only have s-channel contribution, eq.(3.85) implies:

$$s\ channel: \quad a_{k,q} = \sum_a \mathsf{p}_a \frac{v_{\ell_a,q}}{(m_a^2)^{k+1}} \quad \mathsf{p}_a \geq 0. \quad (3.87)$$

If u-channel contributions are present, we redefine the coupling in terms of expanding in (t, z), i.e. $a_{k,q} z^{k-q} t^q$, we find eq.(3.85) can instead be rewritten as:

$$s{-}u\ channel: \quad a_{k,q} = \sum_a \mathsf{p}_a \frac{u_{\ell_a,k,q}}{(m_a^2)^{k+1}} \quad \mathsf{p}_a \geq 0. \quad (3.88)$$

where $u_{\ell,k,q}$ is a linear combination of $v_{\ell,q}$ with its explicit form given in eq.(3.194). For $q = 0$, $u_{\ell,k,0} > 0$ and we are back to the old forward limit positivity constraint. For $q \neq 0$, $u_{\ell,k,q}$ can have either sign and we no longer have strict positive bounds for individual $a_{k,q}$, and naively there is no constraint. However, while there may no longer be constraint for individual $a_{k,q}$ with $q \neq 0$, there are non-trivial constraints as a collective. For example collecting the coefficients with fixed k but distinct q into a vector $\mathbf{a}_k$, we find

$$\mathbf{a}_k \equiv \begin{pmatrix} a_{k,0} \\ a_{k,1} \\ a_{k,2} \\ \vdots \end{pmatrix}, \quad \vec{u}_{\ell,k} \equiv \begin{pmatrix} u_{\ell,k,0} \\ u_{\ell,k,1} \\ u_{\ell,k,2} \\ \vdots \end{pmatrix} \quad \Rightarrow \quad \mathbf{a}_k = \sum_a \mathsf{p}_a \vec{u}_{\ell_a,k} \quad \mathsf{p}_a \geq 0\,, \tag{3.89}$$

where we absorbed the positive factors $(m_a^2)^{k+1}$ into p_a. In other words, $\mathbf{a}_k$ must be in the *convex hull* of the vectors $\vec{u}_{\ell,k}$! That is the boundary of "theory space", the space of allowed $\mathbf{a}_k$, is given by the boundaries of the hull.

Let us "see" explicitly examples of what this space looks like. For simplicity consider color ordered EFT amplitude whose UV completion does not include u-channel contributions. Taking $k = 1$ we find that eq.3.88 tells us:

$$\mathbf{a}_2 = \begin{pmatrix} a_{1,0} \\ a_{1,1} \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} v_{\ell_a,0} \\ v_{\ell_a,1} \end{pmatrix}\,. \tag{3.90}$$

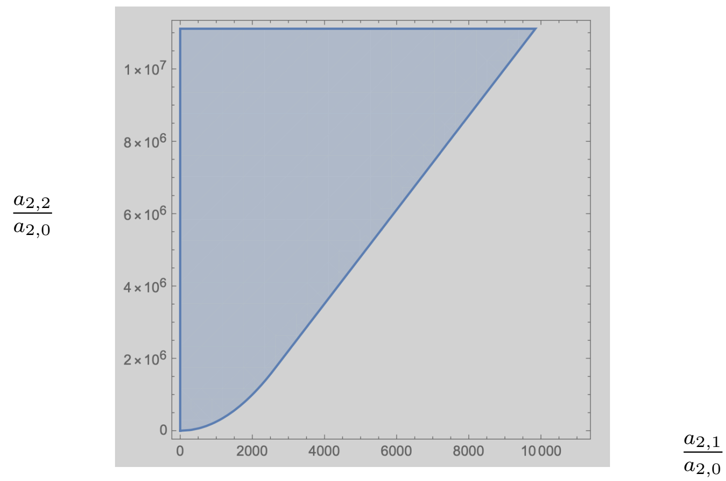
Since p_a is positive, the equality is projective in nature and we can rescale the top component of each vector to be 1. This then implies the following inequality

$$\frac{a_{2,1}}{a_{2,0}} \geq Min\left[\frac{v_{\ell,1}}{v_{\ell,0}}\right]. \tag{3.91}$$

Taking $D = 4$, we have $v_{\ell,0} = 1$ and $v_{\ell,1} = \ell(\ell{+}1)$, and we conclude that $\frac{a_{2,1}}{a_{2,0}} \geq 0$. For $k = 2$, the vector $\mathbf{a}_3$ lives in $\mathbb{P}^2$

$$\mathbf{a}_3 = \begin{pmatrix} a_{2,0} \\ a_{2,1} \\ a_{2,2} \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} v_{\ell_a,0} \\ v_{\ell_a,1} \\ v_{\ell_a,2} \end{pmatrix} \quad \to \quad \begin{pmatrix} a_{2,1}/a_{2,0} \\ a_{2,2}/a_{2,0} \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} v_{\ell_a,1} \\ v_{\ell_a,2} \end{pmatrix}, \tag{3.92}$$

where after the rescaling, besides $\mathsf{p}_a \geq 0$, we further have $\sum_a \mathsf{p}_a = 1$. Using $v_{\ell,2} = \frac{(1)_{\ell+2}}{4(\ell-2)!}$, the allowed region is now given as:



Once again, the positivity bound of [4] simply tells us that $a_{2,0} > 0$ and thus has no constraint for the above plot. As we extend to a higher degree in k, eq.(3.87) and eq.(3.88) become the statement that $a_{k,q}$ lives in the convex of vectors $\vec{v}_\ell$ and $\vec{u}_{\ell,k}$ for fixed k, and the relevant question is what are the boundaries of this hull.

In general the spin is unbounded especially when the UV completion involves massive loops, and thus the number of vectors that constitute the hull is infinite. Naively determining the boundaries of such space is computationally prohibitive. Note that these polytopal constraints, being for fixed k, bound operators of the same dimension. At the same time, we should expect non-trivial constraints that are cross dimensional since operators of different dimension are constrained by the same UV completion. As we will see these fascinating questions have a beautiful geometric answer to be explored in the remaining sections.

3.6 Hidden total positivity from unitarity and locality

In this section we briefly review the positive geometries relevant for our analysis. The spaces that we will be interested in are invariantly constructed as a *positive* sum of a fix set of vectors $\{\mathbf{V}_a\}$:

$$\mathbf{a} \in \sum_a \mathsf{p}_a \mathbf{V}_a, \quad \mathsf{p}_a > 0\,. \tag{3.93}$$

Such construction are referred to as convex hulls and the resulting geometry convex polytopes. Given a convex polytope, we will seek the complete set of inequalities that defines its interior. In other words we would like to "carve out" the subspace satisfying eq.(3.93) through equations of the form:

$$f_i(\mathbf{a}) > 0\,. \tag{3.94}$$

In the above i labels the distinct constraints. Depending on the nature of the vectors, we will find that f_i can be either linear or non-linear functions of $\mathbf{a}$. In the context of constraints for EFT, $\mathbf{a}$ is identified with the space of EFT couplings $\{a_{k,q}\}$ and the vectors $\mathbf{V}_a$ are determined by Lorentz invariance and locality, properties that we assume for the UV completion.

Convex hulls and Cyclic polytopes

Let us begin with the definition of convex hull. Given a set of d+1-dimensional vectors $\mathbf{V}_a$, consider the subspace spanned by its positive weighted sum:

$$\mathbf{a} \in \sum_a \mathsf{p}_a \mathbf{V}_a, \quad \mathsf{p}_a > 0\,. \tag{3.95}$$

The number of vectors will in general be greater than the dimension, and one must first determine whether this spans the whole space. For example, consider three vectors in two dimensions as in fig.(3.4). In the first case the three vectors span the whole space, as any point on the two-dimensional plane can be written as some positive sum of the three vectors. This is not the case for the second configuration since all vectors are on one side of the horizontal axis. Thus in order for the hull to be non-trivial, all the vectors must be on the same side of some hyperplane, or equivalently there are no non-trivial solutions to

$$\sum_a \mathsf{p}_a \mathbf{V}_a, = 0 \quad \mathsf{p}_a > 0\,, \tag{3.96}$$

i.e. the vectors do not enclose the origin.

Clearly for any $\mathbf{a}$ that satisfies eq.(3.95), so will $\rho\mathbf{a}$ with $\rho > 0$. Thus the solution space is naturally projective, and we identify $\mathbf{a} \sim \rho\mathbf{a}$ and $\mathbf{V}_a \sim \rho_a \mathbf{V}_a$. Since all the vectors lie on the same side of some hyperplane, we can choose our coordinates such that the top component is always positive, which we choose to normalize to 1:

$$\mathbf{V}_a = \left(\begin{array}{c} 1 \\ \vec{v}_a \end{array}\right), \quad \mathbf{X} = \left(\begin{array}{c} 1 \\ \vec{x} \end{array}\right)\,. \tag{3.97}$$

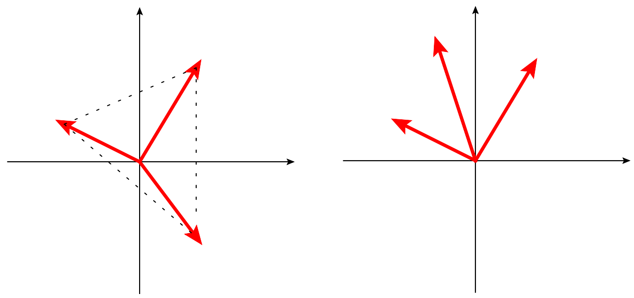


Figure 3.4: The convex hull of these three vectors encloses the origin, and hence trivially covers the entire two-dimensional plane.

In terms of $(\vec{v}_a, \vec{x})$ the canonical definition of convex hull is written as:

$$Conv[\vec{v}_a] = \left\{ \sum_a \mathsf{p}_a \, \vec{v}_a, \quad \middle| (\forall\, a : \mathsf{p}_a > 0) \wedge \sum_a \mathsf{p}_a = 1 \right\} . \tag{3.98}$$

As we will see it will be useful to retain the use of homogeneous coordinates, i.e. considering the vectors in their full $(d+1)$- component, and consider the hull as a projective polytope in $\mathbb{P}^d$:

$$Conv[\mathbf{V}_a] = \left\{ \sum_a \mathsf{p}_a \, \mathbf{V}_a, \quad \middle| (\forall\, a : \mathsf{p}_a > 0) \right\} . \tag{3.99}$$

The advantage of this is that it allows us to define various co-plane or incidence conditions projectively with the help of the $d+1$-dimensional Levi-Cevita tensor, $\epsilon_{I_1 I_2 \cdots I_{d+1}}$. For example, for the three vectors to be on a line in $\mathbb{P}^2$ we have

$$\langle a, b, c \rangle \equiv \epsilon_{I_1 I_2 I_3} V_a^{I_1} V_b^{I_2} V_c^{I_3} = 0 \,, \tag{3.100}$$

where $I_i = 1, 2, 3$. Similarly for $d+1$ vectors to lie on a $d-1$-dimensional plane in $\mathbb{P}^d$, tells us that the bracket $\langle a_1, a_2, \cdots, a_{d+1} \rangle = 0$. In this paper, the dimension of the angle brackets $\langle \cdots \rangle$ will be implicit from the number of entires or the surrounding discussions.

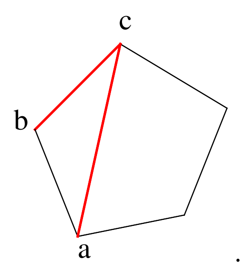
While eq.(3.99) gives us a d-dimensional polytope, not all vectors in $\mathbf{V}_a$ are vertices of the polytope—some might be *inside*. Thus given a convex hull, one needs to identify the vectors that constitute the vertices which ultimately define the polytope. The polytope can equivalently be defined through its boundaries, which are a set of co-dimension one hyper-planes or facets. The advantage of such facet point of view is that the polytope can be carved out successively one facet at a time. Not surprisingly, these facets can also be defined through the vertices of the polytope. More precisely, a co-dimension one plane is defined by a set of d distinct vectors, say $(\mathbf{V}_{a_1}, \mathbf{V}_{a_2}, \cdots, \mathbf{V}_{a_d})$. We can represent this plane as a $d{+}1$ component dual vector $\mathbf{W}_i$, where i labels the set of $\{a_i\}$ that defined the plane, and its components given by:

$$(W_i)_I \equiv \epsilon_{II_1I_2\cdots I_d} V_{a_1}^{I_1}, V_{a_2}^{I_2}, \cdots, V_{a_d}^{I_d} = \langle *, a_1, a_2, \cdots, a_d\rangle\,. \tag{3.101}$$

Then the inside of polytope is then given by the condition that $\mathbf{a}$ lies on one side of the facet $\mathbf{W}_i$. This constraint can be phrased in terms of a positivity condition:

$$\mathbf{W}_i \cdot \mathbf{a} = (W_i)_I a^I = \langle \mathbf{a}, a_1, a_2, \cdots, a_d\rangle > 0, \quad \forall \mathbf{a} \in Conv[\mathbf{V}_a]\,. \tag{3.102}$$

It is useful to see how such constraint arrises in simple setup. Consider a polygon in $\mathbb{P}^2$:



The line $\overline{bc}$ is a boundary since the interior of the polygon is on one side of the line. This is not the case for $\overline{ac}$. Not only does points of the interior lie on both sides, it can be *on* the line, i.e. collinear with (a, c). Since collinear means $\langle \mathbf{a}, a, c\rangle = 0$, this implies that $\langle \mathbf{a}, a, c\rangle$ is positive on one side of $\overline{ac}$, and negative one the other. Thus if $\mathbf{W}_i$ is a boundary, $\mathbf{W}_i \cdot \mathbf{a}$ must have the same sign for all $\mathbf{a}$, which we can always chose to be positive by appropriately arranging the sequence of vectors in $\{a_i\}$ eq.(3.101).

Given the complete set of $\{\mathbf{W}_i\}$, we now have a set of inequalities $f_i(\mathbf{a}) > 0$ that carves out the space. The function f_i in this case is linear in $\mathbf{a}$:

$$f_i(\mathbf{a}) = \mathbf{W}_i \cdot \mathbf{a} \geq 0\,. \tag{3.103}$$

The equal sign refers to points that are *on* the boundary. Now one can see that given a set of vectors $\mathbf{V}_a$, to determine the full set of $\{\mathbf{W}_i\}$, one would need to compute the sign of $\langle a_1, a_2, \cdots, a_{d+1}\rangle$ for all $d{+}1$-tuples. The sign patterns will tell us which vectors are vertices that form facets, and which ones are inside. For n vectors, this involves the computation of $\begin{pmatrix} n \\ d{+}1 \end{pmatrix}$ number of $d{+}1\times d{+}1$ determinants, which becomes intractable for large n. In the context of our EFT setup, n is associated with the number of Gegenbauer polynomials which is infinite. Thus the problem appears intractable, unless some reasonable truncation can be established. As we will now see, if the vectors satisfy special positivity conditions, the boundary and the vertices can be straightforwardly determined before hand. Remarkably, for us these properties are readily satisfied as a consequence of Lorentz invariance and locality of the UV completion!

Cyclic polytopes Let's start with a set of vectors $\mathbf{V}_a$ that are endowed with some preferred ordering. If all "ordered" $d{+}1 \times d{+}1$ determinants are positive:

$$\langle a_1, a_2, \cdots, a_{d+1}\rangle > 0, \quad \forall a_1 > a_2 > \cdots > a_{d+1}\,, \tag{3.104}$$

then the convex hull $Conv[\mathbf{V}_a]$ yields a *cyclic polytope.* The canonical example for a cyclic polytope is the convex hull of points on a moment curve. A moment curve is the embedding of the real line in d-dimensional space, such that each point on the line maps to a d-component vector with successive "moments", i.e. $(z, z^2, \cdots, z^d)$, with $z \in \mathbb{R}$. The convex hull of points on a moment curve is then a positive weighted sum of vectors taking the form:

$$\mathbf{V}_a = \begin{pmatrix} 1 \\ z_a \\ z_a^2 \\ \vdots \\ z_a^d \end{pmatrix}. \tag{3.105}$$

Naturally, $\mathbf{V}_a$ can be ordered by the value of z_a. In such case $\langle a_1, a_2, \cdots, a_{d+1}\rangle$ is simply the determinant of the Vandermonde matrix:

$$\mathrm{D}et \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_{d+1} \\ (z_1)^2 & (z_2)^2 & \cdots & (z_{d+1})^2 \\ \vdots & \vdots & \vdots & \vdots \\ (z_1)^d & (z_2)^d & \cdots & (z_{d+1})^d \end{bmatrix} = \prod_{i<j}(z_j - z_i)\,. \tag{3.106}$$

Indeed this determinant is positive for *ordered* points, $z_1 < z_2 < \cdots < z_{d+1}$.

Given eq.(3.104) one can straightforwardly see that the boundaries for a cyclic polytope in $\mathbb{P}^d$ are simply given as:

$$\begin{aligned} d \,\in\, even \qquad &\to \mathbb{P}^2: \;\; \langle *, i, i{+}1\rangle, \quad \mathbb{P}^4: \;\; \langle *, i, i{+}1, j, j{+}1\rangle, \\ d \,\in\, odd \qquad &\to \mathbb{P}^3: \;\; \langle 0, *, i, i{+}1\rangle, \;\; \langle *, i, i{+}1, \infty\rangle, \\ &\quad\;\; \mathbb{P}^5: \;\; \langle 0, *, i, i{+}1, j, j{+}1\rangle, \;\; \langle *, i, i{+}, j, j{+}1, \infty\rangle, \end{aligned} \tag{3.107}$$

where $i, i{+}1$ represents vectors that are adjacent in the ordering, and 0, ∞ is the first and final vector. To see that these are true boundaries, we must show for each of the walls in eq.(3.107), any point inside the hull $\mathbf{a} \in Conv[\mathbf{V}_a]$ will satisfy $\langle \mathbf{a}, \cdots \rangle \geq 0$ or $\langle 0, \mathbf{a}, \cdots \rangle \geq 0$. Let's take $\langle 0, *, i, i{+}1, j, j{+}1 \rangle$ as an example:

$$\langle 0, \mathbf{a}, i, i{+}1, j, j{+}1 \rangle = \sum_a \mathsf{p}_a \langle 0, a, i, i{+}1, j, j{+}1 \rangle \,, \tag{3.108}$$

since each bracket in the sum is an even permutation away from canonical ordering, they are positive due to eq.(3.104). As $\mathsf{p}_a > 0$ the RHS is a sum of positive terms and thus establishes $\langle 0, *, i, i{+}1, j, j{+}1 \rangle$ being a boundary of $Conv[\mathbf{V}_a]$. Note that a similar argument also tells us that there are no other boundaries.

Thus in summary, if the vectors $\mathbf{V}_a$ satisfy eq.(3.104), then the boundaries for $Conv[\mathbf{V}_a]$ is completely determined and constructed from consecutive pairs as illustrated in eq.(3.107). Furthermore since eq.(3.107) are boundaries for any $i, j, \cdots$, all vectors are vertices.

Hankel matrix total positivity

Let us consider a simple example where the positive geometry of cyclic polytopes arises in our EFT discussion. Take the following four point amplitude:

$$M(s) = \sum_a -\frac{\mathsf{p}_a}{s - m_a} \,. \tag{3.109}$$

This arrises naturally as the dispersive representation of the four-point amplitude in the forward limit. Note that the positivity of p_a is a reflection of unitarity and the simple pole in s is a reflection of locality. Thus the geometry that arrises from eq.(3.109) will have its origin in the union of unitarity and locality.

Expanding eq.(3.109) in small s we find

$$\sum_k a_k \, s^k \; = \sum_a \frac{\mathsf{p}_a}{m_a^2} \left(1 + \frac{s}{m_a^2} + \frac{s^2}{m_a^4} + \cdots \right) \,. \tag{3.110}$$

Matching both sides of the above equation we immediately see that the a_ks are positive. But there is more! If we collect the couplings into a vector $\vec{a}$, eq.(3.110) becomes:

$$\mathbf{a} = \begin{pmatrix} 1 \\ a_1/a_0 \\ a_2/a_0 \\ \vdots \\ a_k/a_0 \end{pmatrix} = \sum_a \mathsf{p}'_a \begin{pmatrix} 1 \\ x_a \\ \vdots \\ x_a^k \end{pmatrix} \,, \quad x_a \equiv \frac{1}{m_a^2} \,, \tag{3.111}$$

where we've used the projective nature of the problem to rescale the top component to be 1. We find that eq.(3.110) tells us that $\vec{a}$ lies in the convex hull of moment curves! Note that since $m_a^2 > 0$, we are really considering the "half" moment curve where $x_a \in \mathbb{R}^+$. Using what we've learned in the previous subsection, we have

$$\mathbf{W}_i \cdot \mathbf{a} \geq 0. \tag{3.112}$$

where $\mathbf{W}_i$ are the boundaries listed in eq.(3.107) with V_a determined by x_a and we have an infinite number of constraints on the couplings! However these constraints are not ideal as they rely on the explicit vectors V_a and for a low energy theorist we are not privy to the information of the UV spectrum, i.e. we do not know what the x_as are. It would be desirable to find constraints $f_i(\mathbf{a}) \geq 0$, such that the functions f_i do not depend on the explicit values x_a, while reflecting the fact that $x_a \in \mathbb{R}^+$.

Let's start by assuming the knowledge of the spectrum and see if we can rewrite $\mathbf{W}_i \cdot \mathbf{a} > 0$ in such a way that the information of the spectrum decouples. We can assume the spectrum to be continuous without lost of generality, since any of the p_as can be set to be arbitrarily matched with any specific spectrum. Beginning with $d = 1$, we have $\mathbf{a} = (1, \frac{a_1}{a_0})$ and there is only one boundary $\mathbf{W} = (1, 0)$. Thus we have:

$$\mathbf{W} \cdot \mathbf{a} = \langle 0, \mathbf{a} \rangle = \frac{a_1}{a_0} > 0, \tag{3.113}$$

which is trivial since we know that $a_0, a_1 > 0$. For $d = 2$, $\mathbf{a} = (1, \frac{a_1}{a_0}, \frac{a_2}{a_0})$ and the constraint is

$$\langle \mathbf{a}, a, a+1 \rangle > 0\,. \tag{3.114}$$

Since the spectrum is continuous, given a point x_a on the moment curve we can take $a+1$ to be arbitrarily close to a, such that $\langle *, a, a+1 \rangle \to \langle *, a, \dot{a} \rangle$, where $\dot{a}$ represents the derivative. The determinant then becomes

$$\langle \mathbf{X}, a, a+1 \rangle = \mathrm{D}et \begin{bmatrix} 1 & 1 & 0 \\ \frac{a_1}{a_0} & x_a & 1 \\ \frac{a_2}{a_0} & x_a^2 & 2x_a \end{bmatrix} = \frac{a_2 - 2a_1 x_a + a_0 x_a^2}{a_0}\,. \tag{3.115}$$

We see that the minimum occurs at $x_a = \frac{a_1}{a_0}$, and thus for eq.(3.114) to hold we must have:

$$a_0 a_2 - a_1^2 = \mathrm{D}et \begin{bmatrix} a_0 & a_1 \\ a_1 & a_2 \end{bmatrix} > 0\,. \tag{3.116}$$

Note that is non-linear in $\mathbf{a}$ and no longer depends on the point x_a! Moving on to $d = 3$, the analysis for $\langle 0, \mathbf{a}, a, a+1 \rangle$ is identical to that for the $d = 2$ case, leading to

$$a_1 a_3 - a_2^2 = \mathrm{D}et \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix} > 0\,. \tag{3.117}$$

Two comments are in order. First note that we have not considered constraints involving the infinity vertex. This is because projectively, the infinity vector is simply $(0,\cdots,0,1)$ and when plugged into $\langle\cdots,a,a{+}1,\infty\rangle$, it reduces to the constraint one dimension lower. Second, as we move from even to odd dimensions, we obtain the same constraint as before only with $a_i\to a_{i+1}$, for example eq.(3.116) and eq.(3.117). This can be understood as follows: the facets in both cases are comprised of the same set of vertices, just with the inclusion of the origin 0 for the odd case. In taking the determinant, 0 removes the first component of each vector, and the remaining part is proportional to the vector one dimension lower. Thus the condition in the odd dimension is simply an overall factor multiplying that of one dimension lower. Importantly, since we are on a half moment curve, the overall prefactor will be positive. For example:

$$\langle 0,\mathbf{a},a,a{+}1\rangle=\mathrm{D}et\begin{pmatrix} x_0 & 1 & 1 & 1\\ a_1 & 0 & x_a & x_{a+1}\\ a_2 & 0 & x_a^2 & x_{a+1}^2\\ a_3 & 0 & x_a^3 & x_{a+1}^3\end{pmatrix}=x_{a+1}x_a\mathrm{D}et\begin{pmatrix} a_1 & 1 & 1\\ a_2 & x_a & x_{a+1}\\ a_3 & x_a^2 & x_{a+1}^2\end{pmatrix}\,. \tag{3.118}$$

Since $x_a,x_{a+1}>0$, the fact that the very LHS is positive translate to the positivity on the very RHS, i.e. in eq.(3.117). Let's consider one more example before moving on to the general constraint. For $d=4$ we have

$$\begin{aligned}\langle\mathbf{a},a,a{+}1,b,b{+}1\rangle&=\mathrm{D}et\begin{bmatrix} 1 & 1 & 0 & 1 & 0\\ \frac{a_1}{a_0} & x_a & 1 & x_b & 1\\ \frac{a_2}{a_0} & x_a^2 & 2x_a & x_b^2 & 2x_b\\ \frac{a_3}{a_0} & x_a^3 & 3x_a^2 & x_b^3 & 3x_b^2\\ \frac{a_4}{a_0} & x_a^4 & 4x_a^3 & x_b^4 & 4x_b^3\end{bmatrix}\\ &=(x_a{-}x_b)^4(a_4{-}2\alpha a_3{+}a_2(\alpha^2{+}2\beta){+}\beta(a_0\beta{-}2a_1\alpha))\,,\end{aligned} \tag{3.119}$$

where $\alpha=(x_a{+}x_b)$ and $\beta=x_ax_b$. The minima in terms of α occurs at $\alpha=\frac{\beta a_1+a_3}{a_2}$. Plugging into the RHS of the above and requiring it to be positive leads to:

$$\mathrm{D}et\begin{bmatrix} a_0 & a_1 & a_2\\ a_1 & a_2 & a_3\\ a_2 & a_3 & a_4\end{bmatrix}>0\,. \tag{3.120}$$

We are now ready to give the result for general d. Collecting the coefficients of $\vec{a}$ into the symmetric Hankel matrix:

$$K(\vec{a}) = \begin{pmatrix} a_0 & a_1 & \cdots & a_{p-1} \\ a_1 & a_2 & \cdots & a_p \\ \vdots & \vdots & \vdots & \vdots \\ a_{p-1} & a_p & \cdots & a_{2p-2} \end{pmatrix}, \tag{3.121}$$

then the coefficients are in the convex hall of the half-moment curve if and only if the Hankel matrix is a totally positive matrix! A totally positive matrix has the property that all of its minors are non-negative. This is the well known solution to the Stieltjes moment problem. Note that due to K being a symmetric matrix, not all minors are independent. The independent constraints are the positivity of the principle minors of $K(\vec{a})$ and $K(\vec{a})_{i\to i+1}$. That is

$$\begin{aligned} i \in even: \quad & \text{Det}\begin{bmatrix} a_0 & a_1 & \cdots & a_{\frac{i}{2}} \\ a_1 & a_2 & \cdots & a_{\frac{i}{2}+1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{\frac{i}{2}} & a_{\frac{i}{2}+1} & \cdots & a_i \end{bmatrix} \geq 0, \\ i \in odd: \quad & \text{Det}\begin{bmatrix} a_1 & a_2 & \cdots & a_{\frac{i+1}{2}} \\ a_2 & a_3 & \cdots & a_{\frac{i+3}{2}} \\ \vdots & \vdots & \vdots & \vdots \\ a_{\frac{i+1}{2}} & a_{\frac{i+3}{2}} & \cdots & a_i \end{bmatrix} \geq 0. \end{aligned} \tag{3.122}$$

Its validity can be seen by the analytic representation of eq.(3.122):

$$\begin{aligned} i \in even: \quad & \sum_{\{b_1,b_2,\cdots,b_{\frac{i}{2}+1}\}} \left(\prod_{k=1}^{\frac{i}{2}+1} \mathsf{p}_{b_k}\right) \prod_{1\leq k<l\leq \frac{i}{2}+1} (x_{b_k}-x_{b_l})^2, \\ i \in odd: \quad & \sum_{\{b_1,b_2,\cdots,b_{\frac{i+1}{2}}\}} \left(\prod_{k=1}^{\frac{i+1}{2}} \mathsf{p}_{b_k} x_{b_k}\right) \prod_{1\leq k<l\leq \frac{i+1}{2}} (x_{b_k}-x_{b_l})^2 \end{aligned} \tag{3.123}$$

For $i \in even$ it is manifestly positive, and thus must hold for the convex hull of general moment curves. Indeed this was already noted in [149]. For $i \in odd$, its positivity then relies on $x_a > 0$, and thus only hold for the convex hull of *half* moment curves.

The Gegenbauer cyclic polytopes

We now turn to the positivity associated with the Gegenbauer polynomials. From the its definition from the generating function in eq.(3.81), it is straightforward to see that

$G_\ell^{(n)}(1) \equiv \frac{1}{n!}\partial_z^n G_\ell(z)|_{z=1} \geq 0$. However, just as the case with moments and Vandermonde determinants, further positive properties can be found when the components are organized into matrices. Let us consider the following Gegenbauer matrix

$$\mathrm{D}et\begin{bmatrix} G_{\ell_1}(z_1) & G_{\ell_2}(z_1) & \cdots & G_{\ell_n}(z_1) \\ G_{\ell_1}(z_2) & G_{\ell_2}(z_2) & \cdots & G_{\ell_n}(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ G_{\ell_1}(z_n) & G_{\ell_2}(z_n) & \cdots & G_{\ell_n}(z_n) \end{bmatrix}. \tag{3.124}$$

It turns out, the above matrix is totally positive if $1 \leq z_1 < z_2 < \cdots z_n$ and $\ell_1 < \ell_2 < \cdots < \ell_n$. For Chebychev polynomials, which are the Gegenbauer polynomials in $D = 3$, this can be straightforwardly proven, and we present the result in appendix B.5. For general D, the proof follows from that presented by Karlin and McGregor for general orthogonal polynomials [111]. In appendix B.5, we also give a direct computation of the relevant determinants for the Gegenbauer case of interest to us, allowing us to see the positivity explicitly

Such "position space" positivity, where the z_is are evaluated at separate points, is not convenient for our EFT analysis. In anticipating the Taylor expansion in eq.(3.84), we would like to instead extract conditions on the derivatives of the polynomials. This can be done by taking the positions to be close to some common point, say 1. Then the determinant of the Gegenbauer matrix becomes that for derivatives of Gegenbauer polynomial evaluated at $z_i = 1$. For example, defining

$$\mathbf{G}_\ell \equiv \begin{pmatrix} G_\ell^{(0)}(1) \\ G_\ell^{(1)}(1) \\ G_\ell^{(2)}(1) \\ \vdots \\ G_\ell^{(n)}(1) \end{pmatrix}, \tag{3.125}$$

the determinant of the Gegenbauer matrix with $1 \leq z_1 < z_2 < \cdots z_n < 1 + \epsilon$ becomes the determinant of the "Taylor" scheme matrix

$$(\mathbf{G}_{\ell_1}(1),\, \mathbf{G}_{\ell_2}(1),\, \cdots,\, \mathbf{G}_{\ell_{n+1}}(1))\,. \tag{3.126}$$

Thus the positivity of the Gegenbauer matrix in position space will imply the determinant of the above matrix is positive. Let's write out the explicit Taylor coefficients:

$$G_\ell(1+2\delta) = \sum_{q=0}^{\ell} v_{\ell,q}\delta^q\,, \quad v_{\ell,q} = \begin{cases} \frac{2^q}{q!(\ell-q)!}\frac{(\alpha)_{\ell+q}}{\prod_{a=1}^{q}(\alpha+2a-1)} & for\; q \leq \ell \\ 0 & for\; q > \ell \end{cases}, \tag{3.127}$$

where $\alpha = D{-}3$. Note that the coefficients are all positive, which reflects the fact that the derivative of $G_\ell(x)$ is again a positive function.[3] Using this one can show that the determinant of eq.(3.126) is (see appendix B.5):

$$\mathrm{Det}\left[\mathbf{G}_{\ell_1},\cdots,\mathbf{G}_{\ell_{n+1}}\right] = 2^{\frac{n(1+n)}{2}}\left(\prod_{i=1}^{n+1}\frac{(\alpha)_{\ell_i}}{\ell_i!}\frac{1}{\prod_{a=1}^{i-1}(\alpha{+}2a{-}1)a!}\right)\prod_{i<j}(\ell_j{-}\ell_i)(\alpha{+}\ell_j{+}\ell_i)\,, \tag{3.128}$$

which is manifestly positive for ordered spins, $\ell_1 < \ell_2 < \cdots < \ell_{d+1}$. This immediately tells us that

$$\boxed{\textit{the convex hull of the } \mathbf{G}_\ell \textit{ is a cyclic polytope!}} \tag{3.129}$$

Thus just as for the convex hull of points on the moment curve, the boundaries for $Conv[\mathbf{G}_\ell]$ are simply given by:

$$\begin{aligned}
d \in \textit{even} \to \mathbb{P}^2:&\quad \langle *,\ell_i,\ell_i{+}1\rangle,\quad \mathbb{P}^4:\ \langle *,\ell_i,\ell_i{+}1,\ell_j,\ell_j{+}1\rangle,\\
d \in \textit{odd} \to \mathbb{P}^3:&\quad \langle 0,*,\ell_i,\ell_i{+}1\rangle,\ \ \langle *,\ell_i,\ell_i{+}1,\infty\rangle,\\
\mathbb{P}^5:&\quad \langle 0,*,\ell_i,\ell_i{+}1,\ell_j,\ell_j{+}1\rangle,\ \ \langle *,\ell_i,\ell_i{+},\ell_j,\ell_j{+}1,\infty\rangle\,.
\end{aligned} \tag{3.130}$$

Going back to the position space Gegenbauer matrix, instead of setting all of the positions close to 1, lets have $z^* \leq z_1 < z_2 < \cdots < z_n < z^*+\delta$, with $1 < z^*$, the eq.(3.124) becomes

$$\mathrm{Det}\begin{bmatrix} G_{\ell_1}(z_1) & G_{\ell_2}(z_1) & \cdots & G_{\ell_n}(z_1)\\ G_{\ell_1}(z_2) & G_{\ell_2}(z_2) & \cdots & G_{\ell_n}(z_2)\\ \vdots & \vdots & \vdots & \vdots\\ G_{\ell_1}(z_n) & G_{\ell_2}(z_n) & \cdots & G_{\ell_n}(z_n)\end{bmatrix} = \mathrm{Det}\left[\mathbf{G}_{\ell_1}(z^*),\cdots,\mathbf{G}_{\ell_n}(z^*)\right] > 0\,. \tag{3.131}$$

Thus the convex hull of $\mathbf{G}_\ell(z^*)$ is in fact a cyclic polytope for all $z^* \geq 1$! Now consider a series of cyclic polytope,

$$Poly_i = Conv[\mathbf{G}_\ell(z_i)]\,. \tag{3.132}$$

defined with $1 \leq z_1 < z_2 < \cdots$. Since the derivative of $G_\ell(z)$ is a positive function, i.e.

$$\frac{dG_\ell(z)}{dz} = \sum_{\ell'} c_{\ell\ell'}G_{\ell'}(z) \quad c_{\ell\ell'} \geq 0 \tag{3.133}$$

[3] This can be deduced by taking the derivative on the generating function. Such extended positivity away from the forward limit was suggested long ago in [127], and utilized as consistency conditions for EFT in [137], deriving bounds in [23].

we can deduce

$$\mathbf{G}_\ell(z+\delta)=\begin{pmatrix}\mathbf{G}_\ell(z)+\delta\mathbf{G}'_\ell(z)\\ \mathbf{G}'_\ell(z)+\delta\mathbf{G}''_\ell(z)\\ \vdots\end{pmatrix}+\mathcal{O}(\delta^2)=\mathbf{G}_\ell(z)+\sum_{\ell'}c_{\ell\ell'}\mathbf{G}_{\ell'}(z)+\mathcal{O}(\delta^2)\,. \tag{3.134}$$

That is, a positively shifted $\mathbf{G}_\ell(z)$ can be positively re-expanded on $\mathbf{G}_\ell(z)$. Now starting with $z_1<z_2$, since we've concluded $\mathbf{G}_\ell(z_2)$ is positively expanded on $\mathbf{G}_\ell(z_1)$, its convex hull is *inside* the polytope Pol_1. Thus given a series of ordered points, $z_1<z_2<z_3$, the corresponding $Poly_i$ defined in eq.(3.132) satisfies:

$$Poly_3\subset Poly_2\subset Poly_1\quad for\ z_1<z_2<z_3. \tag{3.135}$$

In other words, as we push z away from 1, not only is the convex hull of $\mathbf{G}_\ell(z)$ a cyclic polytope, it goes deeper and deeper *inside* the original polytope!

Spinning Gegenbauer cyclic polytope Recall that the Gegenbauer polynomial being the unique polynomial for scalar amplitude with a spin-ℓ exchange is rooted in the three-point amplitude of two scalars and a spin-ℓ particle is unique. For general three-point amplitudes with spins this is no longer true. However as discussed in [10], in four-dimensions given the helicities of the two massless particles and the spin of the massive particle, the amplitude is fixed. This allows one to define a set of "spinning" Gegenbauer polynomial basis.

To see this, let's consider the three-point amplitude involving a massive spin-ℓ particle and massless particles with helicity h_1,h_2. We again have a polarization tensor $\epsilon_{\mu_1\mu_2\cdots\mu_\ell}$ needing ℓ vectors to contract. Due to $h_1,h_2\neq 0$, besides p_{12} we now have two new vectors, $q=\lambda_1\tilde\lambda_2$ and $\tilde q=q^*=\lambda_2\tilde\lambda_1$, that can be used to contract with the polarization tensor. Up to an overall constant, the amplitude is fixed by $\{h_1,h_2,\ell\}$ as:

$$\begin{aligned}&q^{\mu_1}q^{\mu_2}\cdots q^{\mu_{h_2-h_1}}(p_{12})^{\mu_{h_2-h_1+1}}\cdots(p_{12})^{\mu_\ell}\epsilon_{\mu_1\cdots\mu_\ell},\quad for\ h_2{-}h_1>0\\ &\tilde q^{\mu_1}\tilde q^{\mu_2}\cdots\tilde q^{\mu_{h_1-h_2}}(p_{12})^{\mu_{h_1-h_2+1}}\cdots(p_{12})^{\mu_\ell}\epsilon_{\mu_1\cdots\mu_S},\quad for\ h_1{-}h_2>0\,.\end{aligned} \tag{3.136}$$

We can now glue the two three-point amplitudes together to construct the residue for a spin-ℓ exchange. As discussed in [10], since the polarization tensors form irreps of the little group, the gluing of the three-point amplitude is simplified by first rewriting it in SL(2,C) irreps as:

$$[12]^{\ell+h_1+h_2}\left(\lambda_1^{\ell+h_2-h_1}\lambda_2^{\ell+h_1-h_2}\right)_{\{\alpha_1\cdots\alpha_{2\ell}\}}\,, \tag{3.137}$$

and then contract the SL(2,C) indices between both sides of the factorization channel. In the center of mass frame, we can parameterize the spinors as:

$$\lambda_1 = m^{\frac{1}{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \lambda_2 = m^{\frac{1}{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \lambda_3 = im^{\frac{1}{2}} \begin{pmatrix} \sin\frac{\theta}{2} \\ -\cos\frac{\theta}{2} \end{pmatrix}, \lambda_4 = im^{\frac{1}{2}} \begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} \end{pmatrix}. \tag{3.138}$$

We can identify the the three-point coupling in eq.(3.137) involving legs $1,2$ as a spin-ℓ state with "J_z" quantum number $m = h_1 - h_2$. Replacing $1,2$ with $3,4$ we then have a spin-ℓ state with quantum number $m = h_3 - h_4$, acted upon a rotation matrix in the "y"-axes by θ. The gluing of the three-point amplitude on both sides then simply corresponds to computing the overlap of the two states, which is nothing by the Wigner d-matrix! Thus we see that for general spinning particles the polynomial is simply:

$$d^{\ell}_{h_1-h_2,h_3-h_4}(\theta)\,. \tag{3.139}$$

where $d^j_{m',m}(\theta)$ is the Wigner d-matrix defined by $d^j_{m',m}(\theta) = \langle j, m'|e^{-i\theta\mathcal{J}_y}|j,m\rangle$.

Let us consider as an example the residue for a spin-ℓ exchange in the helicity configuration $(+h,-h,+h,-h)$. Writing it as a product of three point amplitudes, we find:

$$\begin{aligned} n_\ell^{\{+h,-h,+h,-h\}} &= A(1^{+h}2^{-h}\mathbf{P}^\ell)A(3^{+h}4^{-h}\mathbf{P}^\ell) \\ &= \frac{|c_\ell|^2([12][34])^\ell}{m^{4\ell-2}}\left(\lambda_1^{\ell-2h}\lambda_2^{\ell+2h}\right)^{\{\alpha_1\cdots\alpha_{2\ell}\}}\left(\lambda_3^{\ell-2h}\lambda_4^{\ell+2h}\right)_{\{\alpha_1\cdots\alpha_{2\ell}\}} \\ &= |c_\ell|^2[13]^{2h}\langle 24\rangle^{2h}\frac{([12][34])^\tau}{m^{4\ell-2-4h}}\left[(\lambda_1^\tau\lambda_2^\tau)\cdot(\lambda_3^\tau\lambda_4^\tau)\right], \\ &= |c_\ell|^2\, m^2 d^\ell_{2h,2h}(\theta)\,, \end{aligned} \tag{3.140}$$

where $\mathbf{P}^\ell$ indicates a spin-ℓ state with $P^2 = (p_1+p_2)^2 = m^2$, $\tau = \ell-2h$ and we've normalized the amplitudes such that the coupling constant c_ℓ is dimensionless. Note that the ℓ-independent prefactor $[13]^{2h}\langle 24\rangle^{2h}$ is required from helicity constraints, indicating that $d^\ell_{2h,2h}(\theta) \propto \cos^{4h}\frac{\theta}{2}$. Exchanging $3,4$ one obtains the residue for other helicity configurations:

$$n_\ell^{\{+h,-h,-h,+h\}} = \sum_\ell |c_\ell|^2\, m^2(-1)^\ell d^\ell_{2h,-2h}(\theta)\,. \tag{3.141}$$

Note that $n_\ell^{\{+h,-h,-h,+h\}} = (-1)^\ell n_\ell^{\{+h,-h,+h,-h\}}|_{\theta\to\theta+\pi}$.[4] For example, the polynomials for

[4] We thank Z. Bern, A. Zhiboedov, and D. Kosmopoulos for pointing out this relation.

the first few spins in $n_\ell^{\{+1,-1,+1,-1\}}$ are:

$$\begin{aligned} d^2_{2,2}(\theta) &= \cos^4\frac{\theta}{2} \\ d^3_{2,2}(\theta) &= \cos^4\frac{\theta}{2}(3\cos\theta-2) \\ d^4_{2,2}(\theta) &= \cos^4\frac{\theta}{2}(1-7\cos\theta+7\cos^2\theta) \\ d^5_{2,2}(\theta) &= \cos^4\frac{\theta}{2}(1+3\cos\theta-18\cos^2\theta+15\cos^3\theta)\,. \end{aligned} \tag{3.142}$$

Note that one starts from $\ell = 2$ a reflection of Landau-Yang's theorem.

Now following the previous discussion, since the Wigner d-matrices are also orthogonal polynomials, we expect that their Taylor vectors yield a positive definite matrix when the spins are ordered. Indeed consider the Taylor vectors for $d^\ell_{2,2}(\theta)$ expanded around $\theta = 0$. The Taylor vectors for spins $2, 3, \cdots, 9$ are given as:

$$h{=}1: \quad \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{7}{4} & 3 & \frac{9}{2} & \frac{25}{4} & \frac{33}{4} & \frac{21}{2} \\ 0 & 0 & \frac{7}{4} & \frac{27}{4} & \frac{135}{8} & \frac{275}{8} & \frac{495}{8} & \frac{819}{8} \\ 0 & 0 & 0 & \frac{15}{4} & \frac{165}{8} & \frac{275}{4} & \frac{715}{4} & \frac{3185}{8} \\ 0 & 0 & 0 & 0 & \frac{495}{64} & \frac{3575}{64} & \frac{15015}{64} & \frac{47775}{64} \\ 0 & 0 & 0 & 0 & 0 & \frac{1001}{64} & \frac{9009}{64} & \frac{5733}{8} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1001}{32} & \frac{10829}{32} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1989}{32} \end{pmatrix}. \tag{3.143}$$

It is straightforward to verify that, just as the vectors from Gegenbauer polynomials, the above is a totally positive matrix. Thus we see that the convex hull of the Taylor vectors from the spinning polynomial yields a cyclic polytope.

3.7 The s-channel EFT-hedron

In the previous section we've seen that for $\mathbf{a}$ to reside inside a convex hull, the geometry set up in eq.(3.87, 3.88), it can be cast into a (infinite) set of positivity conditions:

$$f_i(\mathbf{a}) \geq 0\,. \tag{3.144}$$

The explicit function f_i depends on the vectors that constitute the hull, and can be linear or non-linear functions of $\mathbf{a}$. Let us now explore the geometry for the simplest class of EFTs where the massless degrees of freedom are colored state. We can then focus on color ordered four-point amplitude and assume the absence of UV states in the u-channel. In such case we have eq.(3.87)

$$a_{k,q} = \sum_a \mathsf{p}_a \frac{v_{\ell_a,q}}{(m_a^2)^{k+1}} \qquad \mathsf{p}_a \geq 0\,. \tag{3.145}$$

where once again $v_{\ell,q}$ is the q-th Taylor coefficient in expanding $G_\ell(1+2\delta)$. The couplings $a_{k,q}$ are naturally dimensionful, but since our bounds will be projective in nature, only dimensionless ratios will be constrained. Note that since we are considering color ordered amplitudes, cyclic symmetry implies that the amplitude is symmetric under $s \leftrightarrow t$. Translated to the EFT couplings we have that they must lie on the "cyclic plane" $\mathbf{X}_{\mathrm{C}yc}$ defined by

$$a_{k,q} = a_{k,k-q}\,. \tag{3.146}$$

Thus the geometry of interest will be the intersection of the convex hull in eq.(3.145), with the cyclic plane $\mathbf{X}_{\mathrm{C}yc}$.

Recall that the origin of eq.(3.145) is the fact that the low energy amplitudes can be reproduced from the s-channel singularities. This can be recast into the following equivalence:

$$\sum_{k,q} a_{k,q} \;\; s^{k-q}t^q \;=\; \sum_a -\frac{\mathsf{p}_a\, G_{\ell_a}\left(1+2\frac{t}{m_a^2}\right)}{s-m_a^2} \quad for\ s,t \ll m^2\,, \tag{3.147}$$

where the equality is understood as the matching of Taylor series in s, t on both sides, with $n \geq 2$. Thus the sum on the RHS is only expected to be reproduced $a_{k,q}$ with $q \leq k-2$. Writing out the Taylor series for the RHS,

$$\begin{aligned}\sum_{k,q} a_{k,q} \;\; s^{k-q}t^q \;&= \sum_a \frac{\mathsf{p}_a}{m_a^2}\left(1+\frac{s}{m_a^2}+\frac{s^2}{m_a^4}+\cdots\right)\\ &\cdot\left(v_{\ell_a,0}+v_{\ell_a,1}\frac{t}{m_a^2}+v_{\ell_a,2}\left(\frac{t}{m_a^2}\right)^2+\cdots\right)\,,\end{aligned} \tag{3.148}$$

we immediately see the emergence of two types of geometries: one is the coefficients associated with the expansion in t and the other is the expansion in s. The geometry encoded in the former is a reflection of UV Lorentz invariance, since the convex hull depends on the details of the Gegenbauer polynomials, while the geometric series of the later reflects locality, i.e. that the only singularities of the four-point amplitude are in the Mandelstam variables. We will begin our analysis by disentangling the two geometry, taking the point of view of either fixed k or fixed q, and end in the geometry that is defined by its union.

Fixed k: the Gegenbauer cyclic polytope

Identifying the coefficient for $s^{k-q}t^q$ on both sides of eq.(3.148), we have

$$a_{k,q} = \sum_a \mathsf{p}_a \left[x_a^{k+1}\, v_{\ell_a,q} \right] \quad x_a \equiv \frac{1}{m_a^2}. \tag{3.149}$$

Now consider terms with the same mass-dimension, corresponding to fixed k. We write

$$\mathbf{a}_k = \begin{pmatrix} a_{k,0} \\ a_{k,1} \\ \vdots \end{pmatrix} = \sum_a \mathsf{p}_a x_a^{k+1} \begin{pmatrix} v_{\ell_a,0} \\ v_{\ell_a,1} \\ \vdots \end{pmatrix}. \tag{3.150}$$

Since $\mathsf{p}_a, x_a > 0$, this says that

$$\mathbf{a}_k \in Conv[\mathbf{G}_\ell]\,, \tag{3.151}$$

that is, the coefficients for the distinct polynomials associated with the mass-dimension $2k{+}4$ operator must live inside the Gegenbauer cyclic polytope! We will refer to $Conv[\mathbf{G}_\ell]$ as the unitary polytope $\mathbf{U}_k$, where the subscript k indicates that the polytope is in $\mathbb{P}^{k-2}$. The dimension is projectively $k{-}2$, since there are $k{+}1$ distinct polynomials at given k, with $a_{k,k}$ and $a_{k,k-1}$ not subject to the constraints implied by eq.(3.145).

Furthermore, cyclic symmetry requires that the couplings lie on the cyclic plane $\mathbf{X}_{cyc}$. For $k < 5$ cyclic symmetry simply relates the coefficients $a_{k,k}$ and $a_{k,k-1}$ to those that are constrained by $\mathbf{U}_k$. For $k \geq 5$ the cyclic plane $\mathbf{X}_{cyc}$ defines a $\lceil\frac{k+1}{2}\rceil{-}1$-dimensional subspace inside $\mathbf{U}_k$, i.e. the space of allowed couplings are now given by the intersection of the cyclic plane $\mathbf{X}_{cyc}$ with the unitary polytope $\mathbf{U}_k$, i.e. $\mathbf{U}_k \cap \mathbf{X}_{cyc}$, as illustrated in fig.(3.5). In the following, we will consider explicit examples up to $k = 5$.

- $k{=}2$: $D^4\phi^4$:

  $$M_{D^4\phi^4}(s,t) = (a_{2,0}s^2{+}a_{2,1}st{+}a_{2,2}t^2). \tag{3.152}$$

  We will only be able to bound $a_{2,0}$ and the geometry is $\mathbb{P}^0$. From the fact that $v_{\ell,0}$ is a positive number, we simply have $a_{2,0} > 0$, the forward limit positivity bound discussed in [4].

- $k{=}3$: $D^6\phi^4$

  $$M_{D^6\phi^4}(s,t) = (a_{3,0}s^3{+}a_{3,1}s^2t{+}\cdots). \tag{3.153}$$

  From now on we'll suppress listing the couplings that cannot be bounded. The geometry is now $\mathbb{P}^1$, and $\mathbf{a}_3 = (1, \frac{a_{3,1}}{a_{3,0}})$ is bounded by the minimum and maximum value of $\frac{v_{\ell,1}}{v_{\ell,0}}$, which are 0 and ∞ respectively. Thus we simply have $a_{3,0}, a_{3,1} > 0$.

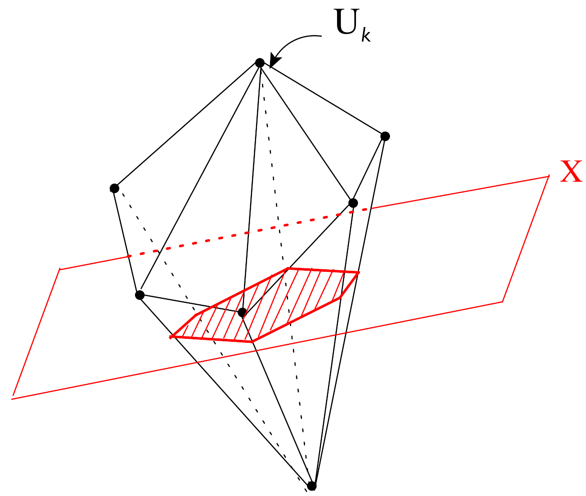


Figure 3.5: The s-channel geometry at fixed k. The vector $\mathbf{a}_k$ must live on the intersection between the cyclic plane $\mathbf{X}_{cyc}$ with the unitary polytope $\mathbf{U}_k$.

- k=4 : $D^8\phi^4$

$$M_{D^8\phi^4}(s,t) = (a_{4,0}s^4 + a_{4,1}s^3t + a_{4,2}s^2t^2 + \cdots). \tag{3.154}$$

We have $\mathbf{a}_4 = (1, \frac{a_{4,1}}{a_{4,0}}, \frac{a_{4,2}}{a_{4,0}}) \equiv (1, x, y)$. The boundaries of the two-dimensional polygon are given by $(i, i{+}1)$, and the constraint on $\mathbf{a}_4$ is given by $\langle \mathbf{a}_4, i, i{+}1\rangle > 0$ and $\langle \mathbf{a}_4, \infty, 0\rangle > 0$, where

$$\langle \mathbf{a}_4, i, i{+}1\rangle = \mathrm{D}et \begin{pmatrix} 1 & v_{i,0} & v_{i+1,0} \\ x & v_{i,1} & v_{i+1,1} \\ y & v_{i,2} & v_{i+1,2} \end{pmatrix} \tag{3.155}$$

Listing the first sets of constraints:

$$\begin{aligned} \langle \mathbf{a}_4, 0, 1\rangle > 0 \Rightarrow y > 0,\ \langle \mathbf{a}_4, 1, 2\rangle > 0 \Rightarrow 6{-}3x{+}2y > 0, \\ \langle \mathbf{a}_4, 2, 3\rangle > 0 \Rightarrow 18{-}4x{+}y > 0\,. \end{aligned} \tag{3.156}$$

The combined constraint is plotted in fig.3.6.

- k=5 : $D^{10}\phi^4$

$$M_{D^{10}\phi^4}(s,t) = (a_{5,0}s^5 + a_{5,1}s^4t + a_{5,2}s^3t^2 + a_{5,3}s^2t^3 + \cdots). \tag{3.157}$$

In this case, the cyclic plane $\mathbf{a}_5 \in \mathbf{Y} = (1, x, y, y)$ is two-dimensional and thus represents a subspace of the three-dimensional unitary polytope $\mathbf{U}_5$. There are two

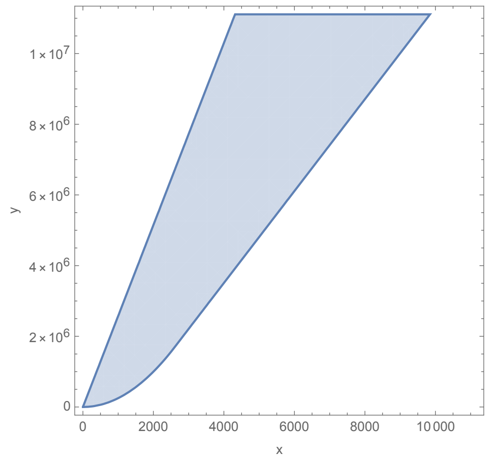


Figure 3.6: The allowed region satisfying $\langle \mathbf{a}_4, i, i{+}1\rangle > 0$. We have plotted the combined constraint for $i \leq 40$. For larger is, the constraint does not appear for the range of (x, y) displayed in the plot.

sets of constraints coming from $\langle 0, \mathbf{a}_5, i, i{+}1\rangle > 0$ and $\langle \mathbf{a}_5, i, i{+}1, \infty\rangle > 0$, given as:

$$\langle 0, \mathbf{a}_5, i, i{+}1\rangle = \begin{pmatrix} 1 & 1 & v_{i,0} & v_{i+1,0} \\ 0 & x & v_{i,1} & v_{i+1,1} \\ 0 & y & v_{i,2} & v_{i+1,2} \\ 0 & y & v_{i,3} & v_{i+1,3} \end{pmatrix},$$

$$\langle \mathbf{a}_5, i, i{+}1, \infty\rangle = \begin{pmatrix} 1 & v_{i,0} & v_{i+1,0} & 0 \\ x & v_{i,1} & v_{i+1,1} & 0 \\ y & v_{i,2} & v_{i+1,2} & 0 \\ y & v_{i,3} & v_{i+1,3} & 1 \end{pmatrix}. \tag{3.158}$$

The first set of constraints simply leads to $y \geq 0, x \geq \frac{y}{3}$, while the second set is shown in fig. 3.7. The combined constraint leads to a finite region composed of boundaries $(i, i{+}1, \infty)$ with $i = 0, 1, \cdots, 4$ and $(0, 4, 5)$ as shown in fig.3.8.

The fact that the ratio of coefficients $\frac{a_{k,q}}{a_{k,0}}$ are bounded within finite regions tells us that, in the on-shell basis, it is not only unnatural to have two distinct operators with the same dimension yet large differences in their coupling constants, *unitarity in the UV tells us that it is impossible to do so !*

Let's see where explicit EFTs sit inside $\mathbf{U}_k \cap \mathbf{X}_{cyc}$. Consider the open superstring four-gluon amplitude in eq.(3.14), where its low-energy expansion is given in eq.(3.16). Strip-

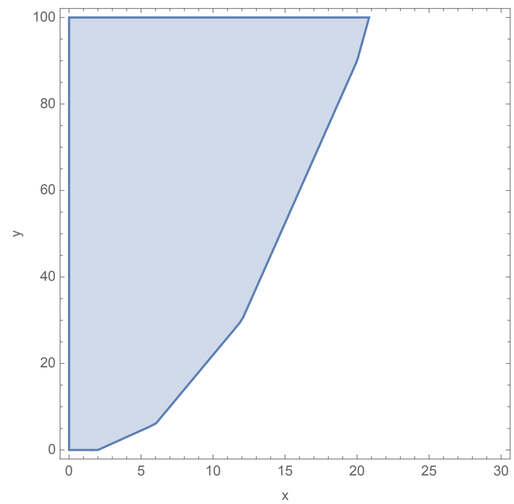


Figure 3.7: The constraints curved out from $\langle \mathbf{a}_5, i, i{+}1, \infty\rangle > 0$.

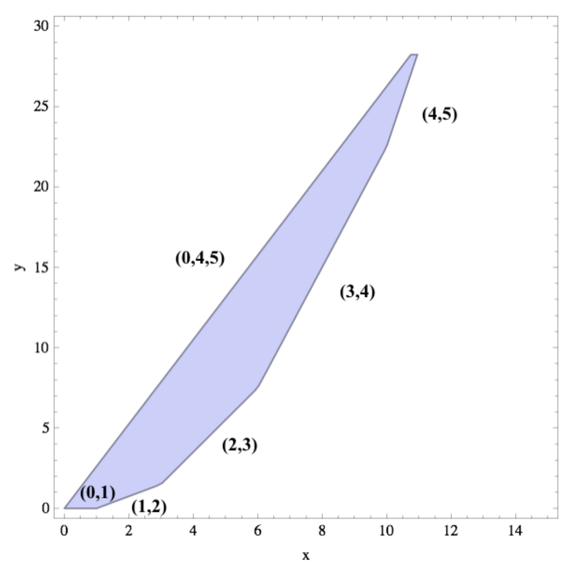


Figure 3.8: The projection of the unitary polytope onto the cyclic plane at $k = 5$. The boundary is given by $(0, 4, 5)$ as well as $(i, i{+}1, \infty)$ for $i = 0, \cdots, 4$, displayed as $(i, i{+}1)$.

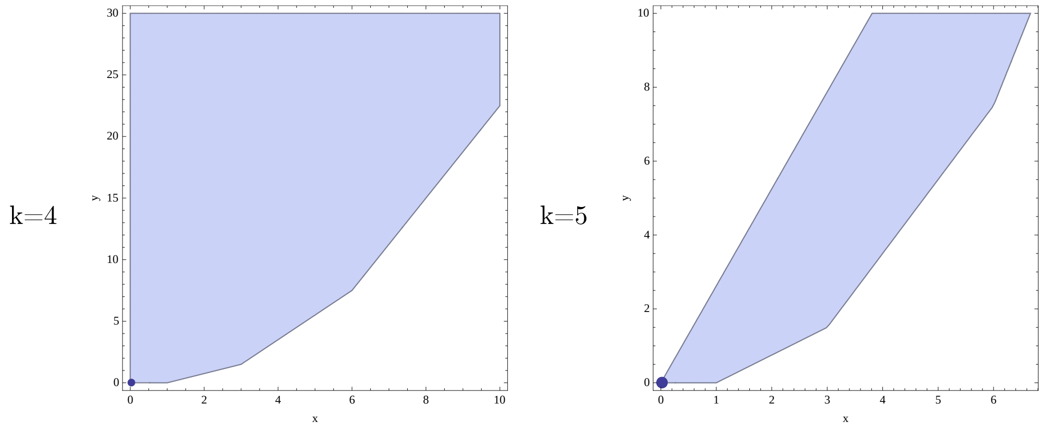


Figure 3.9: The position of the string theory coefficients given in eq.(3.159) inside the region $\mathbf{U}_k \cap \mathbf{Y}$, for $k = 4, 5$ respectively.

ping off the spinor brackets and considering the expansion up to $k = 5$ we find,

$$
\begin{aligned}
k=2: &\qquad a_{2,0} = \frac{2\zeta_2^2}{5}, \quad k=3: \quad a_{3,0} = \zeta(5), \quad a_{3,1} = 2\zeta(5) - \zeta(3)\zeta(2) \\
k=4: &\qquad (x,y) = \left(\frac{a_{4,1}}{a_{4,0}}, \frac{a_{4,2}}{a_{4,0}}\right) = \left(\frac{3}{4} - \frac{945\zeta_3^2}{2\pi^6}, \frac{23}{20160} - \frac{3\zeta_3^2}{4\pi^6}\right) \\
k=5: &\qquad (x,y) = \left(\frac{a_{5,1}}{a_{5,0}}, \frac{a_{5,2}}{a_{5,0}}\right) = \left(3 - \frac{\pi^4\zeta_3 + 15\pi^2\zeta_5}{90\zeta_7}, 5 - \frac{\pi^4\zeta_3 + 24\pi^2\zeta_5}{72\zeta_7}\right).
\end{aligned}
\tag{3.159}
$$

For $k = 2, 3$ the coefficients are not only inside $\mathbf{U}_k$, it close to the "boundary". This behaviour is more prominent for $k = 4, 5$ where the EFT couplings are close to the boundary composed of low spins, as we display in fig.(3.9). This indicates that the p_as in eq.(3.145) is dominated by contributions from low spin sector. In fact, in section 3.10 we will see that such behaviour is common amongst all known EFTs.

Fixed q: Hanekl matrix constraints

Instead of fixed k and considering the constraint on $\mathbf{a}_k$, let's now examine the geometry associated with fixed q, i.e. that associated with the first parentheses on the RHS of eq.(3.148). First taking $q = 0$, we have

$$
a_{k,0} = \sum_a \mathsf{p}'_a (x_a)^k \,, \tag{3.160}
$$

where $\mathsf{p}'_a = x_a \mathsf{p}_a \, v_{\ell,0}$, and the equality holds for the $k \geq 2$. Since $v_{\ell,0} = G_\ell(1)$ is positive, $\mathsf{p}'_a > 0$. We immediately see that eq.(3.160) implies $a_{k,0} > 0$, which is the forward limit positivity bound discussed in [4] extended to higher derivatives. We've seen this before in section 3.6, where the vector

$$\tilde{\mathbf{a}}_0 = \begin{pmatrix} 1 \\ \frac{a_{3,0}}{a_{2,0}} \\ \frac{a_{4,0}}{a_{2,0}} \\ \vdots \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} 1 \\ x_a \\ \vdots \\ x_a^k \end{pmatrix} , \quad x_a \equiv \frac{1}{m_a^2} , \tag{3.161}$$

lies in the convex hull of points on a half moment curve, and thus the Hankel matrix of its entries $K[\tilde{a}_0]$ is a totally positive matrix. Note that since $v_{\ell,q} > 0$ for all q, the same holds true for any fixed q. Thus in general we have:

$$\boxed{K\left[\tilde{\mathbf{a}}_q\right] \in \text{T}otal\ positive\ matrices\ \ \forall q\,.} \tag{3.162}$$

Once again, lets us demonstrate this for the Type-I string amplitude. Collecting the coefficients as

$$\vec{a}_0 = \begin{pmatrix} \frac{2}{5}\zeta_2^2 \\ \zeta_5 \\ \frac{8}{35}\zeta_2^3 \\ \zeta_7 \\ \frac{24}{175}\zeta_2^4 \end{pmatrix} , \quad \vec{a}_1 = \begin{pmatrix} 2\zeta_5 - \zeta_2\zeta_3 \\ \frac{6}{35}\zeta_2^3 - \frac{1}{2}\zeta_3^2 \\ 3\zeta_7 - \zeta_2\zeta_5 - \frac{2}{5}\zeta_2^2\zeta_3 \\ \frac{6}{35}\zeta_2^4 - \zeta_3\zeta_5 \\ 4\zeta_9 - \zeta_2\zeta_7 - \frac{2}{5}\zeta_2^2\zeta_5 - \frac{8}{35}\zeta_2^3\zeta_3 \end{pmatrix} , \tag{3.163}$$

The corresponding Hankel matrix are,

$$\begin{aligned} K\left[\vec{a}_0\right] &= \begin{pmatrix} \frac{2}{5}\zeta_2^2 & \zeta_5 & \frac{8}{35}\zeta_2^3 \\ \zeta_5 & \frac{8}{35}\zeta_2^3 & \zeta_7 \\ \frac{8}{35}\zeta_2^3 & \zeta_7 & \frac{24}{175}\zeta_2^4 \end{pmatrix} \\ K\left[\vec{a}_1\right] &= \begin{pmatrix} 2\zeta_5 - \zeta_2\zeta_3 & \frac{6}{35}\zeta_2^3 - \frac{1}{2}\zeta_3^2 & 3\zeta_7 - \zeta_2\zeta_5 - \frac{2}{5}\zeta_2^2\zeta_3 \\ \frac{6}{35}\zeta_2^3 - \frac{1}{2}\zeta_3^2 & 3\zeta_7 - \zeta_2\zeta_5 - \frac{2}{5}\zeta_2^2\zeta_3 & \frac{6}{35}\zeta_2^4 - \zeta_3\zeta_5 \\ 3\zeta_7 - \zeta_2\zeta_5 - \frac{2}{5}\zeta_2^2\zeta_3 & \frac{6}{35}\zeta_2^4 - \zeta_3\zeta_5 & 4\zeta_9 - \zeta_2\zeta_7 - \frac{2}{5}\zeta_2^2\zeta_5 - \frac{8}{35}\zeta_2^3\zeta_3 \end{pmatrix} . \end{aligned} \tag{3.164}$$

It is straightforward to check that all minors of the above Hankel matrix are indeed positive. A more detailed study of the Hankel matrix constraint for superstring amplitude was recently done in [89].

It is interesting to ask which theories lie on boundaries of the Hankel constraints, i.e., for which theories do all the minors of the Hankel matrix greater than some size all vanish? The answer is extremely simple and satisfying. Only UV amplitudes with a finite number of poles satisfy this property; that is, only UV theories with N massive states

$$
\begin{array}{cccccc}
 & v_{\ell_a,0} & v_{\ell_a,1} & v_{\ell_a,2} & v_{\ell_a,3} & \cdots \\
1 & a_{0,0} & & & & \\
x_a & a_{1,0} & a_{1,1} & & & \\
x_a^2 & a_{2,0} & a_{2,1} & a_{2,2} & & \\
x_a^3 & a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
$$

Figure 3.10: We organize the information that each state contributes to the determination of $a_{k,q}$. For each fixed row (fixed k), for example the red box, each state's contribution is proportional to a Gegenbauer vector multiplied by a universal factor x_a^k. For a fixed column (fixed q), the purple box, each state contributes to a point on a half moment curve multiplied by universal factor $v_{\ell_a,q}$.

exchanged at tree-level lie on the boundary of the Hankel constraints. This can be seen from the analytic expression of the determinants in eq.(3.123), where it is proportional to the Vandermonde determinant of the masses of the UV state x_a. This gives us a way to "detect" the number of massive states: if there are a massive states, then the $(a+1)\times(a+1)$ determinant vanishes.

The s-channel EFT-hedron

Up to now, we've been considering the constraints from the two parentheses in eq.(3.148) separately. These, however, are not the full set of constraints. To see this it is useful to organize the information each state contributes to $a_{k,q}$ as in fig.3.10. For a given row, each state contributes a fixed positive factor x_a^k multiplying the Gegenbauer vector, which led to the constraint that the row vectors must lie in the convex hull of a cyclic polytope. For a fixed column, each state contributes a point on the half moment-curve weighted by a positive factor $v_{\ell,q}$, and thus implying the constraint that the Hankel matrix of the column vector is a totally positive matrix.

As one can see from the above description, these are not the complete constraints. For example, the cyclic polytope constraint does not tell us that the positive proportionality factor takes the form x_a^k, which is only visible if we consider different ks at the same time. Put in another way, if we truncate our expansion of t to a fix order, say the first order, we should see that for different moments (x_a^k), each state contributes the **same vector** $(v_{\ell_a,0}, v_{\ell_a,1})$, as illustrated in fig.3.11. In other words, not only does each row must lie in the cyclic polytope, but it must be the same point after scaling away the moment factors!

To recap, the space of higher dimensional operator is given by the tensor product of two positive geometries, the Gegenbauer cyclic polytope and convex hull of half moment curve, and we would like to find the *full* set of inequalities that carve out this space.

$$
\begin{array}{cccccc}
 & v_{\ell_a,0} & v_{\ell_a,1} & v_{\ell_a,2} & v_{\ell_a,3} & \cdots \\
1 & a_{0,0} & & & & \\
x_a & a_{1,0} & a_{1,1} & & & \\
x_a^2 & a_{2,0} & a_{2,1} & a_{2,2} & & \\
x_a^3 & a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
$$

Figure 3.11: For a given state, its contribution to each row is the same vector $(v_{\ell_a,0}, v_{\ell_a,1})$ after scaling away the moment factor x_a^k.

This is reminiscent of the (tree) Amplituhedron which gives the scattering amplitude of $\mathcal{N} = 4$ SYM [11]. There we have a subspace of k-planes in k+4 dimensions, Y_α^I, given by the product of two positive geometries

$$
Y_\alpha^I = \sum_{i=1,n} C_{\alpha,i} Z_i^I, \qquad C_{\alpha,i} \in Gr_{>0}(k,n), \quad Z_i^I \in M_+(n, k{+}4) \tag{3.165}
$$

where the $C_{\alpha,i}$ is in the positive Grassmannian $Gr_{>0}(k,n)$, a $k \times n$ matrix with all ordered minors positive mod GL(k), and Z_i^I is a $n \times k{+}4$ positive matrix with positive ordered minors. The Zs are the "external data" that is given and already in the positive region. Note that for $k = 1$, this is simply a polytope in $\mathbb{P}^4$. To carve out this space via inequalities, we require that Y satisfies:

$$
\langle Y_1 Y_2 \cdots Y_k Z_i Z_{i+1} Z_j Z_{j+1} \rangle > 0\,. \tag{3.166}
$$

To see this note that we can interpret eq.(3.165) as expanding Y_α^I on the the "basis" Z_i^I, with coefficients $C_{\alpha,i}$. Then the above condition implies

$$
\langle Y_1 Y_2 \cdots Y_k Z_i Z_{i+1} Z_j Z_{j+1} \rangle = \sum_{i_1 < i_2 < \cdots < i_k} \langle C_{i_1} C_{i_2} \cdots C_{i_k} \rangle \langle Z_{i_1} Z_{i_2} \cdots Z_{i_k} Z_i Z_{i+1} Z_j Z_{j+1} \rangle > 0. \tag{3.167}
$$

For this to hold for any choice of $Z_i^I \in M_+(n, k{+}4)$, forces $C_{\alpha,i} \in Gr_{>0}(k,n)$.

For our case, the fixed external data is the Gegenbauer vectors, which automatically yield positive matrices. This motivates us to first organize all the states with the same spin together and rewrite eq.(3.149) as:

$$
a_{k,q} = \sum_a \mathsf{p}_a \left[x_a^{k+1}\, v_{\ell_a,q} \right] \equiv \sum_\ell C_{k,\ell} V_{\ell,q}\,. \tag{3.168}
$$

Here $V_{\ell,q} = v_{\ell,q}$, and $C_{k,\ell} = \sum_{\{a:\ell_a=\ell\}} p_a\, x_a^{k+1}$, where one sums over all the states with the same fixed spin ℓ. Collecting the Cs into a column vector $\mathbf{C}_\ell = \{C_{1,\ell}, C_{2,\ell}, \cdots, C_{k,\ell}\}$, we see that $\mathbf{C}_\ell$ is inside the convex hull of the half moment curve. We are now ready to define the EFT-hedron: the space of consistent coefficients of higher dimension operators is given by the product (with $k \geq q$)

$$a_{k,q} = \sum_{\ell} C_{k,\ell} V_{\ell,q} \tag{3.169}$$

where $C_{k,\ell}$ is positive in the sense that $K[\mathbf{C}_\ell]$ is a totally positive matrix for each ℓ, and $V_{\ell,q}$ is positive in that any ordered minor of the vectors are positive. Let us make a comparison with the amplituhedron [11]. For the EFT-hedron the positivity property in C is defined for each column (spin) independently, while for the amplituhedron the C being in $Gr_{>0}(k,n)$, the positivity condition mixes the columns. For the amplituhedron I is locked in with k being $4{+}k$ dimensional, while for the EFT-hedron q can be any dimension independent of k.

Now let us carve out the space via inequalities. Consider a set of "walls", which are dual vectors $\mathcal{W}_I^q$, labelled by I, satisfying

$$\sum_q \mathcal{W}_I^q V_{\ell,q} \geq 0, \quad \forall \ell. \tag{3.170}$$

Unit vectors $\{0,0,1,\cdots,0\}$ trivially satisfies this criteria due to the positivity of the Gegenbauer Taylor coefficients. We denote these as $\mathcal{W}_{I_{\mathbb{I}}}^q$. There are also walls comprised of the facets of $Conv[V_\ell]$, taking the form $(i, i{+}1)$, $(1, i, i{+}1)$, e.t.c, which in dual vector form is given by $\langle *, i, i+1\rangle$, $\langle *, 1, i, i+1\rangle$. We denote these as $\mathcal{W}_{I_b}^q$. Given these walls we take the inner product with the higher dimension operators. Define

$$A_{k,I} \equiv \sum_q a_{k,q} \mathcal{W}_I^q, \quad \forall\ \mathcal{W}_I^q \in \{\mathcal{W}_{I_{\mathbb{I}}}^q, \mathcal{W}_{I_b}^q\} \tag{3.171}$$

and the EFT-hedron is carved out by the inequality

$$\boxed{K[\vec{A}_I] \quad \text{is a totally positive matrix}}\ . \tag{3.172}$$

where $\vec{A}_I = (A_{0,I}, A_{1,I}, \cdots)$. In other words, for any of one of the walls $\mathcal{W}_I^q$, the $A_{k,I}$s satisfies the following infinite set of constraints

$$A_{0,I} \geq 0, \quad A_{1,I} \geq 0, \quad \mathrm{D}et\begin{pmatrix} A_{0,I} & A_{1,I} \\ A_{1,I} & A_{2,I} \end{pmatrix} \geq 0, \quad \mathrm{D}et\begin{pmatrix} A_{1,I} & A_{2,I} \\ A_{2,I} & A_{3,I} \end{pmatrix} \geq 0$$
$$\mathrm{D}et\begin{pmatrix} A_{0,I} & A_{1,I} & A_{2,I} \\ A_{1,I} & A_{2,I} & A_{3,I} \\ A_{2,I} & A_{3,I} & A_{4,I} \end{pmatrix} \geq 0, \quad \mathrm{D}et\begin{pmatrix} A_{1,I} & A_{2,I} & A_{3,I} \\ A_{2,I} & A_{3,I} & A_{4,I} \\ A_{3,I} & A_{4,I} & A_{5,I} \end{pmatrix} \geq 0, \cdots . \tag{3.173}$$

Before closing, let us confirm that the inequalities in eq.(3.172), combined with the information of the walls, indeed carve out the space in eq.(3.168). First take the walls to be the unit vectors, and then $K[\vec{A}_{I_{\mathbb{I}}}]$ being a totally positive matrix simply implies

$$a_{k,q} = \sum_a p_{a,q}(x_a)^k, \quad p_{a,q} > 0 \quad x_a > 0\,, \tag{3.174}$$

i.e. for each fixed q, the vector $\vec{a}_q = (a_{1,q}, a_{2,q}, \cdots)$ lies in the convex hull of half moment curves. Next, we consider the walls that are the boundaries of the $Conv[V_\ell]$. The positivity of individual A_{k,I_b} tells us that each row $a_{k,q}$ is inside $Conv[V_\ell]$. This combined with the previous result tells us that

$$a_{k,q} = \sum_{a,\ell} p_a(x_a)^k\, \mathcal{O}_{a,k,\ell}\, V_{\ell,q}, \quad p_a > 0\,, \quad x_a > 0\,, \quad \mathcal{O}_{a,k,\ell} > 0\,. \tag{3.175}$$

Finally, the total positivity of $K[\vec{A}_{I_b}]$ then tell us that $\mathcal{O}_{a,k,\ell}$ must be such that

$$(x_a)^k\, \mathcal{O}_{a,k,\ell} = (x'_{a,\ell})^k.$$

In other words,

$$a_{k,q} = \sum_{a,\ell} p_a(x'_{a,\ell})^k\, V_{\ell,q}, \quad p_a > 0 \quad x'_{a,\ell} > 0\,. \tag{3.176}$$

We see that indeed eq.(3.168) is recovered.

The geometry of the gap

Let's suppose we have the extra information of the scale of the UV completion, i.e. the UV spectrum starts at M_{Gap} above the massless modes. This allows us to write

$$a_{k,0} = \sum_a \frac{\mathsf{p}_a}{M_{\text{Gap}}^{2(k+1)}} \left(\frac{M_{\text{Gap}}}{m_a}\right)^{2(k+1)} = \frac{1}{M_{\text{Gap}}^{2(k+1)}} \sum_a \mathsf{p}_a\, x_a^{k+1}, \quad x_a \leq 1\,. \tag{3.177}$$

Now since $x_a \leq 1$, we see that the gap implies

$$a_{2,0} \geq M_{\text{Gap}}^2 a_{3,0} \geq \cdots \geq M_{\text{Gap}}^{2(k-2)} a_{k,0} \geq 0\,. \tag{3.178}$$

The fact that $x_a \leq 1$ also tells us that the convex hull of $\mathbf{a}_k$ is now over a restricted region of the half-moment curve:

$$\mathbf{a}_k = \sum_a \mathsf{p}_a \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{M^2_{\text{Gap}}} & 0 & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & 0 & \frac{1}{M^{2(k-2)}_{\text{Gap}}} \end{pmatrix} \begin{pmatrix} 1 \\ \left(\frac{M_{\text{Gap}}}{m_a}\right)^2 \\ \vdots \\ \left(\frac{M_{\text{Gap}}}{m_a}\right)^{2(k-2)} \end{pmatrix}$$
$$\rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & M^2_{\text{Gap}} & 0 & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & 0 & M^{2(k-2)}_{\text{Gap}} \end{pmatrix} \mathbf{a}_k = \sum_a \mathsf{p}_a \begin{pmatrix} 1 \\ x_a \\ \vdots \\ x_a^{k-2} \end{pmatrix}, \quad \mathsf{p}_a > 0,\ x_a \leq 1\,, \tag{3.179}$$

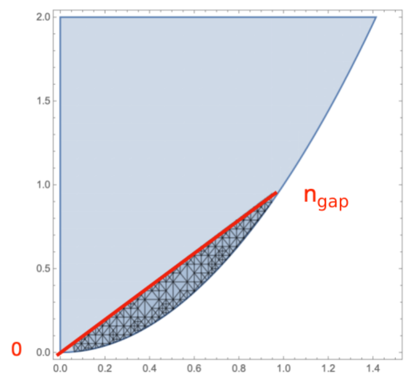
that is, instead of $x \in R^+$ we now have $x \in [0,1]$. For simplicity we set $M^2_{\text{Gap}} = 1$ from now on, and we write:

$$\mathbf{a}_k = \sum_a \mathsf{p}_a \begin{pmatrix} 1 \\ x_a \\ \vdots \\ x_a^{k-2} \end{pmatrix}, \quad \mathsf{p}_a > 0,\ x_a \leq 1\,. \tag{3.180}$$

where the components of $\mathbf{a}_k$ have been rescaled by appropriate factors of M^2_{Gap} to be dimensionless. Now since the curve is bounded by $x_a = 1$, we now have a new boundary vertex

$$n_{\text{Gap}} = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}. \tag{3.181}$$

The change in geometry is fully illustrated in the following $\mathbb{P}^2$ example



where the convex hull now has a new boundaries consisting of $(0,n)$, with 0 denoting the spin-0 vector. Extending to higher dimensions we now have a new set of boundary consists of $(0,i,i{+}1,\cdots,n)$, and thus besides the usual Hankel matrix constraints, $\mathbf{a}$ now must also respect

$$\langle 0,\mathbf{a},i,i{+}1,\cdots,j,j{+}1,n\rangle > 0\,. \tag{3.182}$$

where we recall $(i,i{+}1)\rightarrow(i,\dot{i})$.

Now the new constraint eq.(3.182) can be translated to the geometry projected through the line $(0,n)$. To see this geometry cleanly, we take a GL transformation G that keeps 0 fixed and rotate n to:

$$G\,0=\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix},\quad G\,n=\begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix},\rightarrow\quad G=\begin{pmatrix}1&-1&0&0&0\\0&1&0&0&0\\0&1&-1&0&0\\\vdots&\vdots&\vdots&\vdots&\vdots\\0&0&0&1&-1\end{pmatrix}. \tag{3.183}$$

The action of G on the moment curve yields

$$G\begin{pmatrix}1\\x\\x^2\\x^3\\\vdots\\x^d\end{pmatrix}=\begin{pmatrix}1{-}x\\x\\x(1{-}x)\\x^2(1{-}x)\\\vdots\\x^{d-1}(1{-}x)\end{pmatrix}. \tag{3.184}$$

Thus after the the GL transformation, the presence of $(0,n)$ in the determinant

$$\langle 0,\mathbf{a},i,i{+}1,\cdots,j,j{+}1,n\rangle$$

simply knocks out the first two components of the other vectors, and eq.(3.182) becomes

$$\langle \tilde{\mathbf{a}}, \tilde{i}, \tilde{i}{+}1, \cdots, \tilde{j}, \tilde{j}{+}1\rangle > 0, \tag{3.185}$$

where the " $\tilde{}$ " represents the GL transformed vector with the first two components removed. For example

$$G\,\mathbf{a} = \begin{pmatrix} a_2{-}a_3 \\ a_3 \\ a_3{-}a_4 \\ a_4{-}a_5 \\ a_5{-}a_6 \\ \vdots \end{pmatrix} \to \; \tilde{\mathbf{a}} = \begin{pmatrix} a_3{-}a_4 \\ a_4{-}a_5 \\ a_5{-}a_6 \\ \vdots \end{pmatrix}. \tag{3.186}$$

Now $\tilde{i}$ takes the form :

$$\begin{pmatrix} x_i(1{-}x_i) \\ x_i^2(1{-}x_i) \\ \vdots \\ x_i^{d-1}(1{-}x_i) \end{pmatrix} = x_i(1{-}x_i) \begin{pmatrix} 1 \\ x_i \\ \vdots \\ x_i^{d-2} \end{pmatrix} \tag{3.187}$$

which, since $0 < x \leq 1$, up to a positive factor is once again a moment curve! In other words, the constraint $\langle \tilde{\mathbf{a}}, \tilde{i}, \tilde{i}{+}1, \cdots, \tilde{j}, \tilde{j}{+}1\rangle > 0$ implies that $\tilde{\mathbf{a}}$, which are twisted sum of a_is, also satisfies the non-linear Hankel matrix constraint! For example, starting with $\mathbf{a} \in \mathbb{P}^4$, we have $\tilde{\mathbf{a}} = (a_4{-}a_3, a_5{-}a_4, a_6{-}a_5)$, and the Hanel matrix constraint implies $a_i > a_j$ for $i > j$ and

$$(a_3 - a_4)(a_5 - a_6) - (a_4 - a_5)^2 > 0\,. \tag{3.188}$$

The above argument is not all! We have just noted that $\tilde{i}$ is positively proportional to a moment curve, but once again since $x \leq 1$, it is a capped moment curve and we can reiterate our analysis! The above argument gives an intuitive explanation for the additional gapped Hankel constraints, but with hindsight it is also easy to derive them even more directly. We simply note that if $(a_2, a_3, a_4, a_5, \cdots)$ is in the convex hull of $(1, x, x^2, \cdots)$, then $(a_2{-}a_3, a_3{-}a_4, a_4{-}a_5, \cdots)$ is the the convex hull of $x(1{-}x) \times (1, x, x^2, \cdots)$. Since $x(1{-}x) \geq 0$ for $0 \leq x \leq 1$, this is the same as the hull of $(1, x,^2, \cdots)$. Thus the discrete derivative $(a_2{-}a_3, a_3{-}a_4, a_4{-}a_5, \cdots)$ must have a totally positive Hankel matrix!

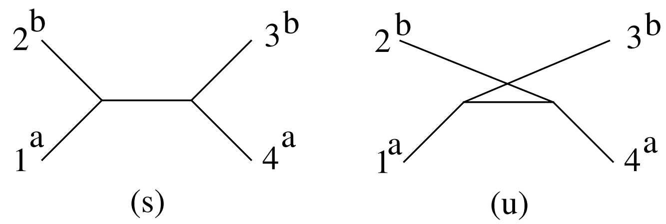
In summary, with a known gap, we can find that the following sequence of "twisted" couplings satisfies the positive Hankel matrix constraint:

$$\begin{pmatrix} a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ \vdots \end{pmatrix}, \quad \begin{pmatrix} a_3 - a_4 \\ a_4 - a_5 \\ a_5 - a_6 \\ a_6 - a_7 \\ \vdots \end{pmatrix}, \quad \begin{pmatrix} (a_4 - a_5) - (a_5 - a_6) \\ (a_5 - a_6) - (a_6 - a_7) \\ \vdots \end{pmatrix}. \tag{3.189}$$

This is known as the Hausdorff moment problem [158, 159]. The extra constraints from the knowledge of the gap are interesting, however, they are obviously of only academic interest to the low-energy observer that has no knowledge of the gap. Any higher-dimension operator measured by a low-energy observer could be produced by arbitrarily weakly coupled, arbitrarily low-mass states, and in the limit where the masses and couplings go to zero we recover the pure Hankel constraints. Note that the pure Hankel constraints are homogeneous in mass dimensions, comparing sums of products of couplings with the same total mass dimension, which are the only sorts of constraints we can talk about without knowledge of an absolute mass scale (such as the gap). For this reason, in the rest of this paper, we will focus on these types of universal constraints that can be sensibly formulated in the low-energy theory, assuming no knowledge of the gap.

3.8 Scalar EFT-hedron

So far we have restricted ourselves to the geometry arising from singularities on the positive real s-axis. For a general $2 \to 2$ process, $M(1^a, 2^b, 3^b, 4^a)$, the amplitude will have poles and discontinuities on *both* positive and negative real s-axes, reflecting s and u-channel exchanges:



The residue or discontinuity on the s-channel as a function of t will be identical to that in the u-channel since the two diagrams are related via $2 \leftrightarrow 3$ exchange. However, while the residues are the same, the u-channel singularities lie on the negative s-axes with a t-dependent shift: $u - m^2 = -s - (t{+}m^2)$. In other words, the low energy couplings are now governed by the Taylor expansion of:

$$-\sum_a \mathsf{p}_{\ell_a} \left[\frac{1}{s - m_a^2} + \frac{1}{-s - t - m_a^2}\right] G_{\ell_a}\left(1 + \frac{2t}{m_a^2}\right). \tag{3.190}$$

Recall that in the previous section, the s-channel EFT-hedron is the direct product of the positive geometry of the Gegenbauer vectors and that of the moment curve. Compared to the above one can see that we now have a new feature: upon Taylor expansion, the t in the u-channel will mix with that from $G_\ell(1 + \frac{2t}{m^2})$, and the two geometries are no longer a direct product, but "entangled".

Due to the s, u symmetry, it will be more convenient to parameterize our kinematics as

$$s = -\frac{t}{2} + z, \quad u = -\frac{t}{2} - z\,, \tag{3.191}$$

and the four-point amplitude is a function of z, t, $M(z,t)$. The low energy couplings are now extracted from the Taylor expansion of:

$$-\sum_a \mathsf{p}_a \left(\frac{1}{-\frac{t}{2} - z - m_a^2} + \frac{1}{-\frac{t}{2} + z - m_a^2}\right) G_{\ell_a}\left(1 + \frac{2t}{m_a^2}\right) \tag{3.192}$$

The resulting Taylor expansion only has even powers of z, which is a reflection of the underlying $s \leftrightarrow u$ symmetry. If we consider the geometry associated with fixed k or fixed

q, then the geometry here is the Minkowski sum of the s- and u-channel convex hull. Thus we have[5]

$$a_{k,q}z^{k-q}t^q = \sum_a \mathsf{p}_a \left[x_a^{k+1}\; u_{\ell_a,k,q}\right] z^{k-q}t^q \quad k-q \in even\,. \tag{3.193}$$

where the coefficients $u_{\ell,k,q}$ are linear combinations of Gegenbauer Taylor coefficients $v_{\ell,q}$s:

$$u_{\ell,k,q} = \sum_{a+b=q} (-)^a \frac{(k-q+1)_a}{a!} 2^{b-a} v_{\ell,b}\,. \tag{3.194}$$

Thus for fixed k, the couplings must live inside $Conv[\vec{u}_{\ell,k}]$, where

$$\begin{aligned} k \in even: \qquad & \vec{u}_{\ell,k} = (u_{\ell,k,0},\, u_{\ell,k,2}, \cdots, u_{\ell,k,k}) \\ k \in odd: \qquad & \vec{u}_{\ell,k} = (u_{\ell,k,1},\, u_{\ell,k,3}, \cdots, u_{\ell,k,k})\,. \end{aligned} \tag{3.195}$$

Importantly, the vectors $\vec{u}_{\ell,k}$ are labeled by both the spin and k. This k-dependence was absent in the s-channel analysis, where $Conv[\vec{v}_\ell]$ only depends on spin. This new feature leads to an important distinction between s-channel and full EFT-hedron.

Due to the absence of z^{odd} terms, at fixed k the dimensionality of $\vec{u}_{\ell,k}$ is smaller than $\vec{v}_\ell$ (half for $k \in$ odd). More precisely, $\vec{u}_{\ell,k}$ is obtained by a GL rotation of $\vec{v}_\ell$ that projects away the odd components. For example, for $k \in even$ we have:

$$\vec{u}_{\ell,k} = \begin{pmatrix} u_{\ell,k,0} \\ 0 \\ u_{\ell,k,2} \\ \vdots \\ 0 \\ u_{\ell,k,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{(k-1)_2}{2}\frac{1}{2^2} & (k-1)_1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{(1)_k}{k!}2^{-k} & -\frac{(1)_{k-1}}{k-1!}2^{2-k} & \frac{(1)_{k-2}}{k-2!}2^{4-k} & \cdots & -2^{k-2} & 1 \end{pmatrix} \begin{pmatrix} v_{\ell,0} \\ v_{\ell,1} \\ v_{\ell,2} \\ \vdots \\ v_{\ell,k-1} \\ v_{\ell,k} \end{pmatrix}. \tag{3.196}$$

Due to this projection, $Conv[\vec{u}_{\ell,k}]$ does not inherit the positivity of $Conv[\vec{v}_\ell]$, and thus we cannot conclude that $Conv[\vec{u}_{\ell,k}]$ is a cyclic polytope. Similarly for fixed q, comparing the coefficient of x_a^{k+1} in eq.(3.193) with the s-channel eq.(3.149), we see that the k-dependence of $u_{\ell_a,k,q}$ results in each moment x_a^{k+1} being weighted differently, and we no longer have a momentum curve. Thus naively, the positivity geometry that defined the s-channel EFT-hedron is lost, and we no longer have control over the geometry. As we will now see, there is in fact a hidden positivity that retains most of the structure of the s-channel cyclic polytope, thus allowing us to carve out *the* EFT-hedron.

[5] Here we define the couplings $a_{k,q}$ with respect to powers of z, t. To avoid proliferation of new couplings, we will continue to use the notation $a_{k,q}$ where the context is obvious.

The $s{-}u$ polytope
Let us consider the boundaries of the $(s{-}u)$ polytope, i.e. $Conv[\vec{u}_{\ell,k}]$. We will be interested in the sign for the determinant of ordered $\vec{u}_{\ell,k}$s. Setting $k=4$ as an example, we find:

$$\begin{aligned}
&\mathrm{Det}\begin{pmatrix} \vec{u}_{\ell_1,4} & \vec{u}_{\ell_2,4} & \vec{u}_{\ell_3,4}\end{pmatrix} = \mathrm{Det}\begin{pmatrix} v_{\ell_1,0} & & \\ v_{\ell_1,2}-\frac{3}{4}v_{\ell_1,1} & \{\ell_2\} & \{\ell_3\} \\ v_{\ell_1,4}-\frac{1}{4}v_{\ell_1,3}+\frac{1}{16}v_{\ell_1,2}-\frac{1}{64}v_{\ell_1,1} & & \end{pmatrix}\\
=\ &\mathrm{Det}\begin{pmatrix} v_{\ell_1,0} & & \\ v_{\ell_1,2} & \{\ell_2\} & \{\ell_3\} \\ v_{\ell_1,4} & & \end{pmatrix} - \frac{3}{4}\mathrm{Det}\begin{pmatrix} v_{\ell_1,0} & & \\ v_{\ell_1,1} & \{\ell_2\} & \{\ell_3\} \\ v_{\ell_1,4} & & \end{pmatrix} - \frac{1}{32}\mathrm{Det}\begin{pmatrix} v_{\ell_1,0} & & \\ v_{\ell_1,1} & \{\ell_2\} & \{\ell_3\} \\ v_{\ell_1,2} & & \end{pmatrix}\\
-\ &\frac{1}{4}\mathrm{Det}\begin{pmatrix} v_{\ell_1,0} & & \\ v_{\ell_1,2} & \{\ell_2\} & \{\ell_3\} \\ v_{\ell_1,3} & & \end{pmatrix} + \frac{3}{16}\mathrm{Det}\begin{pmatrix} v_{\ell_1,0} & & \\ v_{\ell_1,1} & \{\ell_2\} & \{\ell_3\} \\ v_{\ell_1,3} & & \end{pmatrix} + \cdots\,,
\end{aligned} \tag{3.197}$$

where $\{\ell_i\}$ represents the same as the first column just with $\ell_1 \to \ell_i$, and $\ell_1 < \ell_2 < \ell_3$. We see that the determinant for ordered $\vec{u}_{\ell,k}$ is given by a sum of determinant for ordered $\vec{v}_{\ell,k}$ with mixed signs, and thus the positivity of the latter does not imply that for the former.

Amazingly, explicit evaluations of eq.(3.197) reveal that the determinant is positive so long as $\{\ell_i\}$s are larger than some critical spin! That is, above some critical spin, ℓ_c,

$$\mathrm{Det}[\{\vec{u}_{\ell_1,k}, \vec{u}_{\ell_2,k}, \cdots\}] > 0, \quad \forall\ \ell_c \leq \ell_1 < \ell_2 < \cdots\,. \tag{3.198}$$

In other words the convex hull of Gegenbauer vectors above the critical spin yields a cyclic polytope.[6] For example, focusing on four-dimensions, we find the critical spin at different k given as:

k	2	3	4	5	6	7	8	9	10
ℓ_c	1	2	2	3	3	4	4	5	5

(3.199)

It is intriguing to understand how such positivity emerged. In the RHS of eq.(3.197), each term can be identified as a minor of the Gegenbauer matrix with half of the rows removed. Consider the ratio of the first term on the RHS of eq.(3.197), against the next three. The first term has the property that it retains only even Taylor expansion terms. We plot these ratios for spins $(\ell_1,\, \ell_2,\, \ell_3) = (1+n, 2+n, 3+n)$ in fig.(3.12). As we can see, the leading term is dominant to the others as we increase in spin. Thus even though the other determinants in eq.(3.197) may have negative coefficients, their contributions are overwhelmed by the leading term which leads to the observed positivity. In other words, the minors with all even (or odd depending on the dimensions) Taylor coefficients take the maximal value!

[6] A fun "historic" note, the authors actually first observed the positivity of the ordered determinants for $\vec{u}_{\ell,k}$, not $\vec{v}_{\ell}$.

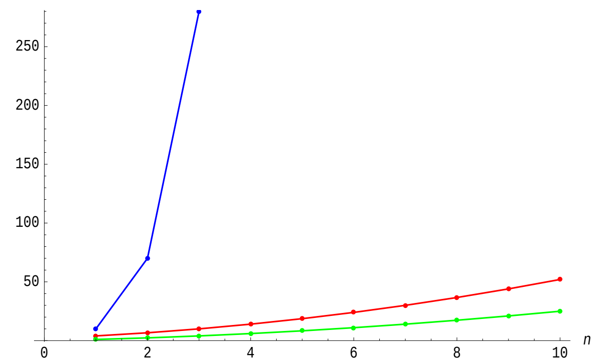


Figure 3.12: We take the ratio of the four determinants in the second and third line in eq.(3.197), denoted as $m_i(\ell_1, \ell_2, \ell_3)$, for $i = 1, \cdots, 4$. We plot $\frac{m_1(1+n,2+n,3+n)}{m_2(1+n,2+n,3+n)}$ (red), $\frac{m_1(1+n,2+n,3+n)}{m_3(1+n,2+n,3+n)}$ (blue), and $\frac{m_1(1+n,2+n,3+n)}{m_4(1+n,2+n,3+n)}$ (green), with $n = 1, \cdots$. As we can see, $m_1(\ell_1, \ell_2, \ell_3)$ is the largest and the ratio is an increasing function with spins.

The fact that $\vec{u}_{\ell,k}$ form a cyclic polytope above the critical spin indicates that for our $s{-}u$ polytope, most of the boundaries are known except for those involving spins below the critical spin, which can be computed straightforwardly. For coefficients that we can reliably bound, i.e. those proportional to z^n with $n \geq 2$. For k=2 ($D^4\phi^4$), we have

$$M_{D^4\phi^4} = (a_{2,0}z^2 + a_{2,2}t^2). \tag{3.200}$$

which simply gives us $a_{2,0} > 0$. For higher k, we have:

- k=3 : $D^6\phi^4$

$$M_{D^6\phi^4} = (a_{3,1}z^2 t + a_{3,3}t^3). \tag{3.201}$$

  Here we again have a single coefficient $a_{3,1}$ to bound. Since

$$u_{\ell,3,1} = \{-3, 1, 9, 21, ...\}, \tag{3.202}$$

  due to the first entry being negative, the positive span of these numbers will cover the whole real line, meaning we have *no bound* for the coefficient $a_{3,1}$.

- k=4 : $D^8\phi^4$

$$M_{D^8\phi^4} = (a_{4,0}z^4 + a_{4,2}z^2t^2 + a_{4,4}t^4). \tag{3.203}$$

  We can hope to bound $(a_{4,0}, a_{4,2})$. The $\vec{u}_{l,k}$ for each spin is

$$\begin{pmatrix} u_{\ell,4,0} \\ u_{\ell,4,2} \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ 27 \end{pmatrix}, \cdots. \tag{3.204}$$

Projectively these are points in $\mathbb{P}^1$, and the boundaries are given by the minimum and maximum value for the ratio $\frac{u_{*,4,2}}{u_{*,4,0}}$, which is given by $-\frac{3}{2}$ and ∞ respectively. Thus we simply have the bound:

$$\frac{a_{4,2}}{a_{4,0}} \geq -\frac{3}{2}\,. \tag{3.205}$$

- $k{=}5 : D^{10}\phi_4$

$$M_{D^{10}\phi_4}(s,t) = (a_{5,1}z^4t{+}a_{5,3}z^2t^3{+}\cdots). \tag{3.206}$$

where we've suppressed the couplings that we cannot bound. We would like to bound $(a_{5,1}, a_{5,3})$ and the space is $\mathbb{P}^1$. However, listing the relevant contributions from each spin

$$\frac{u_{\ell,5,3}}{u_{\ell,5,1}} = \left\{\frac{1}{2}, -\frac{7}{2}, -\frac{5}{14}, -\frac{33}{38}, \ldots\right\}, \tag{3.207}$$

we see that just as in the $k = 3$ case, the positive span will cover the entire $\mathbb{P}^1$, and thus the bound is trivial.

- $k{=}6 : D^{12}\phi_4$

$$M_{D^{12}\phi_4} = (a_{6,0}z^6{+}a_{6,2}z^4t^2{+}a_{6,4}z^2t^4{+}\cdots). \tag{3.208}$$

we can bound $\mathbf{a}_6{=}(a_{6,0}, a_{6,2}, a_{6,4})$ and the geometry is $\mathbb{P}^2$. The boundaries are given by:

$$\langle \mathbf{a}_6, 2, 1\rangle, \quad \langle \mathbf{a}_6, 1, 4\rangle, \quad \langle \mathbf{a}_6, i, i{+}1\rangle_{i\geq 4}, \quad \langle \mathbf{a}_6, \infty, 2\rangle. \tag{3.209}$$

We see that $Conv[\vec{u}_{\ell,6}]$ retains most of the boundaries of a cyclic polytope. Note that since the spin-0 and 3 vector are not involved with any boundary, they are inside the hull.[7]

Moving to higher-ks, in general there are no bounds for $k \in odd$, while for $k \in even$ we have the familiar cyclic polytope boundaries above a critical spin and a few additional boundaries involving spins below the critical spin.

Identical scalars: intersecting with the permutation symmetry plane

When the scalars are identical, the amplitude further respects permutation invariance, and at low energies will be given as a polynomial in $\sigma_2 = (s^2 + t^2{+}u^2)$ and $\sigma_3 = (s^3{+}t^3{+}u^3)$. This translates to the couplings $a_{k,q}$ living on the permutation plane $\mathbf{X}_{perm}$, defined through,

$$\mathbf{X}_{perm}: \quad M\left(z,t\right) = M\left(\frac{z}{2}{+}\frac{3t}{4}, -\frac{t}{2}{+}z\right)\,. \tag{3.210}$$

[7]Here, the critical spin is 4 instead of 3 as listed in table 3.199. This is because here we are only keeping the first three components of $\vec{u_{\ell,6}}$, i.e. $u_{\ell,6,0}, u_{\ell,6,2}$, and $u_{\ell,6,4}$.

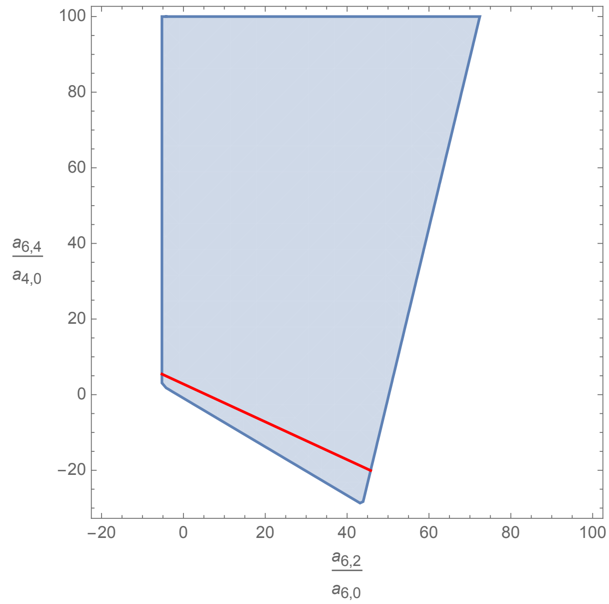


Figure 3.13: The space of allowed $(\frac{a_{6,2}}{a_{6,0}}, \frac{a_{6,4}}{a_{6,0}})$. The shaded region is carved out by the unitary polygon, whose boundary is composed of $(\ell, \ell+2)$ with $\ell \geq 2$, and $(2, \infty)$. Note that spin-0 is inside the hull and thus not part of the boundary. Finally the red-line represents the intersection of the permutation "line" $\mathbf{X}_{perm}$ and the unitary polygon.

Thus the geometry of interest is the intersection between $\mathbf{X}_{perm}$ and the unitary polytope, where the later is now constructed from *even spins* only. The dimensionality of $\mathbf{X}_{perm}$ is the number of independent polynomials built from σ_3 and σ_2. For $k = 2, 4$ the polynomial is unique, and the first place where there are two possibilities is $k = 6$: σ_3^2 and σ_2^3. On $\mathbf{X}_{perm}$ the couplings are parameterize as:

$$\begin{pmatrix} a_{2,0} & a_{2,2} & & & \\ a_{4,0} & a_{4,2} & a_{4,4} & & \\ a_{6,0} & a_{6,2} & a_{6,4} & a_{6,6} & \\ a_{8,0} & a_{8,2} & a_{8,4} & a_{8,6} & a_{8,8} \end{pmatrix} \to \begin{pmatrix} e_2 & \frac{3}{4}e_2 & & & \\ e_4 & \frac{3}{2}e_4 & \frac{9}{16}e_4 & & \\ e_6 & f_6 & \frac{45}{16}e_6 - \frac{1}{2}f_6 & \frac{9}{32}e_6 + \frac{1}{16}f_6 & \\ e_8 & f_8 & \frac{21}{8}e_8 + \frac{1}{4}f_8 & \frac{21}{8}e_8 - \frac{5}{16}f_8 & \frac{45}{256}e_8 + \frac{3}{64}f_8 \end{pmatrix}. \tag{3.211}$$

For $k = 2, 4$ we simply have the bound $e_2, e_4 > 0$. At $k = 6, 8$, the boundaries bound the ratio $\frac{f}{e}$ to be:

$$k = 6: \quad -\frac{21}{4} < \frac{f_6}{e_6} < \frac{183}{4}, \quad k = 8: \quad -8 < \frac{f_8}{e_8} < \frac{223}{4}. \tag{3.212}$$

In fig.3.13 we display the intersection geometry in $\mathbb{P}^2$ for $k = 6$.

These can be explicitly checked against the spinor-bracket stripped type-II closed string amplitude:

$$\frac{\Gamma[-s]\Gamma[-u]\Gamma[-t]}{\Gamma[1+s]\Gamma[1+u]\Gamma[1+t]}\,. \tag{3.213}$$

We can then identify:

$$k=6:\quad \frac{f_6}{e_6}=\frac{(8\zeta_3^3+31\zeta_9)}{12\zeta_9}=3.73895,\quad k=8:\quad \frac{f_8}{e_8}=\frac{2(2\zeta_{11}+\zeta_5\zeta_3^2)}{\zeta_{11}}=6.99512\,. \tag{3.214}$$

We see that it indeed resides in the bounds given by eq.(3.212).

Deformed moment curves and the EFT-hedron

We've seen the k-dependence of $\vec{u}_{\ell,k}$ leads to a deformation of the cyclic polytope discussed in the s-channel geometry. Now we would like to see how such mixing modifies the Hankel constraints, and the EFT-hedron.

Deformed moment curves

Let's again collect the coefficient with different ks and fixed q in to a column vector:

$$\begin{pmatrix} a_{2,q} \\ a_{4,q} \\ a_{6,q} \\ \cdots \\ a_{k,q} \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} u_{\ell_a,2,q} \\ u_{\ell_a,4,q}\, x_a \\ u_{\ell_a,6,q}\, x_a^2 \\ \cdots \\ u_{\ell_a,k,q}\, x_a^{\frac{k-2}{2}} \end{pmatrix}. \tag{3.215}$$

For $q=0$ as $u_{\ell,k,0}=v_{\ell,0}\geq 0$, the vectors on the RHS are just points on a moment curve multiplied by an overall positive factor and the usual Hankel matrix constraint applies. For $q\neq 0$, the k dependence of $u_{\ell,k,q}$ spoils this overall proportionality. This leads us to consider a generalization of moment curves: given a set of distinct positive factors α_i, we define a *deformed* moment curve $(1,x,\alpha_1 x^2,\cdots,\alpha_{n-1}x^n)$. Note that the convex hull of such deformed moment curve can be straightforwardly carved out by the total positivity of the *rescaled* Hankel matrix:

$$\begin{pmatrix} a_{4,q} & a_{6,q} \\ a_{6,q} & \frac{a_{8,q}}{\alpha_1} \end{pmatrix},\quad \begin{pmatrix} a_{6,q} & \frac{a_{8,q}}{\alpha_1} \\ \frac{a_{8,q}}{\alpha_1} & \frac{a_{10,q}}{\alpha_2} \end{pmatrix},\quad \begin{pmatrix} a_{4,q} & a_{6,q} & \frac{a_{8,q}}{\alpha_1} \\ a_{6,q} & \frac{a_{8,q}}{\alpha_1} & \frac{a_{10,q}}{\alpha_2} \\ \frac{a_{8,q}}{\alpha_1} & \frac{a_{10,q}}{\alpha_2} & \frac{a_{12,q}}{\alpha_3} \end{pmatrix}\cdots. \tag{3.216}$$

However, this is not sufficient to describe eq.(3.215) for two reasons: 1.) while each vector on the RHS of eq.(3.215) is a point on a rescaled moment curve, the scaling factors are *distinct* for different spins, and 2.) the rescaled factor $u_{\ell,k,q}$ is not necessarily positive.

Let's instead collect the different qs into row vectors $\vec{u}_{\ell,k}$ and $\vec{a}_k$, and rewrite eq.(3.215) as:

$$\begin{pmatrix} \vec{a}_2 \\ \vec{a}_4 \\ \vec{a}_6 \\ \cdots \\ \vec{a}_k \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} \vec{u}_{\ell_a,2} \\ \vec{u}_{\ell_a,4}\, x_a \\ \vec{u}_{\ell_a,6}\, x_a^2 \\ \cdots \\ \vec{u}_{\ell_a,k}\, x_a^{\frac{k-2}{2}} \end{pmatrix} . \tag{3.217}$$

Here each vector $\vec{u}_{\ell,k}$ will be of the same dimension. Now denote the boundaries of $Conv[\vec{u}_{\ell,k}]$ as $\vec{\mathcal{W}}_I^k$. The inner product $(\vec{u}_{\ell,k} \cdot \vec{\mathcal{W}}_I^{k'})$ by construction will give a positive factor when $k = k'$, but no longer guaranteed for $k' \neq k$. If we find some wall such that $(\vec{u}_{\ell,k} \cdot \vec{\mathcal{W}}_I)$ is always positive, then we are in business. Thus the task at hand is to find the boundary for $Conv[\vec{u}_{\ell,2}, \vec{u}_{\ell,4}, \cdots]$, i.e. we will be interested in the boundary of the Minkowski sum. Remarkably, numerical analysis so far has shown that the boundaries of $Conv[\vec{u}_{\ell,2}, \vec{u}_{\ell,4}, \cdots]$ are simply that of the highest k.

$$Conv[\vec{u}_{\ell,k_1}] \subset Conv[\vec{u}_{\ell,k_2}], \quad \forall k_1 < k_2 \,, \tag{3.218}$$

so in other words the inner product of $\vec{u}_{\ell,k}$ with $\vec{\mathcal{W}}_I^{k'}$ is guaranteed to be positive for $k \geq k'$.

Let us take eq.(3.217) and dotted into the boundaries of the highest k:

$$\begin{pmatrix} \vec{a}_2 \cdot \vec{\mathcal{W}}_I^k \\ \vec{a}_4 \cdot \vec{\mathcal{W}}_I^k \\ \vec{a}_6 \cdot \vec{\mathcal{W}}_I^k \\ \cdots \\ \vec{a}_k \cdot \vec{\mathcal{W}}_I^k \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} (\vec{u}_{\ell_a,2} \cdot \vec{\mathcal{W}}_I^k) \\ (\vec{u}_{\ell_a,4} \cdot \vec{\mathcal{W}}_I^k)\, x_a \\ (\vec{u}_{\ell_a,6} \cdot \vec{\mathcal{W}}_I^k)\, x_a^2 \\ \cdots \\ (\vec{u}_{\ell_a,k} \cdot \vec{\mathcal{W}}_I^k)\, x_a^{\frac{k-2}{2}} \end{pmatrix} . \tag{3.219}$$

Since by construction $\vec{u}_{\ell,k} \cdot \vec{\mathcal{W}}_I^k \geq 0$, the RHS gives a sum over points on a set of deformed moment curves, with the deformation parameters given as $\{\vec{\alpha}_\ell\} = \{\vec{u}_{\ell,2} \cdot \vec{\mathcal{W}}_I^k, \vec{u}_{\ell,4} \cdot \vec{\mathcal{W}}_I^k, \cdots\}$. Note that the $\{\vec{\alpha}_\ell\}$s are distinct for each spin.

Now we have arrived at a well posed positive geometry: the convex hull of an infinite number of deformed half moment curves. To proceed we will construct a "*principle deformed curve*" such that the deformed curves defined by $\{\vec{\alpha}_\ell\}$ reside in the hull of the former, i.e. we will like to find a set of parameters $\{\tilde{\alpha}_i\}$ that defines a deformed moment curve whose convex hull encapsulates the RHS of eq.(3.219) for all ℓ. Note that since $\{\vec{\alpha}_\ell\}$ depends on the boundary $\vec{\mathcal{W}}_I^k$, so will $\{\tilde{\alpha}_i\}$. Let us see how this works in practice.

- **k=6**: Beginning with eq.(3.219) and setting $k=6$, we would like to find a deformed moment curve
$$(1, x, \alpha_1 x^2) \tag{3.220}$$
such that the RHS of eq.(3.219) lies inside its convex hull. Since the being inside its hall translates to total positivity of the deformed Hankel matrix, we conclude that we need to find α_1 such that
$$\begin{pmatrix} \vec{u}_{\ell,2}\cdot\vec{\mathcal{W}}^6_I & \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^6_I \\ \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^6_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^6_I}{\tilde{\alpha}_1} \end{pmatrix}, \tag{3.221}$$
is totally positive for all ℓ, or
$$\frac{(\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^6_I)(\vec{u}_{\ell,2}\cdot\vec{\mathcal{W}}^6_I)}{(\vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^6_I)^2}\geq\tilde{\alpha}_1, \quad \forall\ell\,, \tag{3.222}$$
Thus there is a maximal value for $\tilde{\alpha}_1$ corresponding to the minimal value of the RHS of the above. Importantly, since some of the vectors $\vec{u}_{\ell,6}$ will inevitably be on the boundary $\vec{\mathcal{W}}^6_I$, the upper bound for $\tilde{\alpha}_1$ is actually zero! To this end, it will be natural to consider boundaries that are *outside* of $Conv[\vec{u}_{\ell,6}]$, which we will denote as $\vec{\mathcal{W}}^{6'}_I\equiv\vec{\mathcal{W}}^6_I+\Delta w$. The value for $\tilde{\alpha}_i$ now becomes Δw dependent.

- **k=8**: taking $k=8$ on the RHS of eq.(3.219) for fixed $\vec{\mathcal{W}}^8_I$, the independent positivity constraint will be the total positivity of
$$\begin{pmatrix} \vec{u}_{\ell,2}\cdot\vec{\mathcal{W}}^{8'}_I & \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{8'}_I \\ \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{8'}_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{8'}_I}{\tilde{\alpha}_1} \end{pmatrix}, \quad \begin{pmatrix} \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{8'}_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{8'}_I}{\tilde{\alpha}_1} \\ \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{8'}_I}{\tilde{\alpha}_1} & \frac{\vec{u}_{\ell,8}\cdot\vec{\mathcal{W}}^{8'}_I}{\tilde{\alpha}_2} \end{pmatrix}, \tag{3.223}$$
where once again $\vec{\mathcal{W}}^{8'}_I=\vec{\mathcal{W}}^8_I+\Delta w$. To find a set of suitable $(\tilde{\alpha}_1,\tilde{\alpha}_2)$, we first solve total positivity for the first matrix to determine $\tilde{\alpha}_1$, and use the result to solve the second matrix to determine $\tilde{\alpha}_2$.

For general k one iteratively solves the $\tilde{\alpha}_i$ in sequence. As a final example, for $k=10$ we simply iteratively solve total positivity of the following three matrices
$$\begin{pmatrix} \vec{u}_{\ell,2}\cdot\vec{\mathcal{W}}^{10'}_I & \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{10}_I \\ \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{10'}_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_1} \end{pmatrix}, \quad \begin{pmatrix} \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{10'}_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_1} \\ \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_1} & \frac{\vec{u}_{\ell,8}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_2} \end{pmatrix}, \quad \begin{pmatrix} \vec{u}_{\ell,2}\cdot\vec{\mathcal{W}}^{10'}_I & \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{10'}_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_1} \\ \vec{u}_{\ell,4}\cdot\vec{\mathcal{W}}^{10'}_I & \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_1} & \frac{\vec{u}_{\ell,8}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_2} \\ \frac{\vec{u}_{\ell,6}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_1} & \frac{\vec{u}_{\ell,8}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_2} & \frac{\vec{u}_{\ell,10}\cdot\vec{\mathcal{W}}^{10'}_I}{\tilde{\alpha}_3} \end{pmatrix}. \tag{3.224}$$
In all cases, we need to choose a deformed boundary $\vec{\mathcal{W}}^{k'}_I=\vec{\mathcal{W}}^k_I+\Delta w$.

The EFT-hedron

We now turn to the full EFT-hedron. Again begin with

$$\vec{A}_I = \begin{pmatrix} A_{2,I} \\ A_{4,I} \\ \cdots \\ A_{k,I} \end{pmatrix} = \begin{pmatrix} \vec{a}_2 \cdot \vec{\mathcal{W}}_I \\ \vec{a}_4 \cdot \vec{\mathcal{W}}_I \\ \cdots \\ \vec{a}_k \cdot \vec{\mathcal{W}}_I \end{pmatrix}, \tag{3.225}$$

where we've taken k to be even. Firstly $A_{k,I}$ is positive, whenever $\vec{\mathcal{W}}_I$ is one of the facets of $Conv[\vec{u}_{\ell,k}]$. Furthermore we require total positivity of the deformed Hankel matrix of $\vec{A}_I$, given as

$$\begin{pmatrix} A_{2,I} & A_{4,I} \\ A_{4,I} & \frac{A_{6,I}}{\tilde{\alpha}_1} \end{pmatrix}, \quad \begin{pmatrix} A_{4,I} & \frac{A_{6,I}}{\tilde{\alpha}_1} \\ \frac{A_{6,I}}{\tilde{\alpha}_1} & \frac{A_{8,I}}{\tilde{\alpha}_2} \end{pmatrix}, \quad \begin{pmatrix} A_{2,I} & A_{4,I} & \frac{A_{6,I}}{\tilde{\alpha}_1} \\ A_{4,I} & \frac{A_{6,I}}{\tilde{\alpha}_1} & \frac{A_{8,I}}{\tilde{\alpha}_2} \\ \frac{A_{6,I}}{\tilde{\alpha}_1} & \frac{A_{8,I}}{\tilde{\alpha}_2} & \frac{A_{10,I}}{\tilde{\alpha}_3} \end{pmatrix}, \ e.t.c. \tag{3.226}$$

where $\vec{\mathcal{W}}_I$ is now the deformed boundary of maximal k, $\vec{\mathcal{W}}_I^{k'}$, and the deformation parameters $\{\tilde{\alpha}_i\}$s defined through the total positivity of eq.(3.224). These two constraints are encapsulated as:

$$\boxed{K[\vec{A}_I]_{\{\tilde{\alpha}_i\}} \quad \text{is a totally positive matrix}} \ . \tag{3.227}$$

Let us compare side by side the s-channel EFT-hedron and the general EFT-hedron: starting with $\vec{A}_I$ given in eq.(3.225), they are defined by:

	s-ch EFT-hedron	EFT-hedron
Hankel matrix	Canonical $K[X]$	Deformed $K[X]_{\{\tilde{\alpha}_i\}}$
$\mathcal{W}_I$	boundaries of $Conv[\vec{v}_\ell]$	boundaries of $Conv[\vec{u}_{\ell,k}]$

In the following we will consider the $\mathbb{P}^1$ geometry. **Example:**

Let's consider the explicit example for $k = 4, 6, 8$, where

$$\begin{pmatrix} a_{4,0} & a_{4,2} \\ a_{6,0} & a_{6,2} \\ a_{8,0} & a_{8,2} \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} x_a^4 \vec{u}_{\ell_a,4} \\ x_a^6 \vec{u}_{\ell_a,6} \\ x_a^8 \vec{u}_{\ell_a,8} \end{pmatrix} \qquad \vec{u}_{\ell,k} = (u_{\ell,k,0}, u_{\ell,k,2}), \tag{3.228}$$

Since $u_{\ell,k,0}$ is positive for all ℓ, k, we can use it to positively rescale the first entry to 1 and define $u_\ell^{(k)} = \frac{u_{\ell,k,2}}{u_{\ell,k,0}}$. Then $Conv[\vec{u}_{\ell,k}]$ is simply a line segment in $\mathbb{P}^1$ with its boundary determined by the minimum value of $u_\ell^{(k)}$. From eq.(3.194) one can check that the minimum value of $u_\ell^{(k)}$ for fixed k and arbitrary spin is given as:

$$Min\left[u_\ell^{(4)}\right] = -\frac{3}{2} \ \ (\ell = 1, 2), \quad Min\left[u_\ell^{(6)}\right] = -\frac{21}{4} \ \ (\ell = 2), \quad Min\left[u_\ell^{(8)}\right] = -8 \ \ (\ell = 2)\ . \tag{3.229}$$

Note the above agrees with eq.(3.218), which states that the boundary of the Minkowski sum is given by that of the largest k, here 8. Rescaling $(a_{k,0}, a_{k,2}) = a_k\,(1, \beta_k)$, the above tells us that the boundaries of $Conv[\vec{u}_{\ell,k}]$ for each k translate to

$$a_4 \geq 0, \quad a_6 \geq 0, \quad a_8 \geq 0, \quad \beta_4 \geq -\frac{3}{2}, \quad \beta_6 \geq -\frac{21}{4} \quad, \beta_8 \geq -8\,. \tag{3.230}$$

Furthermore, we also have that $a_{k,0}$ is inside the convex hull of the half-moment curve:

$$a_6^2 - a_4 a_8 \geq 0\,. \tag{3.231}$$

These inequalities correspond to $A_{4,I}, A_{6,I}, A_{8,I}$ being positive with $\mathcal{W}_I$ are chosen to be the boundary of $Conv[\vec{u}_{\ell,k}]$, and

$$\begin{pmatrix} A_{4,I} & A_{6,I} \\ A_{6,I} & \frac{A_{8,I}}{\tilde{\alpha}_1} \end{pmatrix} \tag{3.232}$$

being totally positive, where $\mathcal{W}_I = (1,0)$ and $\tilde{\alpha}_1 = 1$.

Next, we consider the positivity of Det$[eq.(3.232)]$ where $\mathcal{W}_I$ is the boundary of the Minkowski sum. Since the boundary of $Conv[\vec{u}_{\ell,4}, \vec{u}_{\ell,6}, \vec{u}_{\ell,8}]$ is given by $(1,-8)$, the upper bound for $\tilde{\alpha}_1$ is such that

$$\frac{(u_\ell^{(4)} + 8 + \Delta w)(u_\ell^{(8)} + 8 + \Delta w)}{\tilde{\alpha}_1} - \left(u_\ell^{(6)} + 8 + \Delta w\right)^2 \geq 0, \quad \forall \ell\,. \tag{3.233}$$

Note that we have add a small deformation Δw. This is needed since here $\mathcal{W}_I$ is identified with $u_2^{(6)}$, which would cause the first term in the above (with $\Delta w = 0$) to be zero for $\ell = 2$ and invalidate the inequality. Picking $\Delta w = \frac{1}{100}$ we find $\tilde{\alpha}_1 \leq 0.0085$. Equipped with this the positivity of the determinant eq.(3.232) translates to

$$\frac{(\beta_4 + 8 + \frac{1}{100})(\beta_8 + 8 + \frac{1}{100})}{0.0085} - \left(\beta_6 + 8 + \frac{1}{100}\right)^2 \geq 0\,. \tag{3.234}$$

Note that in the above it is necessary to consider walls that are deformed away from the boundary of $Conv[\vec{u}_{\ell,4}, \vec{u}_{\ell,6}, \vec{u}_{\ell,8}]$, and $\tilde{\alpha}_1$ as well as that the non-linear constraint that follows depends on the choice of deformation parameter Δw. As we will see in appendix B.6, the most stringent non-linear constraint does not necessarily correspond to Δw being small! In other words, the true boundary of the EFT-hedron is actually defined by a new wall that can be far from the boundaries of the cyclic polytope. A more complete understanding of the true boundaries will be left to future studies.

When the external particles are identical, we should consider even spins only. However, since the minimum in (3.229) is given by spin-2, the optimal value for $\tilde{\alpha}_1$ remains the same. Thus the problem simply reduces to the intersection of the permutation plane defined in (3.210) with our $\mathbb{P}^1$ geometry. From (3.211), we see that $\beta_4 = \frac{a_{4,2}}{a_{4,0}}$ is fixed to $\frac{3}{2}$. This turns (3.234) into a quadratic bound for β_6 and β_8. Thus for identical scalars, the EFT-hedron bounds are given by eq.(3.230) and

$$(\frac{19}{2}+\frac{1}{100})(\beta_8+8+\frac{1}{100})-0.0085\left(\beta_6+8+\frac{1}{100}\right)^2 \geq 0\,. \tag{3.235}$$

Multiple Species

Let us now return to the scattering of a, b, but now consider the amplitude $M(a, b, b, a)$ in combination with all a and all b scattering. For simplicity we will assume each of a, b have a $\mathbb{Z}_2$ symmetry, so the only non-vanishing amplitude involves even number of a's and b's. Now we can get constraints mixing the a^4, a^2b^2 and b^4 amplitudes, if we consider $ABBA$ scattering of general states $A = \alpha a+\beta b$,and $B = \gamma a+\rho b$. These must satisfy the EFT-constraints for all $(\alpha, \beta, \gamma, \rho)$; in the special case of $A = B$ ($\alpha = \gamma$, $\beta = \rho$) we intersect with the crossing symmetry plane as well. A systematic exploration of the geometry associated with this envelope of constraints is left for future work, but it is easy and illuminating to look at the simplest example.

Consider the leading 4-derivative amplitudes

$$M(a^4) = c_a(s^2{+}t^2{+}u^2),\ M(b^4) = c_b(s^2{+}t^2{+}u^2),\ M(abba) = c(s^2{+}u^2){+}\frac{d}{2}t^2\,. \tag{3.236}$$

Note our analysis of $M(abba)$ just tells us that $c > 0$; d can have any sign. But we will now see that magnitude of d is bounded by $c_{a,b}$ as

$$c_a c_b - d^2 > 0\,. \tag{3.237}$$

To whit, the amplitude for $M(ABBA)$ is given by

$$\begin{aligned} M(ABBA) \;&=\; (\alpha\gamma)^2 M(a^4){+}(\beta\rho)^2 M(b^4){+}(\gamma\beta)^2 M(baab){+}(\alpha\rho)^2 M(abba) \\ &+\; (\alpha\beta\gamma\rho)\left[M(aabb){+}M(baba){+}M(abab){+}M(bbaa)\right]. \end{aligned} \tag{3.238}$$

Note that while the term proportional to d in $M(abba)$ drops out in the forward limit as $t \to 0$, this is not the case e.g. for $M(aabb) = c(u^2{+}t^2){+}\frac{d}{2}s^2$ which becomes $s^2(c + d/2)$ in the forward limit.

Taking the $t \to 0$ limit, the coefficient of s^2 in the $M(ABBA)$ amplitude, which must be positive, is given by

$$\begin{aligned} &(\alpha\gamma)^2 c_a{+}(\beta\rho)^2 c_b{+}(\alpha\beta\gamma\rho)(2d{+}4c){+}2c((\gamma\beta)^2{+}(\alpha\rho)^2) \\ &\quad = (\alpha\gamma)^2 c_a{+}(\beta\rho)^2 c_b{+}2d(\alpha\beta\gamma\rho){+}2c(\gamma\beta{+}\alpha\rho)^2\,. \end{aligned} \tag{3.239}$$

Now of course if we put $\alpha = 1, \beta = 0, \gamma = 0, \rho = 1$, we go back to $A = a$, $B = b$, and we learn that $c > 0$. But now let's put $\gamma\beta + \alpha\rho = 0$. We then have $x^2 c_a + y^2 c_b + 2xyd > 0$, where $x = -\alpha^2\rho/\beta$, $y = \beta\rho$; note that varying over α, β, ρ, (x, y) can be any real numbers. Thus we learn that $c_{a,b} > 0$ and $c_a c_b - d^2 > 0$, or the positivity of the matrix in

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} c_a & d \\ d & c_b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} . \tag{3.240}$$

Note it was important in this analysis to allow general AB states; had we taken only $A = B \to \alpha = \gamma, \beta = \rho$, we would find no constraints on $d > 0$.

This can be straightforwardly generalized to any number of species labelled by the index i. Again assuming $\mathbb{Z}_2$ symmetry for each species, writing

$$M(i^4) = c_i(s^2 + t^2 + u^2),\ M(ijji) = c_{ij}(s^2 + u^2) + d_{ij}t^2 , \tag{3.241}$$

we find that $c_{ij} \geq 0$, and that the matrix

$$\begin{pmatrix} c_{11} & d_{12} & d_{13} & \cdots \\ d_{12} & c_{22} & d_{23} & \cdots \\ d_{13} & d_{23} & c_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} , \tag{3.242}$$

is positive. The positivity of a symmetric matrix S is equivalent to the positivity of all the leading principle minors (determinant of all upper left square matrices) of the matrix (the Sylvester's criterion). As an example we have

$$det \begin{pmatrix} c_{11} & d_{12} & d_{13} \\ d_{12} & c_{22} & d_{23} \\ d_{13} & d_{23} & c_{33} \end{pmatrix} \geq 0, \quad det \begin{pmatrix} c_{11} & d_{12} \\ d_{12} & c_{22} \end{pmatrix} \geq 0, \quad c_{11} \geq 0 . \tag{3.243}$$

3.9 The spinning EFT-hedron

So far we have examined constraints on amplitudes with external scalars. The analysis can be readily extended to external spinning states such as gluons, photons, and gravitons, where the higher dimensional operators of the EFT will be given in terms of field strengths, Riemann tensors, and derivatives thereof. In subsection 3.6 we've seen that the Taylor vectors of spinning polynomials also generate cyclic polytopes, and thus we can simply retrace all of the previous discussion, with $v_{\ell,q}$ replaced by the Taylor coefficient of the spinning polynomials.

An important question is which helicity configuration should one select for the dispersive representation. The choice should be such that one is expanding around a forward process, i.e. the $t \to 0$ limit corresponds to $a, b \to b, a$ scattering. Take for example $M(1^+, 2^-, 3^+, 4^-)$. In the s-channel threshold where $1, 2$ are incoming and $3, 4$ outgoing, the process corresponds to $1^+2^- \to 3^-4^+$. Note that the helicity of legs 3 and 4 are flipped since we've defined the helicity for M with all momenta incoming. For it to be forward, we should identify the state on leg 1 with 4, so we set $p_4 = p_1$ and $p_2 = p_3$ which indeed corresponds to $t = 0$. For the u-channel threshold one instead has $1^+3^+ \to 2^+4^+$, which once again correspond to a forward process with $p_4 = p_1$ and $p_2 = p_3$. Similarly $M(1^+, 2^+, 3^-, 4^-)$ also admits a positive expansion. This is in contrast with $M(1^+, 2^-, 3^-, 4^+)$, where in the s-channel we have $1^+2^- \to 3^+4^-$. In order for this to be forward, we need to take $p_1 = p_3$ and $p_2 = p_4$ which corresponds to $u = 0$ instead of $t = 0$. So in this case the small t expansion of the residue is not an expansion around a forward process, and does not enjoy the positivity properties we wish to exploit.

As a simple example, the s-channel EFT hedron can be generalized to color ordered states. From the previous discussion, we've seen that expanding in t for $M(1^+, 2^-, 3^+, 4^-)$ corresponds to an expansion around the forward limit. Thus the s-channel residue can be positively expanded on $d^\ell_{2,2}(\theta)$ (see eq.(3.140))

$$Res_s[M(1^+, 2^-, 3^+, 4^-)] = \sum_\ell \mathsf{p}_\ell d^\ell_{2,2}(\theta) \quad \mathsf{p}_\ell \geq 0. \tag{3.244}$$

Removing the overall spinor bracket mandated by the helicity weights, we have:

$$\langle 24\rangle^2[13]^2 \left(\sum_{k,q} a_{k,q} s^{k-q} t^q\right) = -\langle 24\rangle^2[13]^2 \left(\sum_a \mathsf{p}_{\ell_a} \frac{\tilde{d}^{\ell_a}_{2,2}(\theta)}{s - m_a^2}\right)\Bigg|_{\theta=\arccos(1+2t/m_a^2)} \tag{3.245}$$

where once again the equality is understood in terms of Taylor expansion, and $\tilde{d}^{\ell_a}_{2,2}(\theta) = \frac{d^{\ell_a}_{2,2}(\theta)}{\cos^4\frac{\theta}{2}}$. We can then bound operators using the boundaries of the cyclic polytopes, as an example, for $k = 2$, which corresponds to D^4F^4, we have

$$\langle \mathbf{a}_2, \ell, \ell+1\rangle \geq 0, \quad \mathbf{a}_2 = (a_{2,0}, a_{2,1}, a_{2,2}). \tag{3.246}$$

The two-dimensional region is then given in fig.3.14. Imposing cyclic symmetry sets $a_{2,2}/a_{2,0} = 1$ and the region becomes a one-dimensional line, and the bound becomes

$$0 \leq a_{2,1}/a_{2,0} \leq \frac{9}{5}. \tag{3.247}$$

For open super-string, we have $\frac{a_{2,1}}{a_{2,0}} = \frac{1}{4}$ and are thus inside the bound.

For photons and gravitons, we need to consider the contributions from both s and u-channel. Here we choose the amplitude $M(1^{+h}, 2^{+h}, 3^{-h}, 4^{-h})$, and the s-channel residue

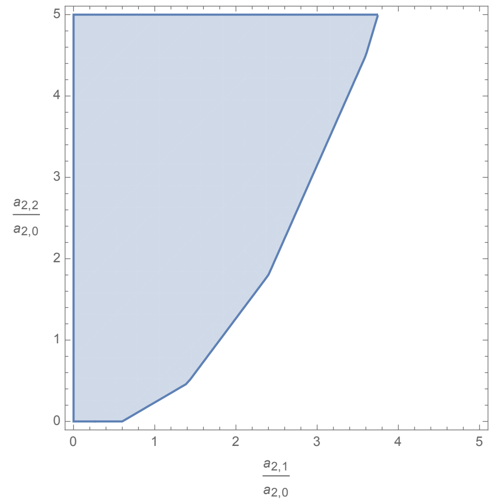


Figure 3.14: The $k = 2$ polygon for $(+-+-)$ gluon scattering.

for a spin-ℓ exchange is written as:

$$Res_s\left[M(1^{+h}, 2^{+h}, 3^{-h}, 4^{-h})\right] \;=\; g_\ell^{++} g_\ell^{--} [12]^{2h} \langle 34\rangle^{2h}\, d^\ell_{0,0}(\theta)\,, \tag{3.248}$$

where $g_\ell^{++/--}$ is the coupling constant of a real spin-ℓ state to a pair of plus/minus helicity photon. CPT requires $g_\ell^{++} = (g_\ell^{--})^*$, and the above yields a positive expansion as expected. Furthermore under 3, 4 exchange $d^\ell_{0,0}(\theta) \to d^\ell_{0,0}(-\theta) = (-)^\ell d^\ell_{0,0}(\theta)$, thus bose symmetry requires $\ell \in even$. The u-channel residue is given as:

$$Res_u\left[M(1^{+h}, 2^{+h}, 3^{-h}, 4^{-h})\right] \;=\; (g_\ell^{+-})^2 [12]^{2h} \langle 34\rangle^{2h} \tilde{d}^{\ell_a}_{2,2} \tag{3.249}$$

where now CPT simply requires g_ℓ^{+-} to be real. Thus we arrive at the following dispersive representation[8]

$$\boxed{[12]^{2h}\langle 34\rangle^{2h}\left(\sum_{k,q} a_{k,q} s^{k-q} t^q\right) = -[12]^{2h}\langle 34\rangle^{2h}\left(\sum_a \mathsf{p}_{\ell_a}\frac{d^{\ell_a}_{0,0}(\theta)}{s-m_a^2} + \sum_b \tilde{\mathsf{p}}_{\ell_b}\frac{\tilde{d}^{\ell_b}_{2,2}}{u-m_b^2}\right)}\,, \tag{3.250}$$

where p_{ℓ_a} and $\tilde{\mathsf{p}}_{\ell_b}$ are distinct positive coefficients and $\ell_a \in even$.

In the following, we will analyze external photons and gravitons separately. For $k = even$ the bounds are listed as:

(-h,-h,+h,+h):		photon	graviton
	$k = 2$	D^4F^4 (3.253)	D^4R^4 (3.259)
	$k = 4$	D^8F^4 (3.256)	D^8R^4 (3.262)

[8]The first version of this paper had an error in the residue polynomials in the spinning dispersion relation, which we correct here, modifying the obtained bounds. We thank Zvi Bern, Alexander Zhiboedov, and Dimitrios Kosmopoulos for pointing out this mistake to us.

Photon EFT

For photons, our analysis can be separated into whether or not gravity decouples. For EFTs whose gravitational dynamics are irrelevant, such as the Euler-Heisenberg theory, one can bound operators of degree 2 or higher in s. If gravity does not decouple, as discussed in sec.3.4 the forward limit graviton pole will obstruct any bound on s^2. In practice, starting with the geometry for gravitationally decoupled EFTs, one can incorporate gravity simply by projecting the geometry onto the directions perpendicular to $a_{k,k-2}$.[9]

Note that now the s- and u-channel have distinct polynomials, we will label the vectors from the s and u channel in eq.(3.250) as ℓ_s and ℓ_u respectively, and the unitary polytope is the Minkowski sum of the two polytopes. Furthermore, this helicity configuration is invariant under $t \leftrightarrow u$ exchange, and thus the amplitude must lie on the "symmetry plane" $\mathbf{X}_{sym}$ parameterized as:

$$\begin{pmatrix} a_{1,0} & a_{1,1} & & & \\ a_{2,0} & a_{2,1} & a_{2,2} & & \\ a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \\ a_{4,0} & a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} \rightarrow \begin{pmatrix} x & 0 & & & \\ x & y & y & & \\ x & y & y & 0 & \\ x & y & z & 2(z-y) & (z-y) \end{pmatrix}. \tag{3.251}$$

We now give the intersection of $\mathbf{X}_{sym}$ with the unitary polytope:

- k=2 : D^4F^4
$$M_{D^4F^4} = \langle 12\rangle^2[34]^2(a_{2,0}s^2 + a_{2,1}st + a_{2,2}t^2). \tag{3.252}$$
Now we would like to bound $\mathbf{a}_2 = (a_{2,0}, a_{2,1}, a_{2,2})$ which live in $\mathbb{P}^2$. The edge of the polygon is given by
$$\langle *, i_u{+}1, i_u, \rangle_{i_u\geq 2}, \quad \langle *, i_s, i_s{+}2\rangle_{i_s\geq 2}, \quad \langle *, 2_u, 2_s\rangle\,, \tag{3.253}$$
where i_s, i_u represents the Taylor vectors from $d^{i_s}_{0,0}$ and $d^{i_u}_{-2,-2}$ respectively. Note that the majority of the edges for the s- and u-channel cyclic polytope remains a facet for the Minkowski sum. The polygon is presented in projective coordinates $\left(\frac{a_{2,1}}{a_{2,0}}, \frac{a_{2,2}}{a_{2,0}}\right)$ in fig.3.15, where we've labeled the vertices from the (purple)s and (red)u channels explicitly.
On $\mathbf{X}_{sym}$ we have $\frac{a_{2,1}}{a_{2,0}} = \frac{a_{2,2}}{a_{2,0}}$ and the geometry reduces to $\mathbb{P}^1$. The region of intersection is given as:
$$-\frac{30}{7} \leq \frac{a_{2,1}}{a_{2,0}} = \frac{a_{2,2}}{a_{2,0}} \leq 6\,. \tag{3.254}$$

[9]We will assume that RF^2 is not relevant for the analysis, although it is straightforward to incorporate.

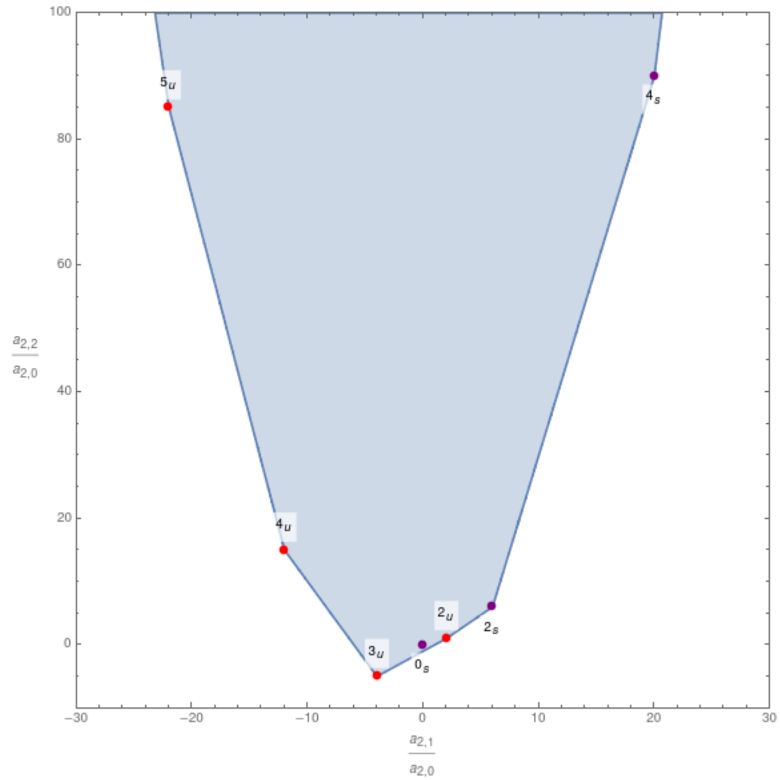


Figure 3.15: The $k = 2$ polygon for $(--++)$ photon scattering. It is bounded by the Minkowski sum of the vectors originated from the s-channel (purple dots) and u-channel (red dots).

Note that similar to the intersection of the scalar $s{-}u$ polytope with the permutation plane, here the intersection yields leads to EFT coefficients being bounded from both sides.

- $k = 4 : D^8F^4$

$$M_{D^8F^4} = \langle 12\rangle^2[34]^2(a_{4,0}s^4 + a_{4,1}s^3t + a_{4,2}s^2t^2 + a_{4,3}st^3 + a_{4,4}t^4). \tag{3.255}$$

The coupling $\mathbf{a}_4 = (a_{4,0}, a_{4,1}, a_{4,2}, a_{4,3}, a_{4,4})$ lives in $\mathbb{P}^4$, and is bounded by

$$\begin{aligned}
&\langle \mathbf{a}_4, 2_u, 3_u, 4_u, 5_u\rangle, \quad \langle \mathbf{a}_4, i_u, i_u{+}1, j_u, j_u{+}1\rangle_{i_u,j_u\geq 3},\\
&\langle \mathbf{a}_4, i_s, i_s{+}2, j_s, j_s{+}2\rangle_{i_s,j_s\geq 2}, \quad \langle \mathbf{a}_4, i_s{+}2, i_s, j_u, j_u{+}1\rangle_{i_s,\geq 4, j_u\geq 3},\\
&\langle \mathbf{a}_4, 4_s, 2_s, 3_u, 2_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 2_s, 2_u, 5_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 2_s, i_u, i_u + 1\rangle_{i_u\geq 5}\\
&\langle \mathbf{a}_4, i_s + 2, i_s, \infty_u, \infty_s\rangle_{i_s\geq 2}, \quad \langle \mathbf{a}_4, i_u, i_u + 1, \infty_u, \infty_s\rangle_{i_u\geq 3}\\
&\langle \mathbf{a}_4, 2_s, \infty_s, 3_u, \infty_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 2_u, 3_u, 4_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 2_u, 4_u, 5_u\rangle,\\
&\langle \mathbf{a}_4, i_s + 2, i_s, 2_s, 3_u\rangle_{i_s\geq 4}, \quad \langle \mathbf{a}_4, 2_s, 3_u, 2_u, 5_u\rangle_{i_s\geq 4}, \quad \langle \mathbf{a}_4, 2_s, 3_u, i_u, i_u + i\rangle_{i_u\geq 5},
\end{aligned} \tag{3.256}$$

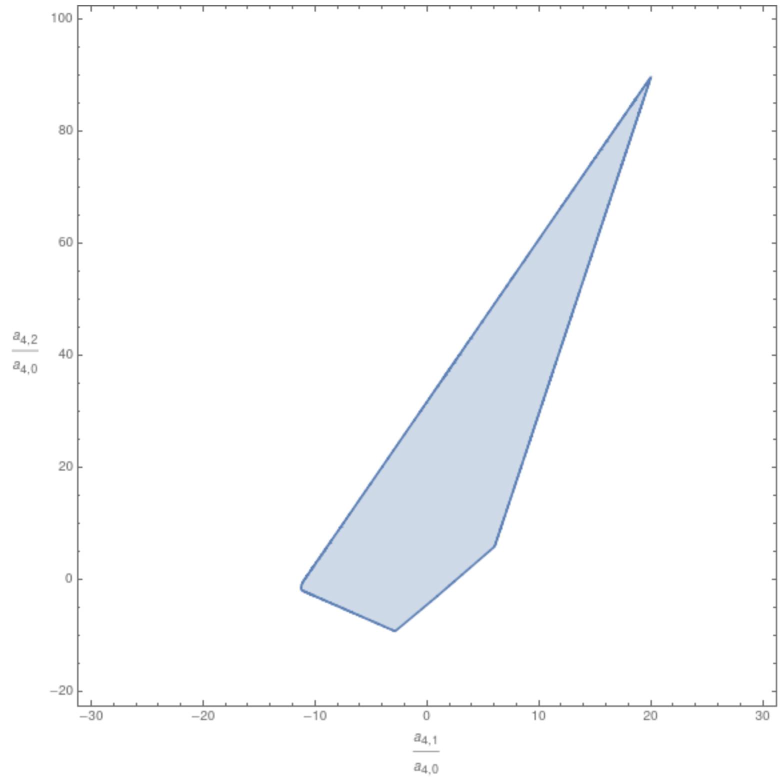


Figure 3.16: The intersection of the $\mathbb{P}^4$ polytope defined by the boundaries in eq.(3.256) with $\mathbf{X}_{sym}$.

being non-negative. Note that the boundary of the Minkowski sum consists of almost all the boundaries of the individual cyclic polytope, label by a pair of consecutive spins, as well as the tensor products of consecutive pair from both sides. At lower spin region we have some irregular boundaries as well. The intersection of the above with $\mathbf{X}_{sym}$ is illustrated in fig.3.16.

Graviton EFT

For gravity the analysis is a straightforward extension of the photon EFT: simply set $h = 2$ in the polynomial basis. From the discussion in sec.(3.4), we've seen that the tree-level four-graviton amplitude does not introduce any t-channel massless obstructions, and thus here we will be able to bound operators proportional to s^n with $n \geq 2$. Once again, we will consider the intersection of the unitary polytope with the symmetry plane $\mathbf{X}_{sym}$ defined in eq.(3.251):

- k=0 : R^4

$$M_{R^4} = \langle 12\rangle^4[34]^4 a_{0,0}. \tag{3.257}$$

and we simply have $a_{0,0} > 0$.

- $k{=}2 : D^4R^4$

$$M_{D^4R^4} = \langle 12\rangle^4[34]^4(a_{2,0}s^2 + a_{2,1}st + a_{2,2}t^2). \tag{3.258}$$

The facets are again given by that of the individual cyclic polytope in the s- and u-channel. The bounds are then given by:

$$\langle \mathbf{a}_2, i_u{+}1, i_u, \rangle_{i_u\geq 4}, \quad \langle \mathbf{a}_2, i_s, i_s{+}2\rangle_{i_s\geq 2}, \quad \langle \mathbf{a}_2, 4_u, 1_s\rangle\,. \tag{3.259}$$

being non-negative, with $\mathbf{a}_2 = (a_{2,0}, a_{2,1}, a_{2,2})$. On $\mathbf{X}_{sym}$ we have

$$-\frac{90}{11} \leq \frac{a_{2,1}}{a_{2,0}} = \frac{a_{2,2}}{a_{2,0}} \leq 6\,. \tag{3.260}$$

- $k{=}4 : D^8R^4$

$$M_{D^8R^4} = \langle 12\rangle^4[34]^4(a_{4,0}s^4 + a_{4,1}s^3t + a_{4,2}s^2t^2 + a_{4,3}st^3 + a_{4,4}t^4). \tag{3.261}$$

The facets are:

$$\begin{aligned}
&\langle \mathbf{a}_4, 4_u, 5_u, 6_u, 7_u\rangle, \quad \langle \mathbf{a}_4, i_u, i_u{+}1, j_u, j_u{+}1\rangle_{i_u,j_u\geq 5},\\
&\langle \mathbf{a}_4, i_s, i_s{+}2, j_s, j_s{+}2\rangle_{i_s,j_s\geq 2}, \quad \langle \mathbf{a}_4, i_s{+}2, i_s, i_u, i_u{+}1\rangle_{i_s\geq 4, i_u\geq 5},\\
&\langle \mathbf{a}_4, i_s{+}2, i_s, \infty_u, \infty_s\rangle_{i_s\geq 4},\\
&\langle \mathbf{a}_4, 4_s, 2_s, 6_s, 5_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 2_s, 5_u, 4_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 2_s, 4_u, 7_u\rangle,\\
&\langle \mathbf{a}_4, 4_s, 2_s, i_u, i_u{+}1\rangle_{i_u\geq 7}, \quad \langle \mathbf{a}_4, 4_s, 2_s, \infty_u, \infty_s\rangle,\\
&\langle \mathbf{a}_4, 2_s, 5_u, 4_u, 7_u\rangle, \quad \langle \mathbf{a}_4, 2_s, 5_u, i_u, i_u{+}1\rangle_{i_u\geq 7},\\
&\langle \mathbf{a}_4, 2_s, 5_u, \infty_u, \infty_s\rangle, \quad \langle \mathbf{a}_4, 2_s, 5_u, i_s{+}2, i_s\rangle_{i_s\geq 4},\\
&\langle \mathbf{a}_4, 4_s, 4_u, 5_u, 6_u\rangle, \quad \langle \mathbf{a}_4, 4_s, 4_u, 6_u, 7_u\rangle, \quad \langle \mathbf{a}_4, i_u, i_u{+}1, \infty_u, \infty_s\rangle_{i_u\geq 5}.
\end{aligned} \tag{3.262}$$

Once again, the facets maintain a cyclic structure at higher spins, while some irregularities occur at lower spin region. Its intersection with the symmetry plane $\mathbf{X}_{sym}$ is displayed in fig.3.17.

In this section we have focused for simplicity on the scattering of a single species—photons or gravitons—but it is easy to constrain photon-graviton couplings as well. The amplitude $M(1^{-1}2^{+2}3^{-2}4^{+1})$ is forward as $t \to 0$ in both the s- and u- channels, and so has a positive expansion. Thus considering the Gegenbauer constraints, the coefficients

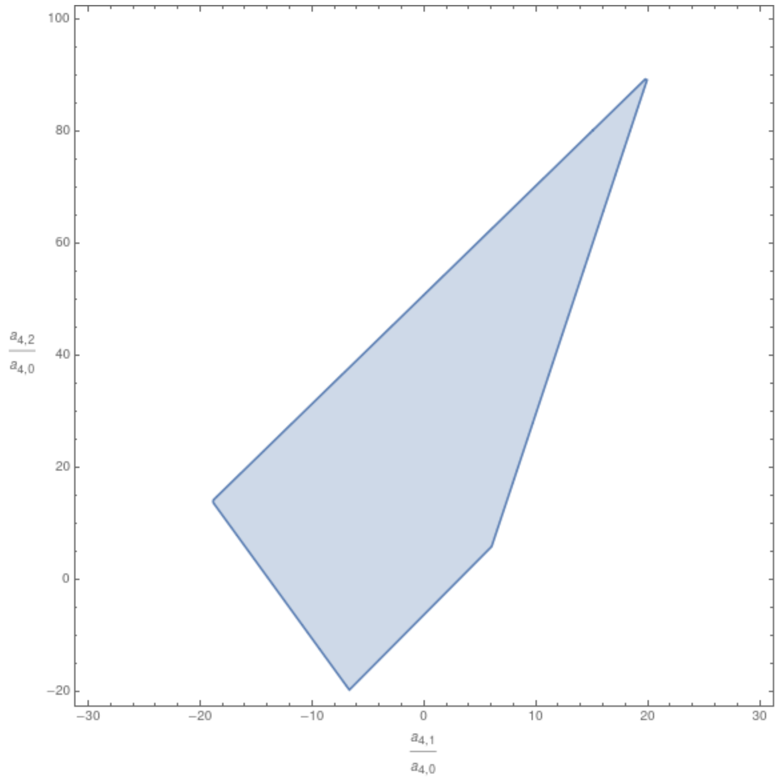


Figure 3.17: The intersection of the $\mathbb{P}^4$ polytope defined by the boundaries in eq.(3.262) with $\mathbf{X}_{sym}$.

must lie inside the unitarity polytopes; but we don't have the extra crossing symmetry constraints enjoyed by pure photon/graviton scattering. While this is all we can say considering only photon-graviton scattering, as with our multi-species discussion for the scalar case, there are clearly constraints relating the pure photon and pure graviton scattering coefficients to those of photon-graviton scattering, considering the scattering of general linear combinations of different species, which would be interesting to further explore.

3.10 Explicit EFTs in the EFT-hedron

So far we have been mostly discussing bounds on general EFTs, derived from the analyticity and unitarity in the UV. In this section we will discuss in more detail how realistic EFTs with explicit UV completions satisfy these bounds.

s-channel EFT-hedron

Let's begin with the s-channel constraints. We will use the tree-level massless open superstring amplitude as an example eq.(3.14), which we display again here

$$M(1^+2^-3^+4^-) = -\langle 24\rangle^2[13]^2\frac{\Gamma[-s]\Gamma[-t]}{\Gamma[1-s-t]} = \langle 24\rangle^2[13]^2\left[-\frac{1}{st}+\sum_{k,q}a_{k,q}s^{k-q}t^q\right]\,. \quad (3.263)$$

The coupling constants, up to $k=4$, are given as

$$\begin{pmatrix} a_{0,0} & & & & \\ a_{1,0} & a_{1,1} & & & \\ a_{2,0} & a_{2,1} & a_{2,2} & & \\ a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3} & \\ a_{4,0} & a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} = \begin{pmatrix} \zeta(2) & & & & \\ \zeta(3) & \zeta(3) & & & \\ \frac{\pi^4}{90} & \frac{\pi^4}{360} & \frac{\pi^4}{90} & & \\ \zeta(5) & 2\zeta(5)-\zeta(3)\zeta(2) & 2\zeta(5)-\zeta(3)\zeta(2) & \zeta(5) & \\ \frac{\pi^6}{945} & \frac{\pi^6-630\zeta^2(3)}{1260} & \frac{23\pi^6}{15120}-\zeta^2(3) & \frac{\pi^6-630\zeta^2(3)}{1260} & \frac{\pi^6}{945} \end{pmatrix}.$$

The s-channel EFT-hedron defined in eq.(3.172) says that the Hankel matrix for $A_{k,I}=\vec{a}_k\cdot\mathcal{W}_I$ must be a totally positive matrix, where $\mathcal{W}_I$ is the facets.

Let us first consider the facets $\mathcal{W}_{I_{\mathbb{I}}}$, the unit vectors. The Hankel matrix for these facets is

$$\mathcal{W}^q_{I_{\mathbb{I}}}=\delta^{q0}:\begin{pmatrix} \zeta_2 & \zeta_3 & \frac{\pi^4}{90} \\ \zeta_3 & \frac{\pi^4}{90} & \zeta_5 \\ \frac{\pi^4}{90} & \zeta_5 & \frac{\pi^6}{945}\end{pmatrix},\quad \mathcal{W}^q_{I_{\mathbb{I}}} = \delta^{q1}:\begin{pmatrix} \zeta_3 & \frac{\pi^4}{360} \\ \frac{\pi^4}{360} & 2\zeta_5-\zeta_3\zeta_2\end{pmatrix},\begin{pmatrix} \frac{\pi^4}{360} & 2\zeta_5-\zeta_3\zeta_2 \\ 2\zeta_5-\zeta_3\zeta_2 & \frac{\pi^6-630\zeta_3^2}{1260}\end{pmatrix}$$

$$\mathcal{W}^q_{I_{\mathbb{I}}} = \delta^{q2}:\begin{pmatrix} \frac{\pi^4}{90} & 2\zeta_5-\zeta_3\zeta_2 \\ 2\zeta_5-\zeta_3\zeta_2 & \frac{23\pi^6}{15120}-\zeta_3^2\end{pmatrix}. \quad (3.264)$$

It is straightforward to check that these matrices are positive semi-definite.

Next we consider facets of the cyclic polytope $\mathcal{W}_{I_b}$. For this we utilize the Taylor vectors for spinning polynomials of $h=1$ listed in eq.(3.143), and denote each column as $\vec{\nu}_\ell$. Recall that due to Yang's theorem, ℓ starts at 2. Since the Taylor vectors forms a cyclic polytope, the boundaries for the $\mathbb{P}^1$, $\mathbb{P}^2$, and $\mathbb{P}^3$ geometry are given by:

$$\mathbb{P}^1:(2),\quad \mathbb{P}^2:(i,i{+}1),\quad \mathbb{P}^3:(2,i,i{+}1)\,. \quad (3.265)$$

When written in terms of dual vectors, they are given by contracting the d vectors with the $d{+}1$ component Levi-Cevita tensor. Explicitly they are given as:

$$\langle *,2\rangle = det\begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix},\quad \langle *,i,i{+}1\rangle = det\begin{pmatrix} * & 1 & 1 \\ * & \frac{\nu_{\ell,1}}{\nu_{\ell,0}} & \frac{\nu_{\ell+1,1}}{\nu_{\ell+1,0}} \\ * & \frac{\nu_{\ell,2}}{\nu_{\ell,0}} & \frac{\nu_{\ell+1,2}}{\nu_{\ell+1,0}}\end{pmatrix},$$

$$\langle 2,*,i,i{+}1\rangle = det\begin{pmatrix} 1 & * & 1 & 1 \\ 0 & * & \frac{\nu_{\ell,1}}{\nu_{\ell,0}} & \frac{\nu_{\ell+1,1}}{\nu_{\ell+1,0}} \\ 0 & * & \frac{\nu_{\ell,2}}{\nu_{\ell,0}} & \frac{\nu_{\ell+1,2}}{\nu_{\ell+1,0}} \\ 0 & * & \frac{\nu_{\ell,3}}{\nu_{\ell,0}} & \frac{\nu_{\ell+1,3}}{\nu_{\ell+1,0}}\end{pmatrix}. \quad (3.266)$$

When taking the inner product with some vector X, $*$s denotes the position where components of X should be placed. For example for $\mathbb{P}^1$, the coupling constants are organized as

$$\vec{a}_k = \begin{pmatrix} 1 \\ \frac{a_{k,1}}{a_{k,0}} \end{pmatrix} \tag{3.267}$$

and identify $\mathcal{W}_I$ as the boundary for $\mathbb{P}^1$ in eq.(3.266), we find (again with $A_k \equiv \vec{a}_k \cdot \mathcal{W}_I$)

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} \frac{a_{1,1}}{a_{1,0}} \\ \frac{a_{2,1}}{a_{2,0}} \\ \frac{a_{3,1}}{a_{3,0}} \end{pmatrix}. \tag{3.268}$$

Then from eq.(3.172), we see that being inside the s-channel EFT-hedron requires

$$K[\vec{A}] = \begin{pmatrix} A_1 & A_2 \\ A_2 & A_3 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{4} \\ \frac{1}{4} & 2 - \frac{\zeta(2)\zeta(3)}{\zeta(5)} \end{pmatrix} \tag{3.269}$$

to be a totally positive matrix. Indeed one can straightforwardly verify that each component and the determinant of the above matrix is positive. Next let's consider the constraint in $\mathbb{P}^2$. Choosing $\mathcal{W}_I$ from eq.(3.266) to be $\langle *, 6, 7\rangle$, we find,

$$\begin{pmatrix} A_2 \\ A_3 \\ A_4 \end{pmatrix} = \begin{pmatrix} \frac{7(45a_{2,0}-20a_{2,1}+6a_{2,2})}{30a_{2,0}} \\ \frac{7(45a_{3,0}-20a_{3,1}+6a_{3,2})}{30a_{3,0}} \\ \frac{7(45a_{4,0}-20a_{4,1}+6a_{4,2})}{30a_{4,0}} \end{pmatrix} = \begin{pmatrix} \frac{161}{15} \\ \frac{7}{90}\left(51 + \frac{7\pi^2\zeta(3)}{\zeta(5)}\right) \\ \frac{721}{80} + \frac{882\zeta^2(3)}{\pi^6} \end{pmatrix}. \tag{3.270}$$

One again finds that the matrix $\begin{pmatrix} A_2 & A_3 \\ A_3 & A_4 \end{pmatrix}$ is totally positive.

Full EFT-hedron

Now let's consider the tree-level closed superstring amplitude in four-dimensions, with $M(1^{+2}2^{-2}3^{+2}4^{-2})$:

$$-\langle 24\rangle^4[13]^4\frac{\Gamma[-s]\Gamma[-t]\Gamma[-u]}{\Gamma[1+s]\Gamma[1+t]\Gamma[1+u]} = \langle 24\rangle^2[13]^2\left[-\frac{1}{stu} + \sum_{k,q} a_{k,q} z^{k-q}t^q\right], \tag{3.271}$$

whose low energy effective coupling constants are:

$$\begin{pmatrix} a_{0,0} & & & \\ a_{2,0} & a_{2,2} & & \\ a_{4,0} & a_{4,2} & a_{4,4} & \\ a_{6,0} & a_{6,2} & a_{6,4} & a_{6,6} \end{pmatrix} =$$

$$\begin{pmatrix} 2\zeta(3) & & & \\ 2\zeta(5) & \frac{3}{2}\zeta(5) & & \\ 2\zeta(7) & 3\zeta(7) & \frac{9}{8}\zeta(7) & \\ 2\zeta(9) & \frac{1}{6}(8\zeta^3(3)+31\zeta(9)) & \frac{1}{24}(-16\zeta^3(3)+73\zeta(9)) & \frac{1}{96}(8\zeta^3(3)+85\zeta(9)) \end{pmatrix}. \tag{3.272}$$

Since the UV states now appear in both $s-u$ channels, the couplings should satisfy the constraints of the full EFT-hedron.

Now let's consider the simplest EFT-hedron constraint in $\mathbb{P}^1$, which was discussed in detail in Appendix.B.6. The difference is that we will use spinning polynomials for our facets. Furthermore, due to the helicity configuration, the s-channel and u-channel will contribute independently and a Minkowski sum over polytopes will be taken. To simplify the discussion, we will assume permutation invariance for the space of amplitudes that we want to constrain here. The absence of $a_{2,1}, a_{4,1}, \cdots$ terms in the above is then just a direct consequence of this, and other amplitudes in this space can be compared with the closed superstring amplitude on equal footing. For each k, the polytope will be a Minkowski sum of the polytopes from s- and u- channels. Let us denote the vertices contributed by spin-ℓ as

$$(x_{\tilde{\ell},k,0}, x_{\tilde{\ell},k,2}), \tag{3.273}$$

where $\tilde{\ell}$ zips together information about spin and channel, for example like

$$\{(1,s),(2,u),\cdots\}.$$

Projectively,

$$\left(1, x^{(k)}_{\tilde{\ell}}\right) = \left(1, \frac{x_{\tilde{\ell},k,2}}{x_{\tilde{\ell},k,0}}\right), \quad for\ k = 2,4,6\ , \tag{3.274}$$

then we have

$$\min x^{(2)}_{\tilde{\ell}} = -\frac{23}{20}, \quad \min x^{(4)}_{\tilde{\ell}} = -\frac{11}{2}, \quad \min x^{(6)}_{\tilde{\ell}} = -\frac{165}{16}, \tag{3.275}$$

and hence we choose $\mathcal{W} = (-w, 1)$, with $w = -\frac{165}{16}$. Note that again we find that the boundary of the Minkowski sum is given by that of maximal k. Now organizing the couplings as

$$\begin{pmatrix} 1 & \frac{a_{2,2}}{a_{2,0}} \\ 1 & \frac{a_{4,2}}{a_{4,0}} \\ 1 & \frac{a_{6,2}}{a_{6,0}} \end{pmatrix} = \begin{pmatrix} \vec{a}_2 \\ \vec{a}_4 \\ \vec{a}_6 \end{pmatrix}, \tag{3.276}$$

the constraint in eq.(3.227) then tells us that

$$\begin{aligned} &(\vec{a}_2 \cdot \mathcal{W})(\vec{a}_6 \cdot \mathcal{W}) - \alpha_{min}(\vec{a}_4 \cdot \mathcal{W})^2 \\ =\ & \frac{177}{16}\left(\frac{619}{48} + \frac{2\zeta^3(3)}{3\zeta(9)}\right) - \alpha_{min}\left(\frac{189}{16}\right)^2 > 0, \end{aligned} \tag{3.277}$$

where α_{min} is defined as the minimum of $\frac{(x^{(6)}_{\tilde{\ell}}-w)(x^{(2)}_{\tilde{\ell}}-w)}{x^{(4)}_{\tilde{\ell}}-w)^2}$. Direct evaluation shows this is indeed true.

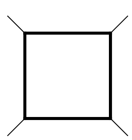
Living near the boundary of unitary polytopes

Now that we've seen how explicit EFTs satisfy our EFT-hedron bounds, we would like to see where do they actually reside. For example, consider the two dimensional region carved out by $\mathbf{X}_{cyc} \cap \mathbf{U}_5$ in fig.(3.8), where $\mathbf{U}_5$ is the s-channel unitary polytope. Now we consider the following scalar EFTs, each with a distinct known UV completion:

- (a) The tree-level exchange of a massive Higgs in the linear Sigma model

$$-\frac{s}{s-m^2} - \frac{t}{t-m^2}\bigg|_{m\to\infty} = \cdots + \frac{1}{m^{10}}(s^5+t^5) + \cdots \tag{3.278}$$

- (b) The one-loop contribution of a massive scalar X coupled to a massless scalar ϕ via $X^2\phi$. The one-loop integrand is simply the massive box, whose low energy expansion is:

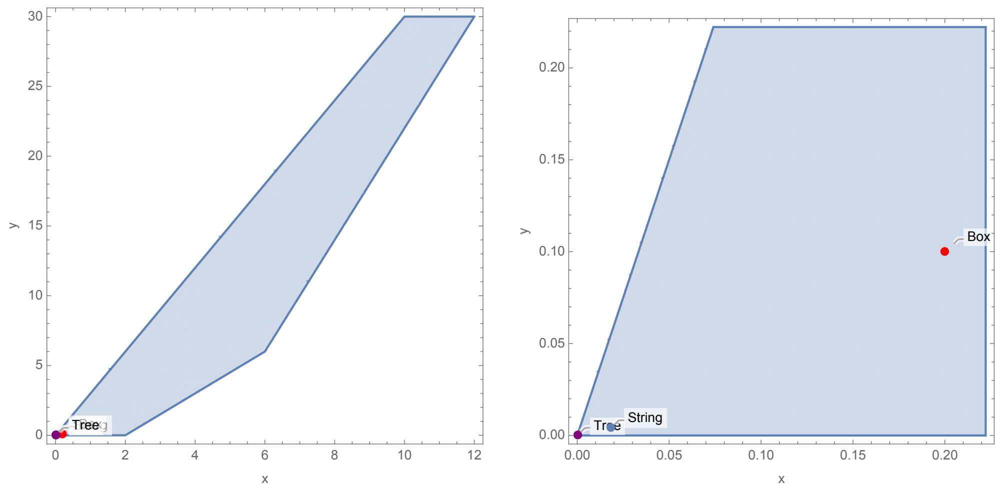


$$\bigg|_{m\to\infty} = \cdots + \frac{(s^5 + \frac{1}{5}s^4 t + \frac{1}{10}s^3t^2 + \frac{1}{10}s^2t^3 + \frac{1}{5}st^4 + t^5)}{1153152 m^{14}\pi^2} + \cdots \tag{3.279}$$

- (c) The type-I stringy completion of bi-adjoint scalar theory:

$$\begin{aligned} -\frac{\Gamma[-\alpha' s]\Gamma[-\alpha' t]}{\Gamma[1-\alpha' s-\alpha' t]}\bigg|_{\alpha'\to 0} &= \cdots + \alpha'^5\left[\zeta_7 s^5 + \left(-\frac{\pi^4\zeta_3}{90} - \frac{\pi^2\zeta_5}{6} + 3\zeta_7\right)s^4 t\right. \\ &\left.+\left(-\frac{\pi^4\zeta_3}{72} - \frac{\pi^2\zeta_5}{3} + 5\zeta_7\right)s^3t^2 + (s\leftrightarrow t)\right] + \cdots \end{aligned} \tag{3.280}$$

where we've listed the coefficients for $k = 5$. Plotting their position within $\mathbf{X}_{Cyc} \cap \mathbf{U}_5$, we find:



$$(3.281)$$

Note that they are sitting extremely close to the bottom tip of the allowed region! Let's consider another example for the graviton $s{-}u$ polytope, parameterized for the MHV configuration as:

$$\langle 24\rangle^4[13]^4 \left(\{massless\ poles\} + \sum_{k,q} z^{k-q}t^q\right) . \tag{3.282}$$

In the most general case, we can have R^3 operator which introduces a t-channel obstruction for operators proportional to z^2. Consider the coefficients $(a_{8,0}, a_{8,2}, a_{8,4})$ such that the geometry is $\mathbb{P}^2$. In principle the odd power coefficients will also be important for comparing spectral densities contributed from each spin. Here we simply wish to visualize certain coefficients in a convenient way. Two theory points that are nearby on this plot can still have very different spectral densities.

We projectively plot the corresponding polygon in the coordinates $(\frac{a_{8,2}}{a_{8,0}}, \frac{a_{8,4}}{a_{8,0}})$. The result as well as the positions of the coefficient for Type-II, Heterotic and bosonic strings is presented in fig.3.18. Labels for lower spin vertices are omitted for clarity. Once again, we see that the three distinct string EFTs are cluttered close to the lowest spins of the entire geometry.

In fact, this behaviour is ubiquitous as we survey other k, as well as the $s{-}u$ channel polytopes: all known EFTs sit close to the boundaries characterized by the low-spin vertices. This implies that the residue or discontinuity induced by the UV completion is generically dominated by low spins! For the linear sigma model, we only have a spin zero exchange so this is trivial. Listing the Gegenbauer coefficients for the residue of the

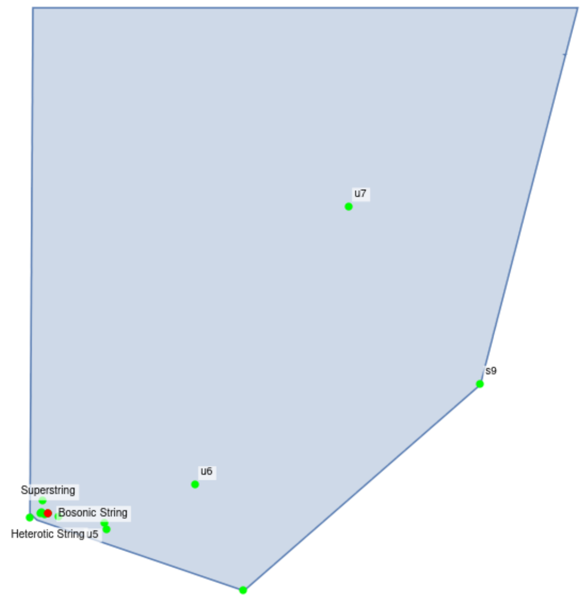


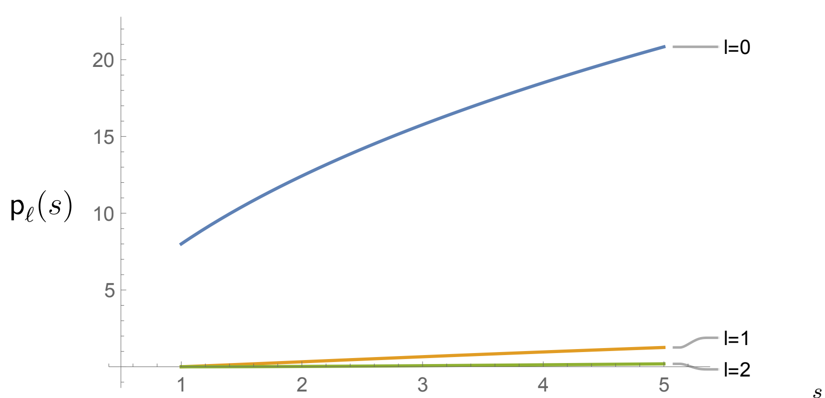
Figure 3.18: The unitary polygon for $(a_{8,0}, a_{8,2}, a_{8,4})$ of the graviton EFT. We see that the string theory EFTs are clustered near the low spin boundaries of the polygon.

open string to level n,

$\ell \backslash n$	1	2	3	4	5
0	1				$\frac{1}{11880}$
1		$\frac{1}{14}$		$\frac{1}{924}$	
2			$\frac{1}{84}$		$\frac{25}{39312}$
3				$\frac{2}{693}$	
4					$\frac{125}{144144}$

, (3.283)

we see that the leading scalar coefficient is dominant over the rest. For the box integral, the spinning spectral function for the discontinuity is discussed in detail in appendix B.4; see eq.(B.73). Plotting the spectral function for spin-$0, 1, 2$ as a function of s we find:



(3.284)

where s is normalized with respect to $4m^2$, and hence the plot begins only at the branch point $s = 1$. Once again the scalar spectral function dominates the contribution from other spins, and the ratio increases as we increase with s. Note that the positivity of the six-dimensional a-anomaly for a free massive scalar was precisely due to such suppression [70]. The suppression of higher spin coefficients can be understood from the polynomial boundedness of the amplitude: as a spin-ℓ exchange in the t-channel will bring a contribution behaving as s^ℓ at large s, polynomial boundedness then implies that higher spin contributions must be suppressed. Indeed the suppression at large spins is precisely what led to the Froissart bound as reviewed in appendix B.1. Thus in general, we expect physical EFTs to lie near the low spin boundaries of the unitary polytope, although a more quantitative understanding of the implications from such suppression is clearly desired, which we leave to future work.

If EFTs naturally live near the low-spin boundaries of the unitary polytope, what is the purpose of the rest? Note that for a given UV completion, there exists an entire family of effective theories for which the EFTs discussed above are in the deep IR. Here, the scale dependence under discussion is not from the running generated from the massless loops, which will be the focus in the next section, but rather from the simple fact that different part of the spectrum is visible depending on the energy. What this means in practice is that at a given energy scale Λ, the couplings for our higher dimensional operators take the form:

$$M(s,t) = \{\textit{massless/massive poles}\} + \sum_{k,q} a^{\Lambda}_{k,q} s^{k-q} t^q \,, \tag{3.285}$$

where the amplitude now contains massless as well as massive poles for all the massive states below Λ. When the couplings are defined in such fashion, they naturally become Λ dependent. Let us consider an explicit example. Imagine that we are studying type -II string theory at some energy scale and we have discovered the first few massive states up to level n. At this scale the amplitude at fixed t should take the form

$$\sum_{a=1}^{n} R_a(t) \left(\frac{1}{s-a} + \frac{1}{u-a} \right) + \sum_{k,q} a^{(n)}_{k,q} z^{k-q} t^q, \tag{3.286}$$

where $R_a(t) = \frac{1}{(a!)^2} \prod_{i=1}^{a-1} (t+i)^2$ is the residue for the resonance $s = a$. The value of the couplings for the higher dimensional operators can be extracted by Taylor expanding both sides of

$$\sum_{k,q} a^{(n)}_{k,q} z^{k-q} t^q = \frac{\Gamma[-s]\Gamma[-t]\Gamma[-u]}{\Gamma[1{+}s]\Gamma[1{+}u]\Gamma[+t]} - \left[\sum_{a=1}^{n} R_a(t) \left(\frac{1}{s-a} + \frac{1}{u-a} \right) \right]. \tag{3.287}$$

Note that by construction, the couplings must reside inside our unitary polytope. Since the massive poles that are "subtracted" from the full UV completion are precisely the

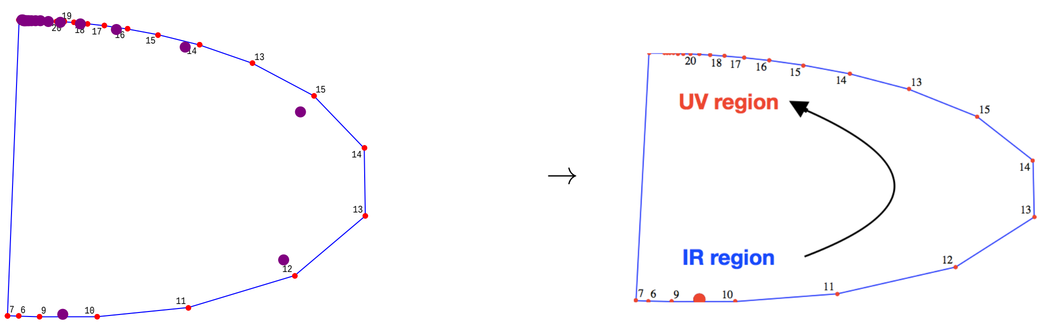


Figure 3.19: On the LHS we the purple dots indicate (x_n, y_n) for $n = 0, \cdots, 20$, representing the position of the type-II string EFT in side the unitary polytope for $s \sim \frac{n}{\alpha'}$. We see that as we go to large s, the EFT tends to the corner with higher spins. This implies that the UV and IR EFTs populate different regions in the polytope, as illustrated on the right.

dominating low spin states, we expect the resulting couplings to *float* towards the upper region of the polytope! When plotting the coefficients for

$$(x_n, y_n) = \left(\frac{a^{(n)}_{8,2}}{a^{(n)}_{8,0} + a^{(n)}_{8,4}/10^3}, \frac{a^{(n)}_{8,4}}{a^{(n)}_{8,0} + a^{(n)}_{8,4}/10^3} \right)$$

in fig.3.19, we see that indeed as we raise the energy scale the corresponding EFT probes deeper in the unitary polytope.

Thus in summary, the low spin regions of the unitary polytope correspond to the EFTs in the deep IR, while the higher spin region corresponds to the EFTs in the UV. We leave the detailed study of this UV-IR relation to future work.

3.11 Running into the EFT-hedron

Let us now turn to discussing the full amplitude including the massless loops that induce the logarithmic running of the EFT couplings. For example, consider again the linear sigma model, whose tree-amplitude is given in eq.(3.12). At one-loop the coefficients of the s^4 starts receiving loop-corrections from the s^2 operators:

$$\begin{aligned} M^{IR}(s,t) \quad = \quad & \frac{\bar{a}_2}{m_h^4}(s^2+t^2+u^2) + \frac{\bar{a}_4}{m_h^8}(s^4+t^4+u^4) - \left[\bar{a}_2^2 \frac{1}{15(4\pi)^2 m_h^8}\left(41s^2+u^2+t^2\right) s^2 \log\frac{s}{s_0}\right. \\ & \left. + (s \leftrightarrow t) + (s \leftrightarrow u)\right] + \mathcal{O}(p^{10}) , \end{aligned} \tag{3.288}$$

where $\bar{a}_i$s are to be understood as renormalized couplings at some scale s_0. In this paper, we will only consider one-loop effects for EFTs that have a well defined S-matrix. The derivative couplings ensures that expansion near the forward limit is well defined, since the t-channel cut appears as $t^n \log t$, as can be seen in the above, and hence there is no singularity at the branch point $t = 0$. The presence of the massless logs leads to two pressing issues, 1.) there is a massless cut coming all the way to the origin, and thus the low energy couplings, analytically extracted from eq.(3.42), are no longer well defined. 2.) the fact that coupling runs also brings into question the fate of our previous positivity bounds as the theory flows to the IR.

Naively, one can simply introduce a mass regulator,[10] which will allow us to push the massless cut away from the origin of the complex s-plane. Since this corresponds to introducing a massive state, all ingredients necessary to the derivation of previous positivity bounds are intact and should hold whenever the EFT is valid. This means that running in the IR will stay within the unitary polytope. However, it is easy to see from explicit examples that this is *not* the case, the massless logs can take us outside of the EFT hedron! This apparent contradiction originated from the fact that the mass deformed theory does not reproduce the correct IR behaviour of the massless loops. It is instructive to see why our intuition was wrong, which in turn will guide us to defining "generalized EFT couplings", for which previous positivity constraints apply.

Running out of bounds Let's consider the EFT of a single massless scalar with the following higher dimension operators turned on:

$$\mathcal{L}_{Int} = \frac{a_2}{\Lambda^4}(\partial\phi)^4 + \frac{a_4}{\Lambda^8}(\partial^2\phi)^4 + \frac{a_6}{\Lambda^{12}}(\partial^3\phi)^4 , \tag{3.289}$$

The one-loop RG equation is then

$$\mu^2\frac{\partial a_4}{\partial\mu^2} = 0, \quad \mu^2\frac{\partial a_4}{\partial\mu^2} = \beta_1 a_2^2, \quad \mu^2\frac{\partial a_6}{\partial\mu^2} = \beta_2 a_2 a_4 . \tag{3.290}$$

[10]This of course can only be consistently done for scalars and vectors, but not gravity.

With the solution, $a_2 = \bar{a}_2$, $a_4 = \bar{a}_4 + \beta_1 \bar{a}_2^2 \log \frac{s_0}{p^2}$ and $a_6 = \bar{a}_6 + \beta_2 \bar{a}_2 \bar{a}_4 \log \frac{s_0}{p^2}$. For simplicity let's consider the forward-limit Hankel matrix constraints, and set $\bar{a}_i$s be the renormalized couplings at some scale M^2 where the constraints hold. For example we have $\bar{a}_i > 0$ and

$$\bar{a}_2 \bar{a}_6 - \bar{a}_4^2 > 0\,. \tag{3.291}$$

Now as we allow the couplings to run in the IR, the determinant of the Hankel matrix becomes:

$$Det \begin{pmatrix} \bar{a}_2 & \bar{a}_4 + \beta_1 \bar{a}_2^2 \delta \\ \bar{a}_4 + \beta_1 \bar{a}_2^2 \delta & \bar{a}_6 + \beta_2 \bar{a}_2 \bar{a}_4 \delta \end{pmatrix} = (\bar{a}_2 \bar{a}_6 - \bar{a}_4^2) + (\beta_2 - 2\beta_1) \bar{a}_4 \bar{a}_2^2 \delta + \mathcal{O}(\delta^2)\,, \tag{3.292}$$

where we have used a short-hand notation $\delta = \log \frac{s_0}{p^2}$. If the running couplings were to stay inside the EFT-hedron, we would have a sharp prediction for the one-loop beta functions, namely $(\beta_2 - 2\beta_1) > 0$. Since for our current theory we only have bubble integrals at one-loop, their coefficients can be directly captured from the two-particle cut, which we derive in appendix B.7, yielding $\beta_1 = \frac{14}{5(4\pi)^2}$ and $\beta_2 = \frac{166}{35(4\pi)^2}$. Immediately we see that $\beta_2 - 2\beta_1 < 0$ in contradiction to the expectation from the Hankel matrix bounds. In other words, the low energy running drives the couplings outside of the EFT hedron!

Let us see why our intuition from the mass regulated picture failed to yield the correct prediction. Consider the explicit low energy amplitude in the forward limit, which is all that is necessary for eq.(3.291). We have:

$$M(s,0) = 2\frac{s^2}{\Lambda^4}\bar{a}_2 + 2\frac{s^4}{\Lambda^8}\left(\bar{a}_4 + \beta_1 \bar{a}_2^2 \log\frac{M^2}{s}\right) + 2\frac{s^6}{\Lambda^{12}}\left(\beta_2 \bar{a}_2 \bar{a}_4 \log\frac{M^2}{s}\right)\,, \tag{3.293}$$

where we've set $\mu^2 = M^2$, representing the scale for which the Hankel constraint holds. Now by deforming the massless loop propagators to be massive, the logs get deformed as:

$$\log\frac{M^2}{s} \to \log\frac{M^2}{m^2} - i\sqrt{\frac{1}{z} - 1}\log(i\sqrt{z} + \sqrt{1 - z^2}) - 1 = \log\frac{M^2}{m^2} - \sum_n \frac{(1)_{n-1}}{3\left(\frac{5}{2}\right)_{n-1}} z^n \tag{3.294}$$

where $z \equiv \frac{s}{4m^2}$. Thus we see that at low energies, $z \ll 1$, the leading log correction appearing at s^4 is $\log \frac{M^2}{m^2}$, reproducing the same running as the massless log if we take $s, m^2 \ll M^2$. However the z expansion in eq.(3.294) introduces correction to the coefficient of $s^6, s^8, \cdots$ that dominates over their original logarithms since:

$$\frac{1}{m^2} \gg \frac{1}{\Lambda^2} \log\frac{M^2}{m^2} \tag{3.295}$$

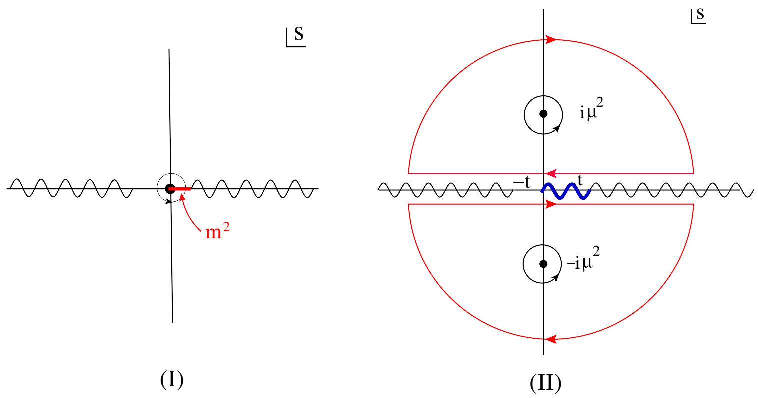


Figure 3.20: In the presence of massless cuts, we can either (I) introduce a small mass regulator and push the cut slightly away from the origin, or (II) we analytically define our generalized couplings by moving the contour at origin onto to the complex plane to $s = \pm i\mu$ in a way that the integration measure is positive definite. After deformation the contour picks up the discontinuity on the real s-axes, which for $|t| < |s|$, is controlled by unitarity. We can analytically continue to $|s| < |t|$ for theories with well behaved soft limits.

as $m^2 \to 0$. Put another way, *the small mass deformation is no longer "small" when one considers subleading contributions.* Note that due to these corrections, the Hankel matrix constraint is trivially satisfied for the mass deformed amplitude. Indeed it is straightforward to check that the Hankel matrix for $a_n \equiv \frac{(1)_{n-1}}{3\left(\frac{5}{2}\right)_{n-1}}$ is total positive, and since the z expansion in eq.(3.294) dominates the contributions for $s^6, s^8, \cdots$ couplings, they trivialize the Hankel matrix constraint on the amplitude.

Generalized EFT couplings and their dispersive representation

The reasons we've introduced the mass regulated theory is so that the massless cut is pushed off the origin, where the couplings are analytically defined. However, we've just seen that by doing so the EFT no longer captures the correct IR physics beyond leading order. Instead of moving the branch point, let's move the pole itself. For example, consider the following contour integral of the amplitude at fixed $t \ll m^2$:

$$\frac{1}{2\pi i}\oint \frac{ds\, s}{(s^2+\mu^4)^{n+1}} M(s,t)\,, \tag{3.296}$$

where the contour encircles the poles at $s = \pm i\mu^2$, and we will take $\mu^2 \ll 1$. Using this contour we can define the following generalized couplings in the forward limit

$$a^{\mu^2}_{2n,0} \equiv \frac{1}{2\pi i}\oint_{\mathcal{C}_0} \frac{ds\, s}{(s^2+\mu^4)^{n+1}} M(s,0)\,, \tag{3.297}$$

where the superscript μ^2 on g^{μ^2} indicates its the position for which the pole has been moved off the origin. Note that we've naturally introduced scale dependence into the definition of the coupling. Now in the forward limit, $M(s,0)$ is finite since the t-channel cut is suppressed by pre-factors proportional to powers of t, guaranteed by the derivative coupling. Again deform the contour $\mathcal{C}_0$ to $\mathcal{C}_\infty$, and this relates the generalized couplings to the discontinuity of the amplitude on the s-axes as illustrated in fig.3.20. In other words, we have

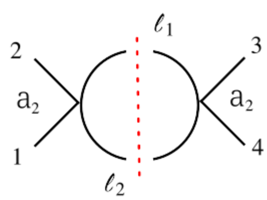
$$a^{\mu^2}_{2n,0} \;=\; -\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{ds\,s}{(s^2+\mu^4)^{n+1}}\,\mathrm{Im}\,M(s,0)\,. \tag{3.298}$$

Once again, let's demonstrate the validity of eq.(3.298) using our linear sigma model amplitude in eq.(3.288). Since the amplitude behaves as $s^4\log s$ as $s\to\infty$, we should expect eq.(3.297) and eq.(3.298) to agree for $a^{\mu^2}_{6,0}$. Using eq.(3.297) the generalized couplings evaluate to:

$$a^{\mu^2}_{4,0} = \frac{\bar{a}_4}{m_h^8} - \frac{7a_2^2}{160\pi^2 m_h^8}\left(3+2\log\frac{\mu^4}{s_0^2}\right),\quad a^{\mu^2}_{6,0} = \frac{7\bar{a}_2^2}{240\pi^2 m_h^8\mu^4}\,. \tag{3.299}$$

As expected, the $a^{\mu^2}_{4,0}$ is given by the combination of tree coefficient $\bar{a}_4$ and the one-loop log proportional to $\bar{a}_2^2$. Moreover, even though we only consider the amplitude up to s^4 terms, all generalized couplings $a^{\mu^2}_{2n,0}$ are nonzero due to the log. Now for eq.(3.298) the imaginary part of the four-point amplitude arising from the s- cut is given by:

$$\mathrm{Im}_{s=[0,\infty]}\,M(s,t) \;=$$



$$\begin{aligned} &= \frac{\bar{a}_2^2}{(4\pi)^3 m_h^8}\int d\phi' d\cos\theta'\,(s^2+t^2+u^2)_L(s^2+t^2+u^2)_R \\ &= -\frac{\bar{a}_2^2}{m_h^8}\frac{s^4}{60(4\pi)^2}(167+\cos 2\theta)\,, \end{aligned} \tag{3.300}$$

where θ is the scattering angle. Taking the forward limit one finds $\mathrm{Im}_{s=[0,\infty]}\,M(s,0) = -\frac{7\bar{a}_2^2}{m_h^8}\frac{s^4}{40\pi^2}$, reproducing the coefficient of the s-channel logarithm in eq.(3.288). Using $\mathrm{Im}_{s=[0,\infty]}\,M(s,0) = -\mathrm{Im}_{s=[-\infty,0]}\,M(s,0)$, one recovers,

$$\frac{1}{2\pi i}\left(-\int_{-\infty}^{0}+\int_{0}^{\infty}\right)\frac{ds\,s}{(s^2+\mu^4)^4}i\pi\frac{7\bar{a}_2^2}{40\pi^2 m_h^8}s^4 = \frac{7\bar{a}_2^2}{240\pi^2\mu^4 m_h^8}\,, \tag{3.301}$$

in agreement with eq.(3.299).

Now deforming the contour one again picks up the discontinuity on the real axes as shown in fig.(3.20) (II). Now the question is whether the discontinuity is given by physical thresholds. For $t<0$, the region $|t|\leq|s|$ corresponds to the physical kinematics and thus its discontinuity is determined from unitarity. Due to the derivative couplings, there are no new singularities at $t=0$, and we can analytically continue to positive t. Thus the entire s-channel discontinuity can be obtained by analytically continuation of that in the physical regime, i.e. it is expressible as a positive sum of the Gegenbauer polynomials in $(D-1)$-spatial dimensions:

$$Dis_{s>0}[M(s,t)]=\sum_{\ell=0,2,4}\mathsf{p}_\ell(s)P_\ell(\theta)\,. \tag{3.302}$$

Let's demonstrate the above in a non-trivial example. The one-loop correction to the scalar theory introduced earlier in this section has one-loop logarithm proportional to $\bar{a}_2^2$, $\bar{a}_2\bar{a}_4$, and $\bar{a}_4^2$. The first two were computed previously while the latter is given by

$$\frac{\bar{a}_4^2 s^8}{M^{16}20160(4\pi)^2}(39843+988\cos 2\theta+\cos 4\theta)\,. \tag{3.303}$$

Summing all three contributions we obtain the discontinuity on the positive real axes given by the following spinning spectral functions

$$\mathsf{p}_0(s)=\frac{s^4(25\bar{a}_2+21\bar{a}_4s^2)^2}{225(4\pi)^2},\quad \mathsf{p}_2(s)=\frac{s^4(7\bar{a}_2+12\bar{a}_4s^2)^2}{2205(4\pi)^2},\quad \mathsf{p}_4(s)=\frac{\bar{a}_4^2s^8}{11025(4\pi)^2}\,, \tag{3.304}$$

and indeed they are positive definite.

In conclusion, the generalized coupling constants defined through the contour integral in eq.(3.296), again subject to appropriate boundary behaviour, will satisfy the same analytic constraint as that before. In the following we will demonstrate with explicit examples that the Hankel matrix constraint is satisfied.

The Hankel matrix constraints:

Let's again take the forward limit four-point amplitude for eq.(3.289)

$$M_4(s,0) \;=\; 2\,\frac{\bar{a}_2s^2}{\Lambda^4}+2\left(\bar{a}_4+\beta_1\bar{a}_2^2\log\frac{M^2}{s}\right)\frac{s^4}{\Lambda^8}+2\left(\bar{a}_6+\beta_1\bar{a}_2\bar{a}_4\log\frac{M^2}{s}\right)\frac{s^6}{\Lambda^{12}}. \tag{3.305}$$

The generalized couplings are then given by

$$
\begin{aligned}
a^{\mu^2}_{2,0} &= \frac{1}{\Lambda^4}\left[\bar{a}_2 + z^4\left(\beta_1\bar{a}_2^2\left(\frac{1}{2} - 2\log y\right) - 2\bar{a}_4\right) + \mathcal{O}\left(z^8\right)\right],\\
a^{\mu^2}_{4,0} &= \frac{1}{\Lambda^8}\left[\bar{a}_4 - \beta_1\bar{a}_2^2\left(\frac{3}{4} - \log y\right) + z^4\left(\beta_2\bar{a}_2\bar{a}_4\left(\frac{5}{4} - 3\log y\right) - 3\bar{a}_6\right) + \mathcal{O}\left(z^8\right)\right]\\
a^{\mu^2}_{6,0} &= \frac{1}{\Lambda^8\mu^4}\left[\frac{\beta_1\bar{a}_2^2}{6} + z^4\left(\bar{a}_6 - \beta_2\bar{a}_2\bar{a}_4\left(\frac{11}{12} - \log y\right)\right) + \mathcal{O}\left(z^8\right)\right],
\end{aligned}
\tag{3.306}
$$

where $z = \frac{\mu^2}{\Lambda^2}$ and $y = \frac{M^2}{\mu^2}$. First of all, we see that the leading contributions for $a^{\mu^2}_{2,0}$ are given by the tree-level coupling $\bar{a}_2$, where as for $a^{\mu^2}_{4,0}$ the tree-level coupling $\bar{a}_4$ mixes with logarithmic contributions $\beta_1\bar{a}_2^2\log y$ at leading order. However, beyond $a^{\mu^2}_{4,0}$ the original tree-couplings become subdominant to terms that were generated from the logarithms in $a^{\mu^2}_{4,0}$. Indeed for $a^{\mu^2}_{6,0}$ the tree-level piece $\bar{a}_6$ is *subleading* to a term proportional to $\beta_1\bar{a}_2^2$, which came from the leading logarithm in $a^{\mu^2}_{4,0}$. The dominance of terms induced by the the leading log for all $a^{\mu^2}_{2n,0}$ with $n > 2$, is reminiscent of the leading $\frac{1}{m}$ corrections flooding the higher-derivative couplings for the mass regulated case discussed previously. As we will see, these effects ensures the positivity constraints on the generalized couplings which we now derive.

Now let us consider the dispersive representation:

$$
a^{\mu^2}_{2n,0} \;=\; -\int_{-\infty}^{\infty}\frac{ds\, s}{(s^2+\mu^4)^{n+1}}\mathrm{Im}\; M(s,0)\,, \tag{3.307}
$$

As discussed above, even in the presence of massless cut, the discontinuity is still given by a positive sum of Gegenbauer polynomials. The only modification is that the s-channel cut now starts at $s = 0$. Incorporating the u-channel cut, we then have a branch cut covering the entire real axes leading to

$$
\begin{aligned}
a^{\mu^2}_{2n,0} &= \left[-\int_{-\infty}^{0} + \int_{0}^{\infty}\right]\frac{ds\, s}{(s^2+\mu^4)^{n+1}}\sum_{\ell}\mathsf{p}_\ell(s)G_\ell^{\frac{D-4}{2}}(1)\\
&= \sum_{\ell}\int_{0}^{\infty}\frac{dx}{(x+\mu^4)^{n+1}}\,\mathsf{p}_\ell(x)G_\ell^{\frac{D-4}{2}}(1)\,,
\end{aligned}
\tag{3.308}
$$

In other words, it is given by a continuous sum of points on the moment curve:

$$
\begin{pmatrix} a^{\mu^2}_{2,0}\\ a^{\mu^2}_{4,0}\\ a^{\mu^2}_{6,0}\\ \vdots\\ a^{\mu^2}_{2n,0}\end{pmatrix} = \sum_i c_i\begin{pmatrix}1\\ y_i\\ y_i^2\\ \vdots\\ y_i^{n-1}\end{pmatrix},\quad c_i > 0,\; y_i > \frac{1}{\mu^4}\;\forall i\,. \tag{3.309}
$$

Note that the moment curve is shifted by $\frac{1}{\mu^4}$, and thus the coefficients will obviously satisfy the original Hankel matrix constraint.

Let us show this in detail for the generalized couplings in eq.(3.306). First of all in the limit $\mu^2 \ll \Lambda^2$, the positivity of $a^{\mu^2}_{2,0}, a^{\mu^2}_{4,0}, a^{\mu^2}_{6,0}$, and $a^{\mu^2}_{2,0}a^{\mu^2}_{6,0}-(a^{\mu^2}_{4,0})^2$ is ensured by the positivity of the tree-level coupling and that of the β_is. An interesting scenario occur when we deform the position of the pole all the way to the renormalization scale $\mu^2 = M^2$, while assuming $M^2 \ll \Lambda^2$. The positivity of $a^{\mu^2}_{4,0}$ then requires that

$$\bar{a}_4-\beta_1\bar{a}_2^2\frac{3}{4} > 0\,, \tag{3.310}$$

where again $\beta_1 = \frac{14}{5(4\pi)^2} \sim 0.002$. It is easy to see that this imposes further constraint on the couplings beyond that of the tree-level Hankel constraints, i.e. the positivity of $\bar{a}_2,\ \bar{a}_4,\ \bar{a}_6$, and $\bar{a}_2\bar{a}_6 - \bar{a}_4^2$.

It is interesting to understand why this new constraint arises. First, note that the effective action considered in the beginning of this section, eq.(3.289), is not the most generic for single scalar theory: it lacks the marginal ϕ^4 interaction. In general, the lack of ϕ^4 interaction is associated spontaneous symmetry breaking in the UV, where the resulting EFT respects a shift symmetry. Now due to boundary contributions, for tree-level couplings we are not privy to the information of the constant piece of the amplitude, or $k = 0$, which translate to the presence/absence of ϕ^4 interaction. However, at loop-level, its presence will affect the pattern of IR running for the couplings. For example, the presence of ϕ^4 would induce logarithmic running already for the s^2 operator, which leads to the modification of $a^{\mu^2}_{4,0}$ to:

$$a^{\mu^2}_{4,0} = \frac{1}{\Lambda_4\mu^4}\left[\frac{\bar{a}_0\bar{a}_2\beta_0}{4}+z^4\left(\bar{a}_4-\beta_1\bar{a}_2^2\left(\frac{3}{4}-\log y\right)\right)+\mathcal{O}(z^8)\right]\,, \tag{3.311}$$

instead of eq.(3.306). Here $\bar{a}_0$ is the tree-level coupling for ϕ^4 and β_0 is the beta function for s^2 operator. We see that the running at s^2 now induces corrections for $a^{\mu^2}_{4,0}$ that dominates the original contributions! Now the positivity of $a^{\mu^2}_{4,0}$ simply implies $\bar{a}_0\beta_0 > 0$, even if we take μ close to the renormalization scale.

Said in another way, the constraint in eq.(3.310) is a reflection of $\bar{a}_0 = 0$! Let's consider an explicit UV completion that realizes such low energy behaviour: the linear sigma model. As discussed previously, the shift symmetry of the EFT ensures that there are no constant piece for the quartic interaction. In IR tree-level couplings can be identified as $\bar{a}_2 = \bar{a}_4 = \lambda$, where λ is the quartic coupling constant of the complex scalar in the UV. Thus we see that in the perturbative regime, where the map between the IR and UV couplings are applicable, eq.(3.310) is trivially satisfied.

Thus we see that when massless loops are included, the positivity bounds allow us to probe details of the EFT previously hidden behind the "Froissart horizon" !

A peek beyond the forward limit

We now consider the extension away from the forward limit, which correspond to taking a Taylor expansion around $t = 0$. Again due to the t-channel log coming in the form $t^n \log t$, the amplitude is finite in the forward limit. Due to the t-channel branch cut, once again we deform the t contour away from the origin to $t = \epsilon$:

$$a^{\mu^2}_{k,q} \equiv \left(\frac{1}{2\pi i}\right)^2 \oint \frac{dt}{(t-\epsilon)^{q+1}} \oint \frac{ds\, s^{\frac{1+(-)^k}{2}}}{(s^2+\mu^4)^{\lfloor\frac{k-q}{2}\rfloor+1}} M(s,t)\,, \tag{3.312}$$

where $\epsilon > 0$. We will be considering the limit where t is much smaller than any massive threshold. Note that since $\epsilon > 0$, we are actually analytically continuing t away from the physical regime $t < 0$. For theories such as those of interacting goldstones, where the massless amplitudes are soft enough, free of soft/collinear singularities, so that massless amplitudes are well-defined, it is reasonable to expect that discontinuities of the amplitude in the $s-$channel are actually analytic in t. Taking this as a working assumption gives us the dispersive representation. We have:

$$a^{\mu^2,\epsilon}_{k,q} = \frac{1}{2\pi i} \oint \frac{dt}{(t-\epsilon)^{q+1}} \sum_\ell \int_0^\infty \frac{ds\, s^{\frac{1+(-)^k}{2}}}{(s^2+\mu^4)^{\lfloor\frac{k-q}{2}\rfloor+1}}\, \mathsf{p}_\ell(s) G_\ell\left(1+2\frac{t}{s}\right)\,, \tag{3.313}$$

Evaluating the t-integral on the pole then gives the Taylor expansion of the Gegenbauer polynomials $G_\ell(x)$ at $x = 1+\epsilon$. Now importantly, since we've set $\epsilon > 0$, the resulting convex hull is *inside* the Gegenbauer polytope! To see this, recall that under the rescaling $x \to ax$ with $a > 1$, Gegenbauer polynomials rescales to a positive function, i.e. :

$$G_\ell((1+\epsilon)x) = \sum_{\ell'=0}^{\ell} \quad c_{\ell'} G_{\ell'}(x), \quad c_{\ell'} > 0\,. \tag{3.314}$$

It then follows that the vector $\vec{G}_\ell(1+\epsilon)$ is a positive sum of $\vec{G}_\ell(1)$, and thus the convex hull of $\vec{G}_\ell(1+\epsilon)$ must be inside Gegenbauer polytope! In fact, from eq.(3.124), we see that the convex hull of $\vec{G}_\ell(1+\epsilon)$ is another cyclic polytope. Thus as we increase in ϵ, the couplings must live in a cyclic polytope that is *contained* in the previous ones. In this precise sense, by increasing ϵ generalized couplings moves deeper inside the original geometry!

3.12 Outlook

We have seen that the constraints on vacuum stability, causality, and unitarity place enormously powerful constraints on low-energy effective field theories. There are a large number of obvious open avenues for future work. Most immediately, there is the question of fully understanding the geometry and boundary structure of the EFT-hedron for four-particle scattering; this mathematical problem has been fully solved for the toy example of the $s-$channel only EFT-hedron where it is already rather non-trivial. We have also bounded the full EFT-hedron for the most general cases of interest, but have still not determined the exact facet structure of the EFT-hedron in complete generality. It would also be interesting to extend the dispersive analysis beyond $2 \to 2$ scattering. Indeed, if we consider a simple theory with Lagrangian $P(X = (\partial\phi)^2)$, we know that subluminality for small fluctuations around background with $\langle \partial\phi \rangle \neq 0$ demands $P''(X) > 0$ for all X, which enforces positivity conditions on higher-point scattering amplitudes. Another obvious avenue is to systematically explore constraints on scattering for multiple species with general helicities.

It is also important to note that, while the EFT-hedron places powerful constraints on the effective field theory expansion, sensible effective field theories do not appear to populate the entire region allowed by the EFT-hedron, but cluster close to its boundaries. The reason is likely that the physical constraints we have imposed, while clearly necessary, are still not enough to capture consistency with fully healthy UV theories. In particular, our dispersive representation at fixed t, does not make it easy to impose the softness of high-energy, fixed-angle amplitudes where both s, t are large with t/s fixed. It would be fascinating to find a way to incorporate this extra information about UV softness into the constraints, along the lines of the celestial sphere amplitude [12], which should further reduce the size of the allowed regions for EFT coefficients.

The unexpected power of stability, causality and unitarity in constraining effective field theory raises the specter of a much greater prize, which was in the fact the question that initially motivated this work. Can the same principles be used to strongly constrain, and perhaps with additional conditions actually uniquely determine, consistent UV complete scattering amplitudes? To sharpen this question, we can begin by thinking about UV completions of gravity amplitudes at "tree-level", assuming the amplitude only has poles. Unlike theories of scalar scattering, which can be UV completed in a myriad of ways such as glueball scattering in large N-gauge theories, the only consistent tree-gravity scattering amplitudes we know of come from string theory, so it is more likely this question has a unique answer. The four particle tree graviton scattering amplitudes in string theory are essentially unique, independent of any details of compactification and fixed by the nature of the worldsheet supersymmetry. Indeed the amplitudes differ only by the massless three particles amplitudes in the low-energy theory, with type II theories having only the usual three-graviton vertex, and the heterotic theory also including the $R^2\phi$ coupling to the dilaton. So it is plausible to conjecture that amplitudes with, say, only the usual three-graviton amplitude at low-energies, have a unique tree-level UV completion given by the Virasoro-Shapiro amplitude.

As an easy first step in this direction, it is easy to see that tree-level UV completions of gravity must contain an infinite tower of massive particles of arbitrarily high spin. In fact gravity is not particularly special in this regard. Consider any theory with fundamental cubic interactions, so that four-particle amplitudes already have $\frac{1}{s,t,u}$ poles at tree-level. Suppose we wish to improve the high-energy behavior of the amplitudes relative to what is seen in the low-energy theory, so for example for gravity/Yang-Mills/ ϕ^3 theory, we would like the high-energy limit to drop more quickly that $s^2/s/s^{-1}$ respectively. It is then easy to see that this is impossible unless the UV theory has an infinite tower of particles with arbitrarily large spin.

Let us briefly sketch the reason for this. It is instructive to contrast the situation with that of simple UV completions for theories whose four-particle interaction begin with contact interactions at low-energies. Consider for instance goldstone scattering in the non-linear sigma model, where the low-energy four-particle amplitude begins as $\mathcal{A} = -\frac{1}{f^2}(s+t)$. It is trivial to UV complete this simply by softening $s \to \frac{s}{(1-s/M^2)}$, $t \to \frac{t}{(1-t/M^2)}$. This is consistent with the causality bounds at large s and fixed t, and keeps the fixed-angle amplitude small so long as $M^2 \ll f^2$. And crucially, thanks to the overall negative sign in front of the amplitude, the residues on the massive poles are positive and are interpreted as the production of a scalar particle with positive probability. This is of course nothing but the linear sigma model UV completion of the non-linear sigma model, with the new massive particle identified as the Higgs. Note that had the overall sign of the amplitude been reversed, we would not be able to do this, as the residue on the massive pole would be negative.

Now, consider instead the amplitude $\mathcal{A} = g^2(\frac{1}{s} + \frac{1}{t} + \frac{1}{u})$ for ϕ^3 theory at tree-level, and let us try to add massive poles to make the amplitude decrease faster than $1/s$ at high-energies. It is easy to see that the same strategy used in the goldstone example can't work. For instance if we again attempt to soften $\frac{1}{s} \to \frac{1}{s(1-s/M^2)}$, the residue on the massive pole will have the opposite sign as that of the (correct, positive) residue on the massless pole at $s = 0$! This will happen for any amplitude that is a rational function (finite number of massive poles) in the Mandelstam variables. If the amplitude is softened in the physical region, it is softened everywhere in the s-plane; so given that the amplitude vanished faster than $1/s$ at infinity, the sum of all the residues must be zero. But that means that some of the massive residues must be negative, to cancel the positive residue at $s = 0$. This can only be avoided if there are infinitely many poles that allow the function to die in the physical region but blow up elsewhere in the s-plane, as is familiar in string theory. A small elaboration of this argument also shows the necessity of an infinite tower of spins, and the same arguments apply to gravity and Yang-Mills amplitudes as well.

It is amusing that theories that only have a life in the UV—such as the weak interactions and the non-linear sigma model, whose low-energy amplitudes are tiny—are "easy" to UV complete with finitely many massive states. It is theories with IR poles, associated with long-range interactions, that are forced to have much more non-trivial UV completions. This is why the most ancient interaction described by physics—gravity—continues to be the most challenging to UV complete, while the weak interactions were discovered and UV completed within about half a century!

One can also easily "discover" the stringy completion of gravity amplitudes, from the bottom-up, as the simplest possible UV completion with an infinite tower of poles satisfying extremely basic consistency conditions, even before imposing the restrictions of causality and unitarity. The tree-level 4-graviton amplitude is $\mathcal{A}^{+-+-} = G_N \langle 13 \rangle^4 [24]^4 \times \frac{1}{stu}$. We know that any tree-level UV completion must have an infinite tower of poles, in the s, t, u channels. Thus, the most general Ansatz for the amplitude would replace $\frac{1}{stu} \to \frac{N(s,t,u)}{stu \prod_i (s-m_i^2)(t-m_i^2)(u-m_i^2)}$. Note that this expression has the property that on the $s-$ channel pole at $s = m_j^2$, the residue has poles at $t = m_i^2$ and $u = m_i^2 \to t = -(m_i^2 + m_j^2)$. These poles must be absent in the physical amplitude, and thus the numerator must have zeroes, when $s = m_j^2$, at these values of t. It is then natural to make the simple assumption that these are the *only* zeroes of the numerator. That tells us that if we write $N(s,t,u) = \prod_j (s + r_i)(t + r_i)(u + r_i)$, then the set of all the roots $\{r_i\}$ must contain all of $\{m_i^2, m_i^2 + m_j^2\}$. And this in turn is most trivially accomplished if $m_j^2 = M_s^2 j$ are just all the integers in the units of a fundamental mass scale M_s!

By this simple reasoning, we are led to the infinite product formula for the Virasoro-

Shapiro amplitude, putting $\alpha' = M_s^{-2}$:

$$\begin{aligned} \mathcal{A} &= G_N \langle 13\rangle^4 [24]^4 \frac{\prod_{j=1}^{\infty}(\alpha' s + j)(\alpha' t + j)(\alpha' u + j)}{\prod_{i=0}^{\infty}(\alpha' s - i)(\alpha' t - i)(\alpha' u - i)} \\ &= G_N \langle 13\rangle^4 [24]^4 \frac{\Gamma(-\alpha' s)\Gamma(-\alpha' t)\Gamma(-\alpha' u)}{\Gamma(1+\alpha' s)\Gamma(1+\alpha' t)\Gamma(1+\alpha' u)}\,. \end{aligned} \tag{3.315}$$

Of course this is not at all a "derivation" of the string amplitude, but it is nonetheless striking to see how easily the amplitude emerges as the simplest possible way of writing an expression with infinitely many poles that passes even the most basic consistency checks.

In fact, it is fascinating that directly checking the consistency known string tree amplitudes is high non-trivial. Causality in the form of the correct Regge behavior is readily verified, but unitarity, in the form of the positivity of the Gegenbauer expansion of the amplitude residues on massive poles, turns into a simple but highly non-trivial statement. For concreteness consider the scattering of colored massless scalars in the type I open superstring theory, where the amplitude is

$$\mathcal{A} = s^2 \frac{\Gamma(-s)\Gamma(-t)}{\Gamma(1-s-t)}. \tag{3.316}$$

The residue on the massive poles at $s = n$ is a polynomial $R_n(x = cos\theta)$, where $t = -\frac{n}{2}(1-x)$, given by

$$P_n(x) = \prod_{i=1}^{n-1}\left(x - \frac{(n-2i)}{n}\right) \tag{3.317}$$

Already at $n = 3$, we learn something striking: $P_3(x) = (x - \frac{1}{3})(x + \frac{1}{3}) = x^2 - \frac{1}{9}$, which we would like to express as a sum over Gegenbauer polynomial. The spin 2 Gegenbauer in d spatial dimension is proportional to $x^2 - \frac{1}{d}$, thus by writing $(x^2 - \frac{1}{9}) = (x^2 - \frac{1}{d}) + (\frac{1}{d} - \frac{1}{9})$, we see a massive spin 2 state with positive norm, but also a spin 0 state with norm $(\frac{1}{d} - \frac{1}{9})$, which is ≥ 0 for $d \leq 9$, but is negative, violating unitarity, for $d > 9$. Thus the critical spacetime dimension $D = d + 1 = 10$ is hiding in plain sight in the four-particle amplitude, purely from asking for unitarity at on this pole at $s = 3$. But of course for unitarity, we must have that

$$P_n(x) = \sum_s p_{n,s} G_s^{(d)}(x), \; with \; p_{n,s} \geq 0 \; for \; d \leq 9. \tag{3.318}$$

This extremely simple statement turns out to be very difficult to prove directly—indeed we are not aware of any direct proof of this fact in the literature! Of course it does follow, more indirectly, from the still rather magical proof of the no-ghost theorem in string theory.

The miraculous way in with which string amplitudes manage to be consistent make it seem even more plausible that these amplitudes emerge as the unique answer to the question of finding consistent four particle massless graviton amplitudes with only poles. But some further constraints other than causality, unitarity, and good high-energy behavior of just massless graviton scattering must be imposed to do this, as we have found candidate four-particle amplitudes satisfying all these rules that deform away from the known string amplitudes. Consider again the Virasoro-Shapiro amplitude for graviton scattering. The residue on the pole at $s = n$ is the square of the open-string residue $P_n(x)^2$, and so the positivity of its Gegenbauer expansion follows directly from the positivity of $P_n(x)$ for the open string. But now consider a deformation by a parameter ϵ of the form

$$\begin{aligned}\frac{\Gamma(-\alpha' s)\Gamma(-\alpha' t)\Gamma(-\alpha' u)}{\Gamma(1+\alpha' s)\Gamma(1+\alpha' t)\Gamma(1+\alpha' u)} \quad &\rightarrow \\ \frac{\Gamma(-\alpha' s)\Gamma(-\alpha' t)\Gamma(-\alpha' u)}{\Gamma(1+\alpha' s)\Gamma(1+\alpha' t)\Gamma(1+\alpha' u)} \quad &+ \quad \epsilon\frac{\Gamma(1-\alpha' s)\Gamma(1-\alpha' t)\Gamma(1-\alpha' u)}{\Gamma(2+\alpha' s)\Gamma(2+\alpha' t)\Gamma(2+\alpha' u)}.\end{aligned} \tag{3.319}$$

This deformed amplitude has the same Regge behavior as the usual string amplitude, and the same exponential softness for high-energy fixed-angle scattering. The residue at $s = n$ is given by

$$\frac{1+n(1-\epsilon)}{n+1}\left(\frac{n^{n-1}}{2^{n-1}n!}\right)^2\left(P_n(x)^2 + \frac{4\epsilon(n-1)}{n(1+(1-\epsilon)n)}P_n(x)P^B_{n-4}(x)\right), \tag{3.320}$$

where $P^B_n(x) \equiv \prod_{i=1}^{n+1}\left(x - \frac{n+2-2i}{n+4}\right)$ is the residue of the Veneziano amplitude. It is straightforward to see that so long as $0 < \epsilon < 1$, the positivity of $P_n(x)$ continues to imply the positivity of the Gegenbauer expansion on the massive poles. Thus this deformed expression satisfies all the constraints we have been imposing on four-particle scattering. It seems very unlikely, however, that this corresponds to amplitudes in some consistent deformation of string theory: the spectrum is exactly the same as the usual (free!) string, and there is no obvious room for an extra parameter ϵ in the quantization of the string.

Thus any claim about consistent UV completion must go beyond merely the consistency of *massless* scattering at four particles, and include consistent expressions for higher-point massless scattering and/or, relatedly, consistent amplitudes for the new massive resonances introduced in the UV completion. This is very reasonable and is after all precisely what happened in the story of the weak interactions, where the four-fermi interaction was UV completed by W particles, which in turn had bad high-energy growth for the scattering of their longitudinal modes that had to be further cured by the Higgs. It is also interesting to note that imposing just a frisson of extra string properties on the four-particle amplitude—such as the monodromy relations relating different color channels [37, 101]—when combined with the EFT-hedron constraints, *do* appear to uniquely fix string amplitudes. These observations all suggest a number of fascinating open avenues for further exploration at the intersection of unitarity, causality, analyticity, string theory, and the UV/IR connection.

3.13 Acknowledgements

We thank Alex Postnikov, Steven Karp, and Congkao Wen for stimulating discussions. We thank Zvi Bern, Alexander Zhiboedov, and Dimitrios Kosmopoulos for discussions, and for pointing out an error in the spinning residue polynomials given in the first version of the paper. N.A-H. is supported by DOE grant de-sc0009988. Y-t H is supported by MoST Grant No. 106-2628-M-002-012-MY3. T-c H is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632.

Chapter 4

THE F-SYMBOLS FOR TRANSPARENT HAAGERUP-IZUMI CATEGORIES WITH $G = \mathbb{Z}_{2n+1}$

[1]

4.1 Introduction

Subfactors [107, 108] and fusion categories [72, 73] provide the mathematical framework underlying various physical objects in quantum field theory, including anyons in (2+1)d Chern-Simons theory [109, 174] and topological defect lines in (1+1)d quantum field theory [36, 50, 160]. Fusion categories with invertible objects encapsulate the notion of symmetries and 't Hooft anomalies in quantum field theory, and those with non-invertible objects generalize such a notion[1, 50, 110, 161]. Due to Ocneanu rigidity [72, 169], a fusion category is an invariant under renormalization group flows connecting short and long distance physics. This generalization of the 't Hooft anomaly matching condition has shed new light on the phases of quantum field theory.

Subfactor theory has an inherent categorical structure [135], and has been a productive factory of fusion categories. Subfactors with Jones indices less than 4 have been classified by Ocneanu [138] and extended to 4 by Popa [145]. Haagerup [96] searched for subfactors with Jones indices *a little bit beyond 4*, and together with Asaeda [14] constructed one with Jones index $\frac{5+\sqrt{13}}{2}$, the smallest above 4. In [105], Izumi generalized the Haagerup fusion ring to a family of fusion rings labeled by a finite abelian group G, and explicitly constructed the subfactors for $G = \mathbb{Z}_3,\ \mathbb{Z}_5$. The constructive classification of subfactors for $|G|$ odd was achieved up to $|G| = 19$ by Evans and Gannon [75] (up to $|G| = 9$ with exact expressions and the rest with numerical estimates), and that of subfactors with $G = \mathbb{Z}_4,\ \mathbb{Z}_2 \times \mathbb{Z}_2,\ \mathbb{Z}_4 \times \mathbb{Z}_2,\ \mathbb{Z}_6,\ \mathbb{Z}_8,\ \mathbb{Z}_{10}$ by Grossman, Izumi, and Snyder [90–92, 104].

The fundamental data underlying a fusion category are the F-symbols, which are solutions to the pentagon identity. Some (almost) equivalent notions exist: associators, quantum $6j$-symbols, and crossing kernels. They underlie the Turaev-Viro theory [165, 166], the Levin-Wen string-net models [119], and large classes of statistical models (see [2] and references within) as well as the associated anyon chains [77]. In [50], one of the present authors showed how the F-symbols strongly constrain (1+1)d (fully extended)

[1]This chapter is adapted from Tzu-Chen Huang and Ying-Hsuan Lin. "The F-Symbols for Transparent Haagerup-Izumi Categories with $G = \mathbb{Z}_{2n+1}$". In: (July 2020). arXiv: `2007.00670 [math.CT]`.

topological quantum field theories [58, 129], which are endpoints of symmetry-preserving renormalization group flows; in many cases, given the F-symbols, the full field theory data could be completely determined by bootstrap.

In this paper, unitary and non-unitary Haagerup-Izumi fusion categories with $G = \mathbb{Z}_{2n+1}$ are constructed up to $G = \mathbb{Z}_{15}$ by computing Groebner bases for the pentagon identity. The notion of a *transparent* fusion category is introduced in Definition 4.3.1, from which various consequential graph equivalences and F-symbol relations are derived to reduce the number of independent F-symbols from $\mathcal{O}(n^6)$ to $\mathcal{O}(n^2)$ and render the pentagon identity practically solvable. These relations are summarized into a system of constraints in Definition 4.4.1, and the solutions to the pentagon identity under said constraints provide a classification of F-symbols for transparent Haagerup-Izumi fusion categories. The results of this classification are stated in Theorems 4.5.1 and 4.5.2.

Some remarks on the comparison of the present results with the existing literature are in order. As mentioned above, the datum equivalent to the F-symbols for several unitary Haagerup-Izumi fusion categories were obtained by Izumi [105], Evans and Gannon [75], and Grossman and Snyder [93] using Cuntz algebra techniques; such constructions were further generalized by Evans and Gannon [74] to fusion categories that need not be unitary. More recently, the F-symbols for all fusion categories realizing the Haagerup fusion ring ($G = \mathbb{Z}_3$) with six simple objects have been computed using the pentagon approach by Titsworth [163], and for the special case of the Haagerup $\mathcal{H}_3$ fusion category (in the nomenclature of Grossman and Snyder [93]) independently by Osborne, Stiegemann and Wolf [139].

The novelty of this paper is twofold. First, it offers the direct pentagon construction for Haagerup-Izumi fusion categories beyond the Haagerup case ($G = \mathbb{Z}_3$); in particular, the Haagerup-Izumi fusion categories classified in Theorem 4.5.2 have not appeared in the literature beyond $G = \mathbb{Z}_5$. Second, the special transparent gauge adopted in this paper—in which all F-symbols involving at least one external invertible object take value one—not only makes the independent F-symbols directly comparable to the Cuntz algebra datum of Izumi [105], Evans and Gannon [74, 75], and Grossman and Snyder [93], but also makes the F-symbols automatically tetrahedral-symmetric (A_4 or S_4 tetrahedral-invariant in the language of this paper), and unitary for pseudo-unitary fusion categories.[2]

[2] In [139, 163], the F-symbols for the Haagerup fusion categories with six simple objects were presented in non-transparent gauges that do not enjoy tetrahedral symmetry. The present authors used the Mathematica package provided by Titsworth [163] to check that the F-symbols in the present paper are indeed gauge-equivalent to his. The authors also thank Yuji Tachikawa for explicitly checking that the four sets of F-symbols in [139] are all gauge-equivalent, and also gauge-equivalent to those presented in

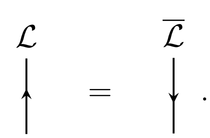
In physical applications, such a gauge satisfies the assumptions of various theoretical constructions—the Levin-Wen string-net models [119], large classes of statistical models (see [2] and references within) and the associated anyon chains [77]—and allows the more effective exploitation of the $G = \mathbb{Z}_{2n+1}$ symmetry. Of course, for a given fusion ring, there may exist non-transparent fusion categories that elude the present approach. However, none of the Haagerup-Izumi fusion categories up to $G = \mathbb{Z}_9$ known in the literature [14, 75, 90–92, 104, 105] was found to be non-transparent!

The outline of this chapter is as follows. Section 4.2 reviews the string diagram calculus, the F-symbols, and their relation to the tetrahedra. Section 4.3 defines the notion of a transparent fusion category, and derives various consequences including invariance relations for the F-symbols. Section 4.4 introduces the Haagerup-Izumi fusion rings, and formulates a set of constraints on F-symbols that must be satisfied for transparent Haagerup-Izumi fusion categories. Section 4.5 states the classification of solutions to the pentagon identity under the said constraints, and presents the explicit F-symbols for unitary Haagerup-Izumi fusion categories with S_4 tetrahedral invariance, as well as for the Haagerup $\mathcal{H}_2$ fusion category. Section 4.6 ends with some concluding remarks.

Note: The authors first obtained the F-symbols for the Haagerup fusion categories with six simple objects from Titsworth [163]. By performing gauge transformations on his solution, a gauge manifesting the transparent property was found. This observation led the present authors to postulate that transparent fusion categories also exist for the subsequent Haagerup-Izumi fusion rings with $G = \mathbb{Z}_{2n+1}$.

4.2 Preliminaries

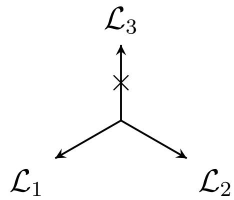
A classic introduction to fusion categories can be found in [72, 73]. The type of fusion categories considered in this chapter are pivotal fusion categories over ground field $k = \mathbb{C}$.[3] The notation for string diagrams is as follows. Each object $\mathcal{L}$ is represented by an oriented string that is equivalent to its dual $\overline{\mathcal{L}}$ with the opposite orientation,



this paper.

[3]For such categories, a physical formulation in the context of topological defect lines in $(1+1)d$ quantum field theory can be found in [50] (see also [36, 160]).

The basic building block for string diagrams is a trivalent vertex with three open edges

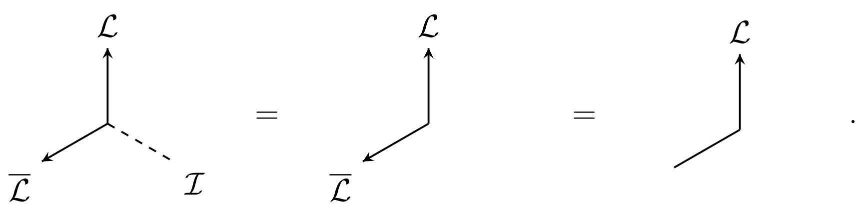


with $\times$ specifying the ordering of edges. It represents the vector space of morphisms

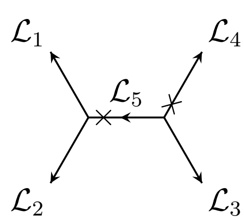
$$V_{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3} \equiv \hom(\overline{\mathcal{L}}_2 \otimes \overline{\mathcal{L}}_1, \mathcal{L}_3) \in \mathbb{C}^{N^{\mathcal{L}_3}_{\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_1}} \,,$$

where $N^{\mathcal{L}_3}_{\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_1}$ is the fusion coefficient, the multiplicity of $\mathcal{L}_3$ in $\overline{\mathcal{L}}_2 \otimes \overline{\mathcal{L}}_1$. A change of basis at this vertex is a *gauge* transformation $g_{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3} \in GL(N^{\mathcal{L}_3}_{\mathcal{L}_1,\mathcal{L}_2}, \mathbb{C})$. To simplify the discussion, it is assumed in the following that the fusion algebra is multiplicity-free, i.e. all nonzero fusion coefficients are one, and hence every nontrivial gauge factor $g_{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}$ is a complex scalar.

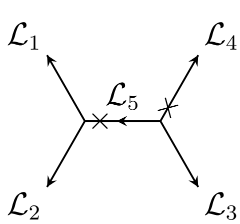
For a trivalent vertex involving at least one unit object, the ordering of edges is irrelevant, and the marking $\times$ can be dropped. Furthermore, by choosing the unitors and counitors to be identity morphisms, the unit object $\mathcal{I}$ can be removed or added at will,



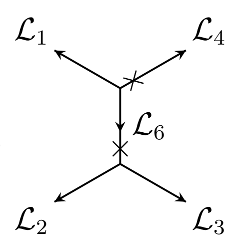
For a string diagram composed of two trivalent vertices



the gauge freedom is $g_{\mathcal{L}_1,\mathcal{L}_2,\overline{\mathcal{L}}_5}\, g_{\mathcal{L}_5,\mathcal{L}_3,\mathcal{L}_4}$. It is related by an *F-move* to a sum of string diagrams in a different configuration,



$$= \sum_{\mathcal{L}_6} (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6}$$

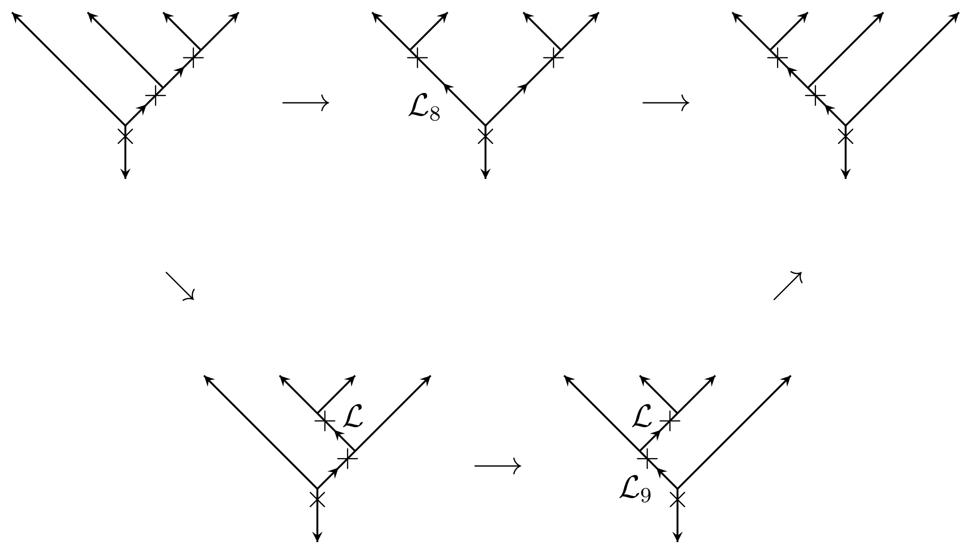


, (4.1)

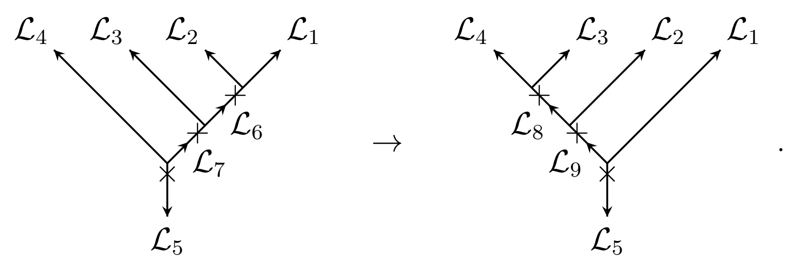
where $(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6}$ are the F-*symbols*. The gauge factor for an F-symbol is

$$\frac{g_{\mathcal{L}_1,\mathcal{L}_2,\overline{\mathcal{L}}_5}\, g_{\mathcal{L}_5,\mathcal{L}_3,\mathcal{L}_4}}{g_{\mathcal{L}_2,\mathcal{L}_3,\overline{\mathcal{L}}_6}\, g_{\mathcal{L}_1,\mathcal{L}_6,\mathcal{L}_4}} \,.$$

The F-symbols must satisfy a consistency condition that is the equivalence of the two different sequences of F-moves



both resulting in



This consistency condition is the *pentagon identity*

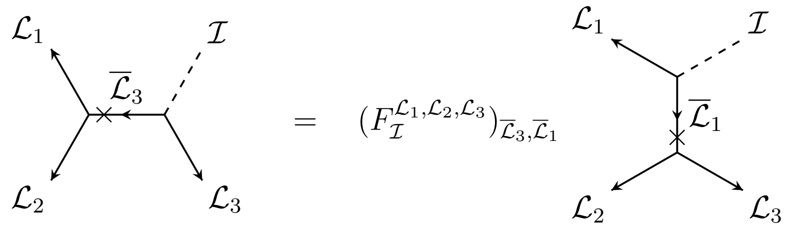
$$(F^{\mathcal{L}_6,\mathcal{L}_3,\mathcal{L}_4}_{\overline{\mathcal{L}}_5})_{\mathcal{L}_7,\mathcal{L}_8}\,(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_8}_{\overline{\mathcal{L}}_5})_{\mathcal{L}_6,\mathcal{L}_9} \;=\; \sum_{\mathcal{L}} (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_7})_{\mathcal{L}_6,\mathcal{L}}\,(F^{\mathcal{L}_1,\mathcal{L},\mathcal{L}_4}_{\overline{\mathcal{L}}_5})_{\mathcal{L}_7,\mathcal{L}_9}\,(F^{\mathcal{L}_2,\mathcal{L}_3,\mathcal{L}_4}_{\overline{\mathcal{L}}_9})_{\mathcal{L},\mathcal{L}_8}\,. \tag{4.2}$$

A solution to the pentagon identity amounts to the construction of a pivotal fusion category. If there are n isomorphism classes of simple objects, then the pentagon identity is a set of $\mathcal{O}(n^9)$ cubic polynomial equations for $\mathcal{O}(n^6)$ variables, modulo $\mathcal{O}(n^3)$ gauge freedom. As n grows, a generic system of this size quickly becomes unmanageable.

The *cyclic permutation map* is the isomorphism relating the three vector spaces

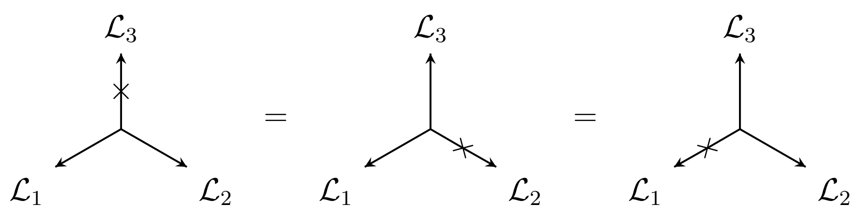
$$V_{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}\,,\quad V_{\mathcal{L}_2,\mathcal{L}_3,\mathcal{L}_1}\,,\quad V_{\mathcal{L}_3,\mathcal{L}_1,\mathcal{L}_2}\,,$$

which pictorially corresponds to moving the $\times$ mark around. It is the F-move with an external edge representing the unit object $\mathcal{I}$:



The net effect is a counter-clockwise rotation of the $\times$ mark accompanied by a factor of $(F_{\mathcal{I}}^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3})_{\overline{\mathcal{L}}_3,\overline{\mathcal{L}}_1}$. Gauge freedom alone cannot guarantee that the F-symbols $(F_{\mathcal{I}}^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3})_{\overline{\mathcal{L}}_3,\overline{\mathcal{L}}_1}$ take value one for all $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$.[4] The temptation to ignore the ordering and marking at trivalent vertices motivates the following definition.

Definition 4.2.1 (Cyclic-permutation invariance) *A pivotal fusion category is called cyclic-permutation invariant if the trivalent vertices are cyclic-permutation invariant, that is, for every triple* $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ *of objects,*[5]

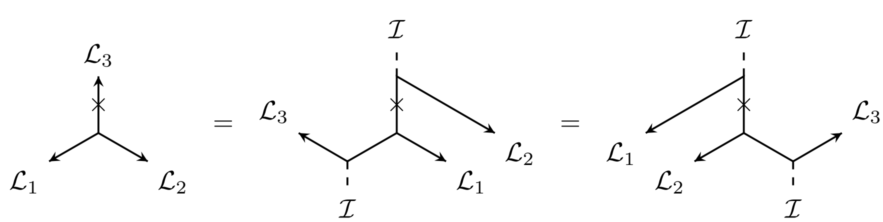


In a cyclic-permutation invariant fusion category $\mathcal{C}$, it is clear by a π-rotation that the F-symbols enjoy an order-two invariance

$$(F_{\overline{\mathcal{L}}_4}^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3})_{\mathcal{L}_5,\mathcal{L}_6} = (F_{\overline{\mathcal{L}}_2}^{\mathcal{L}_3,\mathcal{L}_4,\mathcal{L}_1})_{\overline{\mathcal{L}}_5,\overline{\mathcal{L}}_6} \, .$$

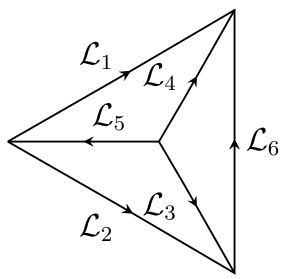
[4]See for instance Appendix A of [50].

[5]A more conventional string diagram is

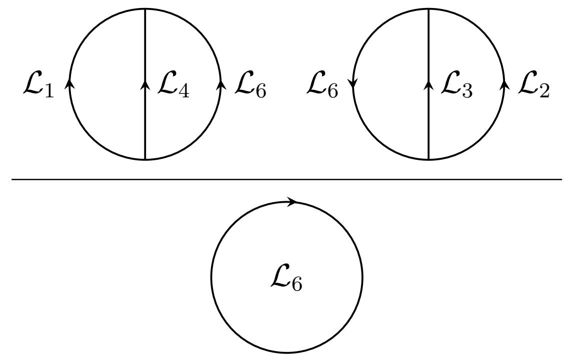


involving evaluation, coevaluation, unitor, and counitor.

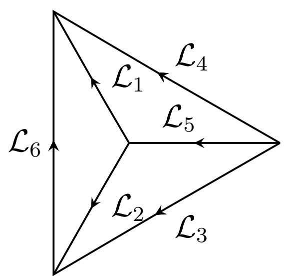
By relating the F-symbols to tetrahedra (as shown in Appendix C.1),



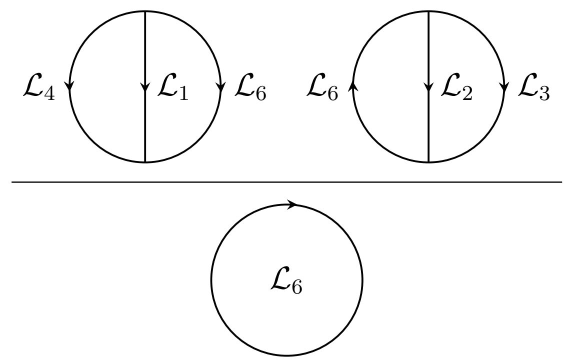
$$= \ (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6}$$



,



$$= \ (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6}$$



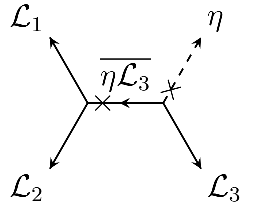
,

(4.3)

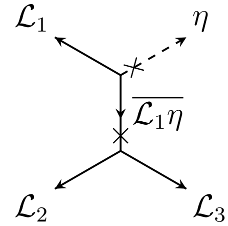
additional relations can be manifested. Each tetrahedron enjoys an S_3 symmetry: it is invariant under the $\mathbb{Z}_3$ rotations and complex conjugate under a reflection. Combined with the aforementioned π-rotation invariance generates an S_4 worth of relations for the F-symbols. However, these relations are generally nonlinear due to the factors of graphs appearing on the right of (4.3).

4.3 Transparent fusion categories

Definition 4.3.1 (Transparency) *A pivotal fusion category $\mathcal{C}$ is called transparent if the associator involving any invertible object is the identity map. In terms of string diagrams, $\mathcal{C}$ is transparent if for every triple $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ of objects in $\mathcal{C}$ and for every invertible object η,*



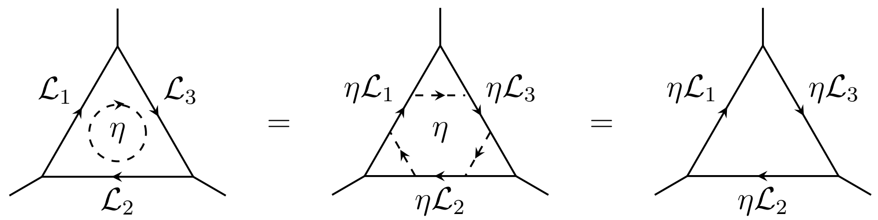
$=$



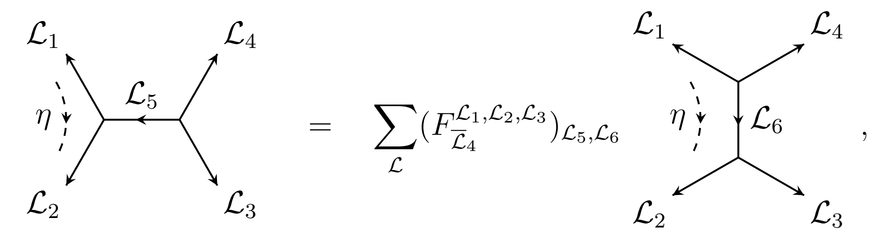
and likewise for η on any of the three other external edges.

Since the unit object is invertible, a transparent fusion category is automatically cyclic-permutation invariant. Hence, the marking $\times$ on the trivalent vertices representing the ordering or edges can be ignored.

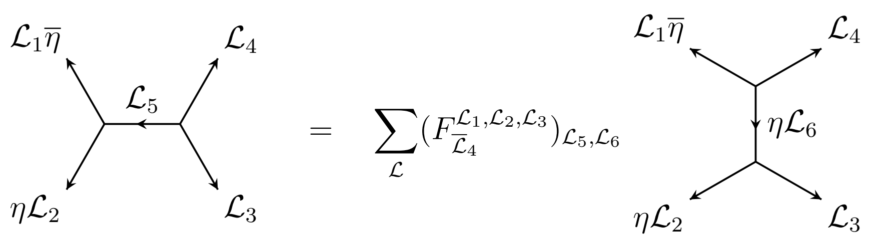
Transparency essentially means that invertible objects can be attached or detached "freely", changing the isomorphism classes of the other involved objects without generating extra F-symbols. Appendix C.2 illustrates some basic operations. The following operation will be referred to as *symmetry nucleation*: given a graph, nucleate an invertible loop on any face and merge it with the bordering edges. For example, on any triangular face,



A slight variant of symmetry nucleation gives rise to invariance relations for F-symbols. Consider the F-move equation and add an invertible object η to an open face



which by transparency is equivalent to



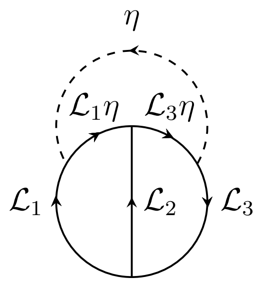
The result is an invariance relation

$$(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6} = (F^{\mathcal{L}_1\overline{\eta},\eta\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\eta\mathcal{L}_6} .$$

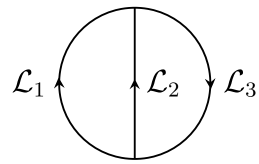
Similar operations on the other three faces give

$$(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6} = (F^{\mathcal{L}_1,\mathcal{L}_2\overline{\eta},\eta\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5\overline{\eta},\mathcal{L}_6} = (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\overline{\eta}}_{\overline{\eta\mathcal{L}_4}})_{\mathcal{L}_5,\mathcal{L}_6\overline{\eta}} = (F^{\eta\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}_4\overline{\eta}}})_{\eta\mathcal{L}_5,\mathcal{L}_6} .$$

Further useful relations between graphs and F-symbols can be derived as follows. Let $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ be any triple of simple objects in $\mathcal{C}$, and η any invertible object. Consider

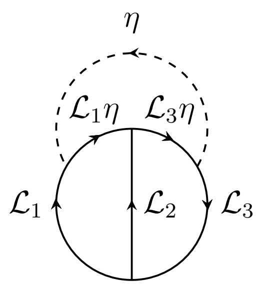


$=$

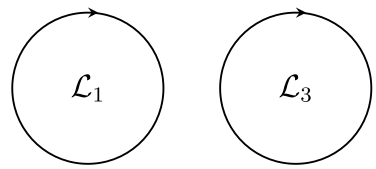


, (4.4)

and perform an F-move on $\mathcal{L}_2$ to obtain



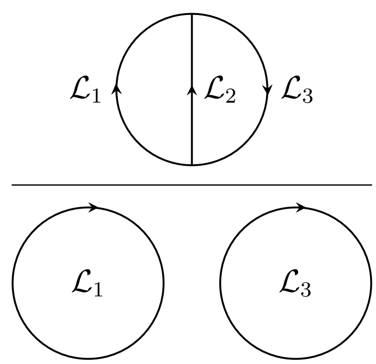
$= \quad (F^{\mathcal{L}_1,\overline{\mathcal{L}}_3,\mathcal{L}_3\eta}_{\mathcal{L}_1\eta})_{\overline{\mathcal{L}}_2,\eta}$



.

Thus

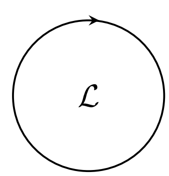
$(F^{\mathcal{L}_1,\overline{\mathcal{L}}_3,\mathcal{L}_3\eta}_{\mathcal{L}_1\eta})_{\overline{\mathcal{L}}_2,\eta} \quad =$



. (4.5)

The special case of $\mathcal{L}_2 = \theta$ invertible, and $\mathcal{L}_1 = \mathcal{L},\ \mathcal{L}_3 = \theta\mathcal{L}$ gives

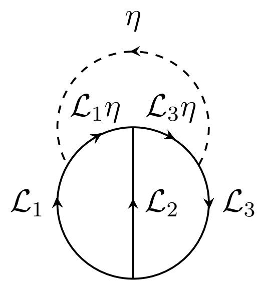
$(F^{\mathcal{L},\overline{\theta\mathcal{L}},\theta\mathcal{L}\eta}_{\mathcal{L}\eta})_{\overline{\theta},\eta}{}^{-1} \quad =$



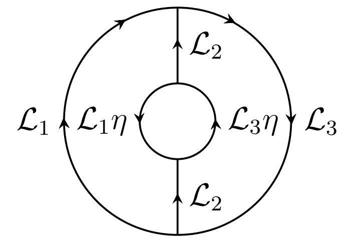
. (4.6)

Consider again the original diagram (4.4). Perform an F-move on η, and then another

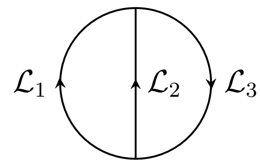
F-move on a unit object connecting the two $\mathcal{L}_2$ edges to obtain

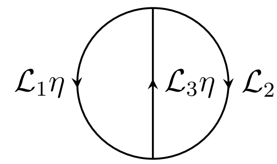


$$= \quad (F^{\overline{\mathcal{L}}_1,\mathcal{L}_1\eta,\overline{\mathcal{L}_3\eta}}_{\overline{\mathcal{L}}_3})_{\eta,\overline{\mathcal{L}}_2}$$



$$= \quad (F^{\overline{\mathcal{L}}_1,\mathcal{L}_1\eta,\overline{\mathcal{L}_3\eta}}_{\overline{\mathcal{L}}_3})_{\eta,\overline{\mathcal{L}}_2} \quad (F^{\mathcal{L}_2,\overline{\mathcal{L}}_2,\mathcal{L}_2}_{\overline{\mathcal{L}}_2})_{\mathcal{I},\mathcal{I}}$$

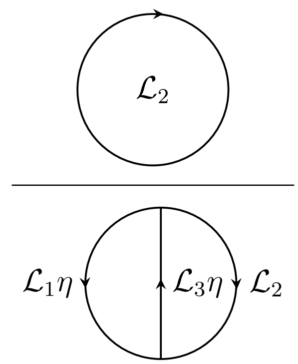




.

From the above it is deduced that

$$(F^{\overline{\mathcal{L}}_1,\mathcal{L}_1\eta,\overline{\mathcal{L}_3\eta}}_{\overline{\mathcal{L}}_3})_{\eta,\overline{\mathcal{L}}_2} \quad = \tag{4.7}$$



.

4.4 Transparent Haagerup-Izumi fusion categories

A Haagerup-Izumi fusion ring can be defined for every finite abelian group G. A key feature is that it is quadratic [91, 162]: the fusion of a single non-invertible simple object with the invertible objects generates all the non-invertible simple objects. In this section, a set of constraints are formulated for classifying transparent Haagerup-Izumi fusion categories with $G = \mathbb{Z}_{2n+1}$.

The Haagerup-Izumi fusion ring with $G = \mathbb{Z}_\nu$ has ν invertible objects

$$\mathcal{I}, \quad \alpha, \quad \alpha^2, \quad \cdots \quad \alpha^{\nu-1}$$

and ν non-invertible simple objects

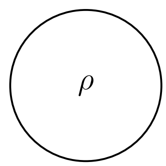
$$\rho, \quad \alpha\rho, \quad \alpha^2\rho, \quad \cdots \quad \alpha^{\nu-1}\rho\,,$$

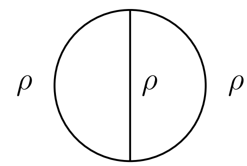
subject to the relations

$$\alpha^\nu = 1\,, \quad \alpha\rho = \rho\,\alpha^{\nu-1}\,, \quad \rho^2 = \mathcal{I} + \sum_{k=0}^{\nu-1} \alpha^k \rho\,.$$

When $\nu = 1$, this is the Fibonacci ring, which is the Grothendieck ring of the Fibonacci category (even sectors of the A_4 subfactor) and Lee-Yang category. When $\nu = 2$, this is the Grothendieck ring of the $\mathcal{C}(sl(2), 8)_{ad}$ fusion category (even sectors of the A_7 subfactor), which is premodular but not modular [43]. When $\nu = 3$, this is the Grothendieck ring of the Haagerup $\mathcal{H}_2$ and $\mathcal{H}_3$ fusion categories [14, 93]. For $\nu \geq 3$, the fusion ring is non-commutative.

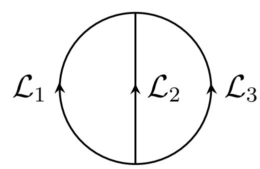
Let $\mathcal{C}$ be a transparent Haagerup-Izumi fusion category with $G = \mathbb{Z}_{2n+1}$. Define ζ and ξ to be the graph values

$\zeta \equiv$ 

, $\quad \xi \equiv$ 

.

On the left, symmetry nucleation implies that all non-invertible loops take value ζ. On the right, symmetry nucleation on the three faces implies that all such graphs with three non-invertible simple objects take the same value ξ. In summary, for any triple $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ of simple objects,



$$= \begin{cases} 1 & \text{all invertible}\,, \\ \zeta & \text{one invertible}\,, \\ \xi & \text{all non-invertible}\,. \end{cases}$$

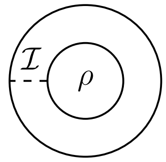
By (4.6), for any pair $(\eta,\, \theta)$ of invertible objects,

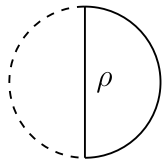
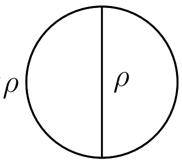
$$(F^{\mathcal{L}, \overline{\theta \mathcal{L}}, \theta \mathcal{L} \eta}_{\mathcal{L} \eta})_{\overline{\theta}, \eta} \;=\; \begin{cases} 1 & \mathcal{L} \ \text{invertible}\,, \\ \zeta^{-1} & \mathcal{L} \ \text{non-invertible}\,. \end{cases}$$

The F-symbols with a single internal invertible object can also be deduced. For any triple $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ of non-invertible simple objects, by (4.5) and (4.7),

$$(F^{\mathcal{L}_1, \mathcal{L}_3, \eta \mathcal{L}_3}_{\mathcal{L}_1 \overline{\eta}})_{\mathcal{L}_2, \overline{\eta}} \;=\; \zeta^{-2}\, \xi\,, \qquad (F^{\mathcal{L}_1, \eta \mathcal{L}_1, \eta \mathcal{L}_3}_{\mathcal{L}_3})_{\overline{\eta}, \mathcal{L}_2} \;=\; \zeta\, \xi^{-1}\,. \tag{4.8}$$

The possible values of ζ can be constrained as follows. Consider two concentric ρ loops and perform an F-move to obtain

$\zeta^2 \;=\; \rho$ 

$= \;(F^{\rho,\rho,\rho}_{\rho})_{\mathcal{I},\mathcal{I}}\;\; \mathcal{I}$  $\rho \;+\; \sum_{i=0}^{2n} (F^{\rho,\rho,\rho}_{\rho})_{\mathcal{I},\alpha^i\rho}\;\; \alpha^i\rho$  ρ

$= \;1 + (2n+1)\,\zeta\,.$

Hence,

$$\zeta \;=\; \frac{2n+1 \pm \sqrt{(2n+1)^2+4}}{2}\,.$$

Finally, a gauge choice can be made such that

$$\xi \;=\; \zeta^{\frac{3}{2}}\,, \qquad (F^{\mathcal{L}_1,\mathcal{L}_3,\eta\mathcal{L}_3}_{\overline{\eta\mathcal{L}_1}})_{\mathcal{L}_2,\overline{\eta}} \;=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_1,\eta\mathcal{L}_3}_{\overline{\mathcal{L}_3}})_{\overline{\eta},\mathcal{L}_2} \;=\; \zeta^{-\frac{1}{2}}\,.$$

Definition 4.4.1 (Transparent constraints) *Let I be the set of isomorphism classes of invertible objects and N the set of isomorphism classes of non-invertible simple objects in the Haagerup-Izumi fusion ring with $G = \mathbb{Z}_{2n+1}$. The transparent constraints are the collection of constraints on the F-symbols*

$$\begin{aligned}
(F^{\eta,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;&=\; (F^{\mathcal{L}_1,\eta,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;=\; (F^{\mathcal{L}_1,\mathcal{L}_2,\eta}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;=\; (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\eta}})_{\mathcal{L}_5,\mathcal{L}_6} \;=\; 1\,,\\
(F^{\eta\mathcal{L}\theta,\overline{\theta}\mathcal{L},\mathcal{L}}_{\eta\mathcal{L}})_{\eta,\overline{\theta}} \;&=\; \zeta^{-1}\,, \qquad (F^{\mathcal{L}_1,\mathcal{L}_3,\eta\mathcal{L}_3}_{\mathcal{L}_1\overline{\eta}})_{\mathcal{L}_2,\overline{\eta}} \;=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_1,\eta\mathcal{L}_3}_{\mathcal{L}_3})_{\overline{\eta},\mathcal{L}_2} \;=\; \zeta^{-\frac{1}{2}}\,,\\
(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;&=\; (F^{\mathcal{L}_1\overline{\eta},\eta\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\eta\mathcal{L}_6} \;=\; (F^{\mathcal{L}_1,\mathcal{L}_2\overline{\eta},\eta\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5\overline{\eta},\mathcal{L}_6}\\
&=\; (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\overline{\eta}}_{\mathcal{L}_4\overline{\eta}})_{\mathcal{L}_5,\mathcal{L}_6\overline{\eta}} \;=\; (F^{\eta\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\eta\mathcal{L}_4})_{\eta\mathcal{L}_5,\mathcal{L}_6}\,,
\end{aligned} \tag{4.9}$$

for all $\eta, \theta \in I$ and $\mathcal{L}, \mathcal{L}_i \in N$.

For the Haagerup-Izumi fusion ring with $G = \mathbb{Z}_{2n+1}$, the number of independent F-symbols after imposing the transparent constraints is $(2n+1)^2+1$, significantly reduced from $\mathcal{O}(n^6)$. This number can be further reduced by exploiting tetrahedral invariance. Since the factors in the relations (4.3) between the tetrahedra and the F-symbols are universally equal to $\zeta^{-1}\xi^2$, the set of F-symbols with all objects non-invertible are *invariant* under the A_4 symmetry of the tetrahedron, and are related by complex conjugation under reflection if ξ is chosen to be real. To facilitate the computation, one may further assume reflection invariance and impose S_4 invariance on the F-symbols.[6]

Table 4.1 lists the numbers of independent F-symbols after imposing the transparent constraints together with A_4 or S_4 tetrahedral invariance. With A_4 invariance (necessary consequence of transparency), the pentagon identity under the transparent constraints can be practically solved up to $G = \mathbb{Z}_9$ by computing a Groebner basis using MAGMA [38]. With S_4 invariance, it can be solved up to $G = \mathbb{Z}_{15}$. The next section presents the results of this classification.

G	A_4	S_4
$\mathbb{Z}_3$	8	7
$\mathbb{Z}_5$	22	16
$\mathbb{Z}_7$	44	29
$\mathbb{Z}_9$	74	46
$\mathbb{Z}_{11}$	112	67
$\mathbb{Z}_{13}$	158	92
$\mathbb{Z}_{15}$	212	121

Table 4.1: The numbers of independent F-symbols for the Haagerup-Izumi fusion rings after imposing the transparent constraints together with A_4 or S_4 tetrahedral invariance.

4.5 Classification of F-symbols

Main theorems

Theorem 4.5.1 *For the Haagerup-Izumi fusion rings with $G = \mathbb{Z}_{2n+1}$, let*

$$\zeta_\pm \equiv \frac{2n+1 \pm \sqrt{(2n+1)^2+4}}{2} .$$

Under the transparent constraints (4.4.1) *and imposing A_4 tetrahedral invariance (necessary by transparency), the pentagon identity has the following solutions:*

(a) There are two solutions for $G = \mathbb{Z}_1$ corresponding to the Fibonacci and Lee-Yang categories.

(b) There are eight solutions for $G = \mathbb{Z}_3$.

(c) There are sixteen solutions for $G = \mathbb{Z}_5$.

[6]The usual notion of tetrahedral symmetry includes complex conjugation under reflections. However, such relations complicate the present approach of computing a Groebner basis for the pentagon identity. Hence, the term tetrahedral invariance in this paper refers to true equality without complex conjugation, and the notion is further subdivided into A_4 invariance (without reflections) and S_4 invariance (with reflections).

(d) *There are twenty-four solutions for* $G = \mathbb{Z}_7$.

(e) *There are forty-eight solutions for* $G = \mathbb{Z}_9$.

(f) *For* $G = \mathbb{Z}_{2n+1}$ *with* $n = 1, 2, 3$, *the solutions form four order-*$2n$ *orbits of the* $\mathbb{Z}_{2n}$ *automorphism group. Two orbits are unitary with* $\zeta = \zeta_+$; *the* F*-symbols are real in one of the two orbits, and complex in the other. The remaining two orbits are the non-unitary Galois associates of the two unitary orbits, with* $\zeta = \zeta_-$. *In particular, for* $G = \mathbb{Z}_3$, *the unitary real orbit corresponds to the Haagerup* $\mathcal{H}_3$ *fusion category, and the unitary complex orbit corresponds to the Haagerup* $\mathcal{H}_2$ *fusion category, in the nomenclature of Grossman and Snyder [93].*

Theorem 4.5.2 *For the Haagerup-Izumi fusion rings with* $G = \mathbb{Z}_{2n+1}$, *let*

$$\zeta_\pm \equiv \frac{2n+1 \pm \sqrt{(2n+1)^2+4}}{2}.$$

Under the transparent constraints (4.4.1) *and imposing* S_4 *tetrahedral invariance, the pentagon identity has the following solutions:*

(a) *There are two solutions for* $G = \mathbb{Z}_1$, *corresponding to the Fibonacci and Lee-Yang categories.*

(b) *There are four solutions for* $G = \mathbb{Z}_3$.

(c) *There are eight solutions for* $G = \mathbb{Z}_5$.

(d) *There are twelve solutions for* $G = \mathbb{Z}_7$.

(e) *There are twenty-four solutions for* $G = \mathbb{Z}_{13}$.

(f) *For* $G = \mathbb{Z}_{2n+1}$ *with* $n = 1, 2, 3, 6$, *the solutions form two order-*$2n$ *orbits of the* $\mathbb{Z}_{2n}$ *automorphism group. One orbit is unitary with* $\zeta = \zeta_+$, *and the other orbit consists of the non-unitary Galois associates with* $\zeta = \zeta_-$. *In particular, for* $G = \mathbb{Z}_3$, *the unitary real orbit corresponds to the Haagerup* $\mathcal{H}_3$ *fusion category in the nomenclature of Grossman and Snyder [93].*

(g) *There are twenty-four solutions for* $G = \mathbb{Z}_9$, *forming four order-six orbits of the* $\mathbb{Z}_6$ *automorphism group. Two orbits are unitary with* $\zeta = \zeta_+$, *and the other two orbits consist of the non-unitary Galois associates with* $\zeta = \zeta_-$.

(h) *There are twenty-four solutions for $G = \mathbb{Z}_{11}$, forming two order-two orbits and two order-ten orbits of the $\mathbb{Z}_{10}$ automorphism group. One order-two orbit and one order-ten orbit are unitary with $\zeta = \zeta_+$, and the other two orbits consist of the non-unitary Galois associates with $\zeta = \zeta_-$.*

(i) *There are forty-eight solutions for $G = \mathbb{Z}_{15}$, forming six order-eight orbits of the $\mathbb{Z}_2 \times \mathbb{Z}_4$ automorphism group. Three orbits are unitary with $\zeta = \zeta_+$, and the other three orbits consist of the non-unitary Galois associates with $\zeta = \zeta_-$.*

(j) *In the above, the F-symbols are real when $\zeta = \zeta_+$, and complex when $\zeta = \zeta_-$. Solutions in a single orbit of the automorphism group have the same $(F^{\rho,\rho,\rho}_{\rho})_{\rho,\rho}$, while different orbits have distinct $(F^{\rho,\rho,\rho}_{\rho})_{\rho,\rho}$. Since $(F^{\rho,\rho,\rho}_{\rho})_{\rho,\rho}$ is gauge-invariant, solutions with distinct values correspond to inequivalent fusion categories.*

Explicit F-symbols for $G = \mathbb{Z}_{2n+1}$ with $1 \leq n \leq 7$

Let I be the set of invertible objects and N the set of non-invertible simple objects of the Haagerup-Izumi fusion ring with $G = \mathbb{Z}_{2n+1}$. By (4.4.1), the F-symbols involving at least one invertible object are given by

$$(F^{\eta\mathcal{L}\theta,\overline{\mathcal{L}\theta},\mathcal{L}}_{\eta\mathcal{L}})_{\eta,\bar{\theta}} \;=\; \zeta^{-1}\,, \qquad (F^{\mathcal{L}_1,\mathcal{L}_3,\eta\mathcal{L}_3}_{\overline{\eta\mathcal{L}_1}})_{\mathcal{L}_2,\overline{\eta}} \;=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_1,\eta\mathcal{L}_3}_{\overline{\mathcal{L}}_3})_{\overline{\eta},\mathcal{L}_2} \;=\; \zeta^{-\frac{1}{2}}\,,$$

for all $\eta, \theta \in I$ and $\mathcal{L}_i \in N$. For the F-symbols with all simple objects being non-invertible, it suffices to specify the $(2n+1)^2$ components $(F^{\rho,\rho,\rho}_{*})_{\rho,*}$ with $*$ running over the non-invertible simple objects. The rest are equal to one of the above by the $\mathbb{Z}^4_{2n+1}$ invariance relations in (4.4.1). In fact, these invariance relations can be equivalently written as

$$\begin{aligned}(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;&=\; (F^{\eta\mathcal{L}_1,\eta\mathcal{L}_2,\eta\mathcal{L}_3}_{\eta\mathcal{L}_4})_{\eta\mathcal{L}_5,\eta\mathcal{L}_6} \;=\; (F^{\eta\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\eta}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \\ &=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4\eta})_{\mathcal{L}_5,\mathcal{L}_6} \;=\; (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\eta\mathcal{L}_5,\mathcal{L}_6\eta}\,,\end{aligned}$$

for all $\eta, \theta \in I$ and $\mathcal{L}_i \in N$.[7]

The explicit F-symbols for the Haagerup $\mathcal{H}_2$ fusion category will first be presented, corresponding to two of the eight solutions in Theorem 4.5.1(b) that are unitary and complex (hence strictly A_4 tetrahedral invariance but not S_4). Then, among the solutions classified by Theorem 4.5.2 (satisfying S_4 tetrahedral invariance), the explicit F-symbols

[7] Note that the equality of the first and the last terms implies that every $(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6}$ with $\mathcal{L}_5$, $\mathcal{L}_6 \in N$ is anti-circulant and therefore symmetric. Together with the gauge choice in (4.4.1), it follows that every F-symbol is symmetric in the appropriate basis.

for the unitary ones—the real ones with $\zeta = \zeta_+$—will now be presented.[8] The F-symbols for the other fusion categories are available by request. In the following, except for the particularly nice ones, the F-symbols will be presented as roots of polynomials, where x_i denotes the i-th smallest real root of some polynomial in x, and likewise for other symbols y, z, ... Note that all the presented polynomials only have simple roots. The simpler polynomials are given in the main text, while the more complicated ones are given in Appendix C.3.

Haagerup $\mathcal{H}_2$ ($G = \mathbb{Z}_3$)

Theorem 4.5.1(b). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_2$, there is exactly one unitary orbit with two complex solutions (strict A_4 tetrahedral invariance). One solution is given by

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$
ρ	x	z	z
$\alpha\rho$	z	z	y_1
$\alpha^2\rho$	z	y_2	z

where

$$x = \frac{7-\sqrt{13}}{6}, \quad y_{1,2} = \frac{1}{12}\left(-1+\sqrt{13} \pm 3i\sqrt{2\left(1+\sqrt{13}\right)}\right), \quad z = \frac{1-\sqrt{13}}{6}.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_2$ exchanges y_1 and y_2, giving the other solution in the orbit.

Haagerup $\mathcal{H}_3$ ($G = \mathbb{Z}_3$)

Theorem 4.5.2(b). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_2$, there is exactly one unitary orbit with two real solutions (S_4 tetrahedral invariance). One solution is given by

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$
ρ	x	y_1	y_2
$\alpha\rho$	y_1	y_2	z
$\alpha^2\rho$	y_2	z	y_1

[8]The unitarity of the F-symbols for these solutions can be understood as follows. First, by transparency, $F_{\mathcal{L}_4}^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}$ is nontrivial only if $\mathcal{L}_1, \ldots, \mathcal{L}_4$ are all non-invertible and self-dual, so $F_{\overline{\mathcal{L}}_4}^{\overline{\mathcal{L}}_1,\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_3} = F_{\mathcal{L}_4}^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}$. Next, S_4 invariance and reality imply that $F_{\mathcal{L}_1}^{\mathcal{L}_2,\mathcal{L}_3,\mathcal{L}_4} = (F_{\overline{\mathcal{L}}_4}^{\overline{\mathcal{L}}_1,\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_3})^t = (F_{\overline{\mathcal{L}}_4}^{\overline{\mathcal{L}}_1,\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_3})^\dagger$. Hence unitarity $F_{\overline{\mathcal{L}}_4}^{\overline{\mathcal{L}}_1,\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_3}(F_{\overline{\mathcal{L}}_4}^{\overline{\mathcal{L}}_1,\overline{\mathcal{L}}_2,\overline{\mathcal{L}}_3})^\dagger = I$ becomes equivalent to the condition $F_{\mathcal{L}_4}^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3} F_{\mathcal{L}_1}^{\mathcal{L}_2,\mathcal{L}_3,\mathcal{L}_4} = I$, which is the pentagon identity (4.2) with $\mathcal{L}_5 = \mathcal{I}$, $\mathcal{L}_7 = \mathcal{L}_4$, and $\mathcal{L}_9 = \mathcal{L}_1$.

where

$$x = \frac{2-\sqrt{13}}{3}\,, \quad y_{1,2} = \frac{1}{12}\left(5-\sqrt{13}\mp\sqrt{6\left(1+\sqrt{13}\right)}\right)\,, \quad z = \frac{1+\sqrt{13}}{6}\,.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_2$ exchanges y_1 and y_2, giving the other solution in the orbit.

$G = \mathbb{Z}_5$

Theorem 4.5.2(c). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_4$, there is exactly one unitary orbit with four real solutions (S_4 tetrahedral invariance). One solution is given by

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$
ρ	x	y_1	y_3	y_2	y_4
$\alpha\rho$	y_1	y_4	z_2	z_4	z_2
$\alpha^2\rho$	y_3	z_2	y_2	z_4	z_4
$\alpha^3\rho$	y_2	z_4	z_4	y_3	z_2
$\alpha^4\rho$	y_4	z_2	z_4	z_2	y_1

where

$$x = \frac{7-\sqrt{29}}{5}\,,$$

y_i are the real roots of

$$P_y^{\mathbb{Z}_5}(y) = 625y^8 - 1375y^7 + 1275y^6 + 245y^5 - 654y^4 + 152y^3 + 75y^2 - 29y - 1\,,$$

and z_i are the roots of

$$P_z^{\mathbb{Z}_5}(z) = 25z^4 - 15z^3 - 9z^2 + 7z - 1\,.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_4$ permutes y_i and exchanges z_2 and z_4 by

$$\tau_y = (1243)\,, \quad \tau_z = (24)\,,$$

giving the other solutions in the orbit.

Note that the polynomial in z factorizes over $\mathbb{Q}(\sqrt{29 = 5^2+4})$, and z_2, z_4 are the roots to one of the factors. This pattern continues in the following solutions. Namely, all polynomials factorize over $\mathbb{Q}(\sqrt{n^2+4})$, and the roots in a single orbit of $\mathrm{Aut}(G)$ will always be roots of the same factor.

$G = \mathbb{Z}_7$

Theorem 4.5.2(d). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_6$, there is exactly one unitary orbit with six solutions. One solution is given by

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$	$\alpha^5\rho$	$\alpha^6\rho$
ρ	x	y_1	y_2	y_6	y_4	y_3	y_5
$\alpha\rho$	y_1	y_5	z_6	w_2	z_3	w_1	z_6
$\alpha^2\rho$	y_2	z_6	y_3	w_1	z_4	z_4	w_2
$\alpha^3\rho$	y_6	w_2	w_1	y_4	z_3	z_4	z_3
$\alpha^4\rho$	y_4	z_3	z_4	z_3	y_6	w_2	w_1
$\alpha^5\rho$	y_3	w_1	z_4	z_4	w_2	y_2	z_6
$\alpha^6\rho$	y_5	z_6	w_2	z_3	w_1	z_6	y_1

where

$$x = \frac{11 - 2\sqrt{53}}{7}\,.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_6 \cong \langle \sigma, \tau \mid \sigma^2 = \tau^3 = 1 \rangle$ permutes the roots by

$$\sigma_y = (15)(23)(46)\,, \quad \sigma_z = \mathrm{id}\,, \quad \sigma_w = (12)\,,$$
$$\tau_y = (356)(142)\,, \quad \tau_z = (346)\,, \quad \tau_w = \mathrm{id}\,,$$

giving the other solutions in the orbit.

$G = \mathbb{Z}_9$

Theorem 4.5.2(g). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_6$, there are two unitary orbits each with six solutions. A solution in one orbit is given by

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$	$\alpha^5\rho$	$\alpha^6\rho$	$\alpha^7\rho$	$\alpha^8\rho$
ρ	x_1	y_1	y_{12}	r_4	y_6	y_8	r_1	y_7	y_5
$\alpha\rho$	y_1	y_5	z_8	w_{10}	w_2	z_{11}	w_5	w_7	z_8
$\alpha^2\rho$	y_{12}	z_8	y_7	w_7	z_4	w_9	w_1	z_4	w_{10}
$\alpha^3\rho$	r_4	w_{10}	w_7	r_1	w_5	w_1	s_4	w_9	w_2
$\alpha^4\rho$	y_6	w_2	z_4	w_5	y_8	z_{11}	w_9	w_1	z_{11}
$\alpha^5\rho$	y_8	z_{11}	w_9	w_1	z_{11}	y_6	w_2	z_4	w_5
$\alpha^6\rho$	r_1	w_5	w_1	s_4	w_9	w_2	r_4	w_{10}	w_7
$\alpha^7\rho$	y_7	w_7	z_4	w_9	w_1	z_4	w_{10}	y_{12}	z_8
$\alpha^8\rho$	y_5	z_8	w_{10}	w_2	z_{11}	w_5	w_7	z_8	y_1

where

$$x_{1,2} = \frac{35 - 4\sqrt{85} \mp \sqrt{517 - 56\sqrt{85}}}{18}$$

are the two negative roots of

$$P_x^{\mathbb{Z}_9}(x) = 81x^4 - 630x^3 + 899x^2 + 210x + 9\,.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_6 \cong \langle \sigma, \tau \mid \sigma^2 = \tau^3 = 1 \rangle$ permutes the roots by

$$\begin{aligned}
&\sigma_x = \mathrm{id}, \quad \sigma_y = (1\ 5)(2\ 4)(3\ 11)(6\ 8)(7\ 12)(9\ 10)\,, \quad \sigma_z = \mathrm{id}\,,\\
&\sigma_w = (1\ 9)(2\ 5)(3\ 8)(4\ 12)(6\ 11)(7\ 10)\,, \quad \sigma_r = (1\ 4)(2\ 3)\,, \quad \sigma_s = \mathrm{id}\,,\\
&\tau_x = \mathrm{id}, \quad \tau_y = (1\ 6\ 7)(2\ 3\ 9)(4\ 11\ 10)(5\ 8\ 12)\,, \quad \tau_z = (3\ 10\ 7)(4\ 8\ 11)\,,\\
&\tau_w = (1\ 7\ 2)(3\ 6\ 12)(4\ 8\ 11)(5\ 9\ 10)\,, \quad \tau_r = \mathrm{id}, \quad \tau_s = \mathrm{id}\,,
\end{aligned}$$

giving the other solutions in the orbit. There is an additional $\mathbb{Z}_2 \cong \langle \iota \mid \iota^2 = 1 \rangle$ action that acts by

$$\begin{aligned}
&\iota_x = (1\ 2)\,, \quad \iota_y = (1\ 2)(3\ 6)(4\ 5)(7\ 9)(8\ 11)(10\ 12)\,, \quad \iota_z = (3\ 4)(7\ 11)(8\ 10)\,,\\
&\iota_w = (1\ 12)(2\ 6)(3\ 7)(4\ 9)(5\ 11)(8\ 10)\,, \quad \iota_r = (1\ 3)(2\ 4)\,, \quad \iota_s = (2\ 4)\,,
\end{aligned}$$

and exchanges the two unitary orbits.

$G = \mathbb{Z}_{11}$

Theorem 4.5.2(h). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_{10}$, there is one unitary orbit with two solutions and one unitary orbit with ten solutions. In the orbit with two solutions, one solution is given by

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$	$\alpha^5\rho$	$\alpha^6\rho$	$\alpha^7\rho$	$\alpha^8\rho$	$\alpha^9\rho$	$\alpha^{10}\rho$
ρ	x	y_1	y_2	y_1	y_1	y_1	y_2	y_2	y_2	y_1	y_2
$\alpha\rho$	y_1	y_2	z_2	w_2	w_2	w_1	z_2	w_2	w_1	w_1	z_2
$\alpha^2\rho$	y_2	z_2	y_1	w_1	z_2	w_2	w_1	w_2	w_1	z_2	w_2
$\alpha^3\rho$	y_1	w_2	w_1	y_2	w_1	w_1	z_2	z_2	z_2	w_2	w_2
$\alpha^4\rho$	y_1	w_2	z_2	w_1	y_2	w_2	w_2	z_2	z_2	w_1	w_1
$\alpha^5\rho$	y_1	w_1	w_2	w_1	w_2	y_2	z_2	w_1	z_2	w_2	z_1
$\alpha^6\rho$	y_2	z_2	w_1	z_2	w_2	z_2	y_1	w_1	w_2	w_1	w_2
$\alpha^7\rho$	y_2	w_2	w_2	z_2	z_2	w_1	w_1	y_1	w_2	z_2	w_1
$\alpha^8\rho$	y_2	w_1	w_1	z_2	z_2	z_2	w_2	w_2	y_1	w_2	w_1
$\alpha^9\rho$	y_1	w_1	z_2	w_2	w_1	w_2	w_1	z_2	w_2	y_2	z_2
$\alpha^{10}\rho$	y_2	z_2	w_2	w_2	w_1	z_1	w_2	w_1	w_1	z_2	y_1

where

$$x = \frac{13 - 5\sqrt{5}}{11}$$

is a root of the polynomial

$$P^{\mathbb{Z}_{11}}_{2|x}(x) = 11x^2 - 26x + 4\,.$$

The $\mathbb{Z}_2$ subgroup of $\mathrm{Aut}(G) \cong \mathbb{Z}_{10}$ exchanges y_1 with y_2 and w_1 with w_2. In the order-ten orbit, one solution is given by

$(F^{\rho,\rho,\rho}_*)_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$	$\alpha^5\rho$	$\alpha^6\rho$	$\alpha^7\rho$	$\alpha^8\rho$	$\alpha^9\rho$	$\alpha^{10}\rho$
ρ	x	y_1	y_{10}	y_9	y_2	y_8	y_3	y_7	y_4	y_6	y_5
$\alpha\rho$	y_1	y_5	z_6	w_{10}	w_3	w_9	z_7	w_1	w_7	w_4	z_6
$\alpha^2\rho$	y_{10}	z_6	y_6	w_4	z_3	w_2	w_6	w_5	w_8	z_3	w_{10}
$\alpha^3\rho$	y_9	w_{10}	w_4	y_4	w_7	w_8	z_4	z_8	z_4	w_2	w_3
$\alpha^4\rho$	y_2	w_3	z_3	w_7	y_7	w_1	w_5	z_8	z_8	w_6	w_9
$\alpha^5\rho$	y_8	w_9	w_2	w_8	w_1	y_3	z_7	w_6	z_4	w_5	z_7
$\alpha^6\rho$	y_3	z_7	w_6	z_4	w_5	z_7	y_8	w_9	w_2	w_9	w_1
$\alpha^7\rho$	y_7	w_1	w_5	z_8	z_8	w_6	w_9	y_2	w_3	z_3	w_7
$\alpha^8\rho$	y_4	w_7	w_8	z_4	z_8	z_4	w_2	w_3	y_9	w_{10}	w_4
$\alpha^9\rho$	y_6	w_4	z_3	w_2	w_6	w_5	w_9	z_3	w_{10}	y_{10}	z_6
$\alpha^{10}\rho$	y_5	z_6	w_{10}	w_3	w_9	z_7	w_1	w_7	w_4	z_6	y_1

where

$$x = \frac{101 - 49\sqrt{5}}{22}$$

is a root of the polynomial

$$P^{\mathbb{Z}_{11}}_{10|x}(x) = 11x^2 - 101x - 41\,.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_{10} \cong \langle \sigma, \tau \mid \sigma^2 = \tau^5 = 1 \rangle$ permutes the roots by

$$\sigma_y = (1\ 5)(2\ 7)(3\ 8)(4\ 9)(6\ 10)\,, \quad \sigma_z = \mathrm{id}\,, \quad \sigma_w = (1\ 9)(2\ 8)(3\ 7)(4\ 10)(5\ 6)\,,$$

$$\tau_y = (1\ 2\ 8\ 6\ 9)(3\ 10\ 4\ 5\ 7)\,, \quad \tau_z = (3\ 4\ 6\ 8\ 7)\,, \quad \tau_w = (1\ 5\ 2\ 10\ 3)(4\ 7\ 9\ 6\ 8)\,,$$

giving the other solutions in the orbit.

$G = \mathbb{Z}_{13}$

Theorem 4.5.2(e). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_{12}$, there is exactly one unitary orbit with twelve solutions. One solution is given by

$(F^{\rho,\rho,\rho}_{*})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$	$\alpha^5\rho$	$\alpha^6\rho$	$\alpha^7\rho$	$\alpha^8\rho$	$\alpha^9\rho$	$\alpha^{10}\rho$	$\alpha^{11}\rho$	$\alpha^{12}\rho$
ρ	x	y_1	y_9	y_{12}	y_8	y_4	y_7	y_3	y_5	y_2	y_{10}	y_6	y_{11}
$\alpha\rho$	y_1	y_{11}	z_6	w_5	s_3	w_8	w_4	z_9	w_{11}	w_9	s_2	w_2	z_6
$\alpha^2\rho$	y_9	z_6	y_6	w_2	z_7	w_{10}	w_{12}	s_1	s_4	w_7	w_1	z_7	w_5
$\alpha^3\rho$	y_{12}	w_5	w_2	y_{10}	s_2	w_1	z_{10}	w_3	z_8	w_6	z_{10}	w_{10}	s_3
$\alpha^4\rho$	y_8	s_3	z_7	s_2	y_2	w_9	w_7	w_6	z_4	z_4	w_3	w_{12}	w_8
$\alpha^5\rho$	y_4	w_8	w_{10}	w_1	w_9	y_5	w_{11}	s_4	z_8	z_4	z_8	s_1	w_4
$\alpha^6\rho$	y_7	w_4	w_{12}	z_{10}	w_7	w_{11}	y_3	z_9	s_1	w_3	w_6	s_4	z_9
$\alpha^7\rho$	y_3	z_9	s_1	w_3	w_6	s_4	z_9	y_7	w_4	w_{12}	z_{10}	w_7	w_{11}
$\alpha^8\rho$	y_5	w_{11}	s_4	z_8	z_4	z_8	s_1	w_4	y_4	w_8	w_{10}	w_1	w_9
$\alpha^9\rho$	y_2	w_9	w_7	w_6	z_4	z_4	w_3	w_{12}	w_8	y_8	s_3	z_7	s_2
$\alpha^{10}\rho$	y_{10}	s_2	w_1	z_{10}	w_3	z_8	w_6	z_{10}	w_{10}	s_3	y_{12}	w_5	w_2
$\alpha^{11}\rho$	y_6	w_2	z_7	w_{10}	w_{12}	s_1	s_4	w_7	w_1	z_7	w_5	y_9	z_6
$\alpha^{12}\rho$	y_{11}	z_6	w_5	s_3	w_8	w_4	z_9	w_{11}	w_9	s_2	w_2	z_6	y_1

where

$$x = \frac{107 - 8\sqrt{173}}{13}$$

is a root of the polynomial

$$P_x^{\mathbb{Z}_{13}}(x) = 13x^2 - 214x + 29\,.$$

$\mathrm{Aut}(G) \cong \mathbb{Z}_{12} \cong \langle \sigma, \tau \mid \sigma^4 = \tau^3 = 1\rangle$ permutes the roots in the following way

$$\sigma_y = (1\ 4\ 11\ 5)(2\ 7\ 8\ 3)(6\ 12\ 9\ 10)\,, \quad \sigma_z = (4\ 9)(6\ 8)(7\ 10)\,,$$
$$\sigma_w = (1\ 5\ 10\ 2)(3\ 7\ 6\ 12)(4\ 9\ 11\ 8)\,, \quad \sigma_s = (1\ 3\ 4\ 2)\,,$$

$$\tau_y = (1\ 2\ 12)(3\ 6\ 5)(4\ 7\ 9)(8\ 10\ 11)\,, \quad \tau_z = (4\ 10\ 6)(7\ 8\ 9)\,,$$
$$\tau_w = (1\ 4\ 7)(2\ 8\ 3)(5\ 9\ 6)(10\ 11\ 12)\,, \quad \tau_s = \mathrm{id}\,.$$

$G = \mathbb{Z}_{15}$

Theorem 4.5.2(i). Under the automorphism group $\mathrm{Aut}(G) \cong \mathbb{Z}_2 \times \mathbb{Z}_4$, there are three unitary orbits each with eight solutions. A solution in one orbit is given by[9]

$(F_*^{\rho,\rho,\rho})_{\rho,*}$	ρ	$\alpha\rho$	$\alpha^2\rho$	$\alpha^3\rho$	$\alpha^4\rho$	$\alpha^5\rho$	$\alpha^6\rho$	$\alpha^7\rho$	$\alpha^8\rho$	$\alpha^9\rho$	$\alpha^{10}\rho$	$\alpha^{11}\rho$	$\alpha^{12}\rho$	$\alpha^{13}\rho$	$\alpha^{14}\rho$
ρ	x_2	y_1	y_9	r_7	y_2	s_5	r_1	y_{23}	y_{16}	r_4	s_4	y_{17}	r_9	y_{19}	y_{18}
$\alpha\rho$	y_1	y_{18}	z_{14}	w_{14}	t_{10}	u_5	v_{19}	w_{13}	z_7	w_4	v_{23}	u_{10}	t_5	w_1	z_{14}
$\alpha^2\rho$	y_9	z_{14}	y_{19}	w_1	z_{15}	v_6	w_{10}	u_2	t_7	t_{12}	u_{12}	w_{20}	v_2	z_{15}	w_{14}
$\alpha^3\rho$	r_7	w_{14}	w_1	r_9	t_5	v_2	a_{11}	w_{18}	v_5	a_4	v_{10}	w_{19}	a_{11}	v_6	t_{10}
$\alpha^4\rho$	y_2	t_{10}	z_{15}	t_5	y_{17}	u_{10}	w_{20}	w_{19}	z_4	v_4	v_{13}	z_4	w_{18}	w_{10}	u_5
$\alpha^5\rho$	s_5	u_5	v_6	v_2	u_{10}	s_4	v_{23}	u_{12}	v_{10}	v_{13}	b_6	v_4	v_5	u_2	v_{19}
$\alpha^6\rho$	r_1	v_{19}	w_{10}	a_{11}	w_{20}	v_{23}	r_4	w_4	t_{12}	a_4	v_4	v_{13}	a_4	t_7	w_{13}
$\alpha^7\rho$	y_{23}	w_{13}	u_2	w_{18}	w_{19}	u_{12}	w_4	y_{16}	z_7	t_7	v_5	z_4	v_{10}	t_{12}	z_7
$\alpha^8\rho$	y_{16}	z_7	t_7	v_5	z_4	v_{10}	t_{12}	z_7	y_{23}	w_{13}	u_2	w_{18}	w_{19}	u_{12}	w_4
$\alpha^9\rho$	r_4	w_4	t_{12}	a_4	v_4	v_{13}	a_4	t_7	w_{13}	r_1	v_{19}	w_{10}	a_{11}	w_{20}	v_{23}
$\alpha^{10}\rho$	s_4	v_{23}	u_{12}	v_{10}	v_{13}	z_{22}	v_4	v_5	u_2	v_{19}	s_5	u_5	v_6	v_2	u_{10}
$\alpha^{11}\rho$	y_{17}	u_{10}	w_{20}	w_{19}	z_4	v_4	v_{13}	z_4	w_{18}	w_{10}	u_5	y_2	t_{10}	z_{15}	t_5
$\alpha^{12}\rho$	r_9	t_5	v_2	a_{11}	w_{18}	v_5	a_4	v_{10}	w_{19}	a_{11}	v_6	t_{10}	r_7	w_{14}	w_1
$\alpha^{13}\rho$	y_{19}	w_1	z_{15}	v_6	w_{10}	u_2	t_7	t_{12}	u_{12}	w_{20}	v_2	z_{15}	w_{14}	y_9	z_{14}
$\alpha^{14}\rho$	y_{18}	z_{14}	w_{14}	t_{10}	u_5	v_{19}	w_{13}	z_7	w_4	v_{23}	u_{10}	t_5	w_1	z_{14}	y_1

$\mathrm{Aut}(G) \cong \mathbb{Z}_2 \times \mathbb{Z}_4 \cong \langle \sigma, \tau \mid \sigma^2 = \tau^4 = 1 \rangle$ permutes the roots in the following way

$$\sigma_y = (1\ 18)(2\ 17)(9\ 19)(16\ 23)\,, \quad \sigma_r = (1\ 4)(7\ 9)\,, \quad \sigma_s = (4\ 5)\,, \quad \sigma_t = (5\ 10)(7\ 12)\,,$$
$$\sigma_r = (2\ 12)(5\ 10)\,, \quad \sigma_r = (1\ 4)(2\ 6)(4\ 13)(10\ 20)\,, \quad \sigma_w = (1\ 14)(4\ 13)(10\ 20)(18\ 19)\,,$$
$$\sigma_z = \mathrm{id}\,, \quad \sigma_a = \mathrm{id}\,, \quad \sigma_b = \mathrm{id}\,,$$

$$\tau_y = (1\ 19\ 2\ 23)(9\ 17\ 16\ 18)\,, \quad \tau_r = (1\ 7\ 4\ 9)\,, \quad \tau_s = \mathrm{id}\,, \quad \tau_t = (5\ 7)(10\ 12)\,,$$
$$\tau_u = (2\ 10\ 12\ 5)\,, \quad \tau_v = (2\ 13\ 5\ 23)(4\ 10\ 19\ 6)\,, \quad \tau_w = (1\ 10\ 19\ 13)(4\ 14\ 20\ 18)\,,$$
$$\tau_z = (4\ 7\ 14\ 15)\,, \quad \tau_a = (4\ 11)\,, \quad \tau_b = \mathrm{id}\,,$$

giving the other solutions in the orbit. There is an additional $\mathbb{Z}_3 \cong \langle \iota | \iota^3 = 1 \rangle$ action that acts by

$$\iota_x = (1\ 2\ 3)\,,$$
$$\iota_y = (1\ 3\ 5)(2\ 10\ 14)(4\ 20\ 18)(6\ 11\ 17)(7\ 9\ 13)(8\ 12\ 19)(15\ 16\ 22)(21\ 24\ 23)\,,$$
$$\iota_r = (1\ 6\ 8)(2\ 7\ 5)(3\ 4\ 10)(9\ 12\ 11)\,, \quad \iota_s = (1\ 5\ 3)(2\ 6\ 4)\,,$$
$$\iota_t = (1\ 6\ 7)(2\ 5\ 8)(3\ 10\ 11)(4\ 12\ 9)\,, \quad \iota_u = (1\ 12\ 4)(2\ 6\ 7)(3\ 8\ 10)(5\ 11\ 9)\,,$$
$$\iota_v = (1\ 23\ 15)(2\ 3\ 12)(4\ 18\ 16)(5\ 7\ 22)(6\ 20\ 8)(9\ 14\ 13)(10\ 17\ 21)(11\ 24\ 19)\,,$$
$$\iota_w = (1\ 8\ 16)(2\ 18\ 11)(3\ 10\ 22)(4\ 9\ 6)(5\ 24\ 20)(7\ 21\ 13)(12\ 14\ 15)(17\ 19\ 23)\,,$$
$$\iota_z = (4\ 5\ 11)(7\ 10\ 8)(14\ 19\ 17)(15\ 16\ 20)\,,$$
$$\iota_a = (3\ 10\ 11)(4\ 8\ 6)\,, \quad \iota_b = (3\ 6\ 5)\,,$$

[9]The polynomials are rather long and thus omitted in writing.

and cycles through the three distinct unitary orbits. The polynomial for x is given by

$$3375x^6 - 116550x^5 + 620280x^4 - 926392x^3 + 41520x^2 + 88128x - 6912\,.$$

4.6 Conclusions and outlook

In this paper, the notion of a transparent fusion category was defined, and the F-symbols for transparent Haagerup-Izumi fusion categories with $G = \mathbb{Z}_{2n+1}$ were constructively classified up to $G = \mathbb{Z}_9$, and further up to $G = \mathbb{Z}_{15}$ by additionally imposing S_4 tetrahedral invariance. Various graph equivalences and F-symbol relations were derived from transparency, reducing the number of independent F-symbols from $\mathcal{O}(n^6)$ to $\mathcal{O}(n^2)$,ing the pentagon identity practically solvable.

In the Cuntz algebra approach to the construction of Haagerup-Izumi fusion categories, the polynomial equations are simpler to solve for Izumi systems [105], corresponding to F-symbols with strict A_4 tetrahedral invariance, than for Grossman-Snyder systems [93], corresponding to F-symbols with S_4 tetrahedral invariance. For the former, the results of this paper (up to $G = \mathbb{Z}_9$) can be directly compared with the solutions obtained by Evans and Gannon [75],and they are in complete agreement. For the latter, whereas solutions to the Grossman-Snyder systems are available from Evans and Gannon [74] only up to $G = \mathbb{Z}_5$, the present paper constructed categories up to $G = \mathbb{Z}_{15}$. To facilitate the comparison of the Cuntz algebra approach, the present authors solved the Grossman-Snyder system up to $G = \mathbb{Z}_9$ by computing Groebner bases, and again found agreement.

Up to $G = \mathbb{Z}_9$, the number of solutions under strict A_4 tetrahedral invariance (Izumi systems) and that under S_4 (Grossman-Snyder systems) are in agreement. A possible explanation is that for any G, there is a one-to-one correspondence between the two, where a fusion category on one side is the bimodule category of another fusion category on the other side, with respect to an appropriate algebra object; in the physics language, they are related by gauging the $G = \mathbb{Z}_{2n+1}$. That this is true for the Haagerup [96] case $G = \mathbb{Z}_3$ was shown by Grossman and Snyder [93]. If this explanation is generally valid, then for $G = \mathbb{Z}_{15}$, the existence of three unitary orbits of the $\mathbb{Z}_2 \times \mathbb{Z}_4$ automorphism group according to Theorem 4.5.2 suggests that Evans and Gannon [75] missed two solutions in their numerical analysis. However, the present authors have not been able to compute a Groebner basis for the $G = \mathbb{Z}_{15}$ Izumi system to verify this speculation.

It would be interesting to construct transparent fusion categories for other fusion rings, especially quadratic (or generalized near-group) fusion rings where the fusion of the invertible objects with a single non-invertible object generates all the non-invertible objects [91, 162]. Partial transparency may already be sufficient to reduce the pentagon identity to being practically solvable. For instance, consider the following family of fusion rings: introduce ν invertible objects

$$\mathcal{I}, \quad \alpha, \quad \alpha^2, \quad \cdots \quad \alpha^{\nu-1}$$

together with $\nu+1$ non-invertible simple objects

$$\rho, \quad \alpha\rho, \quad \alpha^2\rho, \quad \cdots \quad \alpha^{\nu-1}\rho, \quad \mathcal{N}\,,$$

and define the fusion ring

$$\begin{aligned}
&\alpha^\nu = 1\,, \quad \alpha\rho = \rho\,\alpha^{\nu-1}\,, \quad \alpha\,\mathcal{N} = \mathcal{N}\,\alpha = \mathcal{N}\,,\\
&\rho^2 = \mathcal{I} + \mathcal{Z} + \mathcal{N}\,, \quad \mathcal{N}^2 = \mathcal{Y} + \mathcal{Z}\,, \quad \rho\mathcal{N} = \mathcal{N}\rho = \mathcal{Z} + \mathcal{N}\,,
\end{aligned}$$

where

$$\mathcal{Y} \equiv \sum_{k=0}^{\nu-1} \alpha^k\,, \quad \mathcal{Z} \equiv \sum_{k=0}^{\nu-1} \alpha^k \rho\,.$$

When $\nu = 1$, this is the $R_{\mathbb{C}}(\widehat{so}(3))_5$ fusion ring, which is known to admit three inequivalent fusion categories. The generalization of $R_{\mathbb{C}}(\widehat{so}(3))_5$ to the above family of fusion rings parallels the generalization of Fibonacci to Haagerup-Izumi. Because the $\mathcal{N}$ object is similar to the non-invertible object in the $G = \mathbb{Z}_\nu$ Tambara-Yamagami categories, which are not transparent, it is unreasonable to expect that the above family extending the $R_{\mathbb{C}}(\widehat{so}(3))_5$ fusion ring admits fully transparent fusion categories. Nonetheless, partial transparency for ρ, $\alpha\rho$, $\ldots$, $\alpha^{\nu-1}\rho$ may already be sufficient to render the pentagon identity solvable.

Explicit F-symbols have interesting applications. For instance, three-manifold invariants can be defined by the F-symbols alone without the need of braiding [21, 87]. In physics, one could study the gapped phase of $(1{+}1)d$ quantum field theories with Haagerup-Izumi symmetries, by constructing $(1{+}1)d$ topological quantum field theories with the same symmetries, as was done in [50] for fusion categories of smaller ranks. The statistical models of [2] and the associated anyon chains can also be explicitly constructed from the unitary F-symbols. In conformal field theory, the crossing symmetry of defect four-point functions may produce universal bounds on the spectra via the conformal bootstrap [154].

Acknowledgements

The authors are grateful to Yuji Tachikawa for initiating their interest in the Haagerup-Izumi fusion categories, and to Matthew Titsworth for sharing his explicit F-symbols for the Haagerup $\mathcal{H}_3$ fusion category. The authors also thank Chi-Ming Chang, Terry

Gannon, Tobias Osborne, Yuji Tachikawa, and Yifan Wang for helpful comments and suggestions on the draft. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632. YL is supported by the Sherman Fairchild Foundation.

Chapter 5

TOPOLOGICAL FIELD THEORY WITH HAAGERUP SYMMETRY

[1]

5.1 Introduction

The best cultivated terrains in the landscape of (1+1)d conformal field theories (CFTs) are rational conformal field theories (RCFTs) [80], free theories, and orbifolds [63, 64] thereof. Exactly marginal deformations of orbifold twist fields bring us into more interesting realms, and when explored far enough provide candidates with weakly coupled holographic duals. But the full landscape is believed to be vaster. The conformal bootstrap bounds on various quantities such as the twist gap [26, 27, 54] are not saturated by known CFTs, and numerical studies of certain renormalization group flows, such as that from the three-coupled three-state Potts model [66], indicate the existence of fixed points with irrational central charges. However, such fixed points are evasive of current analytic methods. Even for RCFTs, a full classification has not been achieved.

The full set of interesting observables in a (1+1)d CFT is not limited to the correlation functions of local operators. There are boundaries and defects that interact with the local operators in nontrivial ways, and are together subject to stringent consistency conditions. Some of the data, like the fusion category [72, 73] furnished by the topological defect lines (TDLs) [36, 50, 160], are mathematically rigid structures that exist independently of quantum field theory. A simple example of a fusion category is a group-like category, which consists of the specification of a discrete symmetry group together with its anomaly. Fusion categories generalize symmetries and anomalies, and constrain the deformation space of quantum field theory. The preceding remarks beg the following question:

Q1: *Given a fusion category, is there a (1+1)d CFT whose TDLs (or a subset thereof) realize the said category?*

The (2+1)d Turaev-Viro theory [166] or Levin-Wen string-net model [119] constructed out of a fusion category $\mathcal{C}$ is a bulk phase whose anyons are described by the Drinfeld center $\mathcal{Z}(\mathcal{C})$. By placing the bulk phase on a slab between a gapped boundary and another boundary condition B, and further compactifying on a circle, the resulting theory

[1] This chapter is adapted from Tzu-Chen Huang and Ying-Hsuan Lin. "Topological Field Theory with Haagerup Symmetry". In: (Feb. 2021). DOI: 10.1063/5.0079062. arXiv: 2102.05664 [hep-th].

$T[B]$ would be a CFT with TDLs described by $\mathcal{C}$. **Q1** is thus interpreted as the existence/classification problem of boundary conditions for the bulk phase.[2] From a purely $(1{+}1)d$ perspective, statistical height models which take $\mathcal{C}$ (and the choice of a distinguished object) as the microscopic input have recently been shown by Aasen, Fendley, and Mong [2] to host macroscopic TDLs described by $\mathcal{C}$. One can explore the phases of such models in search of a CFT.

A closely related question is the following:

Q2: *Given a modular tensor category (MTC), is there a vertex operator algebra (VOA) whose representations realize the said category?*

The phrase VOA could be replaced by diagonal RCFT, in which the fusion ring of Verlinde lines (TDLs commuting with the VOA) is isomorphic to the fusion ring of the VOA representations. The correspondence between MTC and $(1{+}1)d$ RCFT traces its origin to a seminal series of papers by Moore and Seiberg [130–134], and is conjectured to be one-to-one thought a construction or proof is lacking. Note that an affirmative answer to **Q2** implies an affirmative answer to **Q1**: Given a fusion category $\mathcal{C}$, if one can find a VOA that realizes the Drinfeld center $\mathcal{D}(\mathcal{C})$, then gauging the diagonal RCFT by an algebra object gives a non-diagonal RCFT whose TDLs realize $\mathcal{C} \boxtimes \mathcal{C}^{\mathrm{op}}$.[3]

The explicit realization of many categories in CFT is not known. A famous example is the Haagerup fusion category, which has a special place in the history of category and subfactor theory. Subfactors have inherent categorical structure, and serve as a major source of fusion categories. While Ocneanu [138] and Popa [145] classified subfactors with Jones indices less than or equal to 4, Haagerup and Asaeda [14] constructed one—the Haagerup subfactor—with Jones index $\frac{5+\sqrt{13}}{2}$, the smallest above 4 [96]. As the title of [14] suggests, the Haagerup subfactor was deemed *exotic* since its construction at the time did not fit into any infinite family. Later work by Izumi [105], Evans, and Gannon [75] postulated that the Haagerup subfactor does in fact fit into an infinite family, and furthermore constructed the first few members. This development suggested that the Haagerup may not be exotic after all. Nonetheless, for various categorical conjectures, the explicit demonstration in the case of Haagerup is viewed as a key test of a conjecture's legitimacy and generality.

There are actually three inequivalent unitary Haagerup fusion categories, commonly denoted by $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$. Most of this note concerns the Haagerup $\mathcal{H}_3$ fusion category,

[2] The authors thank Shu-Heng Shao and Yifan Wang for discussions.
[3] The authors thank Sahand Seifnashri for a discussion.

which did not descend directly from the Haagerup subfactor of [14, 96], but was instead constructed by Grossman and Snyder [93]. Because the fusion ring (reviewed in Section 5.3) is non-commutative, the Haagerup $\mathcal{H}_3$ fusion category cannot possibly be realized by Verlinde lines [67, 85, 142, 167] in a diagonal RCFT. To our knowledge, its realization as general TDLs (need not commute with the full VOA) is not known in any CFT. To connect to Verlinde lines, one must consider the MTC that is Drinfeld center of Haagerup. In fact, Evans and Gannon [75] constructed $c = 8$ characters for the Haagerup modular data, and used it to surmise possible constructions of the VOA through the Goddard-Kent-Olive coset construction [88] and its generalizations [65, 84, 120], or through the generalized orbifold construction (gauging an algebra object) of Carqueville, Fröhlich, Fuchs, Runkel, and Schweigert [47, 81, 83] (see [36] for a recent discussion). Recently, Wolf [176] considered the Haagerup anyon chain and numerically searched for CFT phases, but with inconclusive results.[4] To date, a *bona fide* construction remains an important open question. By trying to construct CFTs that realize more exotic fusion categories, one hope is that light would be shed beyond the current borders of known (R)CFTs.

Concerning the gapped phases of $(1+1)d$ quantum field theory, described by $(1+1)d$ topological field theories (TFTs) extended by defects [58, 148], a related but simpler question can be asked:[5]

Q3: *Given a fusion category, is there a (1+1)d TFT whose TDLs (or a subset thereof) realize the said category?*

Thorngren and Wang [161] has argued that $\mathcal{C}$-symmetric defect TFTs are in bijection with $\mathcal{C}$-module categories, and since the regular module category always exists, **Q3** has been affirmatively answered. However, their construction utilizes the Turaev-Viro state-sum [166] or Levin-Wen string-net [119] construction, and it is generally unclear how the axiomatic TFT data can be extracted. We are thus led to the next question:

Q4: *Given a fusion category, can one construct the axiomatic data of a (1+1)d TFT whose TDLs (or a subset thereof) realize the said category?*

[4] Anyon chains generalize the golden chain of Feiguin et al [77], and arise in a limit of the statistical height models of Aasen, Fendley, and Mong [2].

[5] There are various notions of TFT with different amounts of structure, the most common being closed TFT [3, 61, 114, 150] and open/closed TFT [6, 117, 118, 129]. The defect TFT of [58, 148] is an overarching formalism that can incorporate multiple closed TFTs and their boundaries and interfaces. The minimal structure that incorporates the data of TDLs is a defect TFT containing a single closed TFT; mathematically speaking, it is a bicategory with a single object, whose 1-morphisms are the TDLs, and whose 2-morphisms are the local and defect operators. The full enrichment by boundaries and interfaces with other closed TFTs is beyond the scope of this note.

This question has been answered for group-like categories by Wang, Wen, and Witten [168] and by Tachikawa [160], and for categories with fiber functors (the resulting TFT has a unique vacuum) by Thorngren and Wang [161]. In [50], it was shown that for a variety of CFTs, the TFT data can be solved solely from the input of the fusion categorical data, by bootstrapping the consistency conditions. For general categories, a construction of the bulk Frobenius algebra was given by Komargodski, Ohmori, Roumpedakis, and Seifnashri [115], but the full defect TFT data remains unsolved.[6]

The preceding questions are ultimately connected. A CFT realizing a certain fusion category is connected to a TFT realizing the same category under TDL-preserving renormalization group (RG) flows, and this principle strongly constrains the infrared fate of CFTs. In the space of TDL-preserving RG flows and without fine-tuning, they must either flow to gapped phases described by TFTs, or to "dead-end" CFTs [136], which correspond to gapless phases protected by fusion categories [161].[7]

As desirable as fully universal answers to the preceding questions are, a more pragmatic approach may be to first examine fusion categories for which the answers are not known. This note makes a modest offering in this approach of pursuing exotica in the quest for their eventual conformity: the construction of a TFT realizing the Haagerup $\mathcal{H}_3$ fusion category with fully explicit axiomatic TFT data. The construction is of bootstrap nature, by solving the full cutting and sewing consistency conditions. A prerequisite in this approach is the explicit knowledge of the F-symbols. They were implicit in the work of Grossman and Snyder [93] (using a generalization of the approach by Izumi [105]), and also explicitly obtained by Titsworth [163], Osborne, Stiegemann, and Wolf [139]. In [102], the present authors recast the F-symbols in a gauge that manifests the transparent property, which greatly simplify our present computational endeavor.

The remaining sections are organized into steps of the construction and discussions of further ramifications. Section 5.2 reviews the generalities of topological field theory extended by defects, and formulates the defining data and consistency conditions. Section 5.3 introduces the Haagerup fusion ring with six simple objects/TDLs, studies its representation theory, and constrains the vacuum degeneracy using modular invariance. Section 5.4 studies the relations among dynamical data implied by transparency and $\mathbb{Z}_3$ symmetry. Section 5.5 delineates the constraints of associativity and torus one-point

[6] After publication of the first version of this note, Kantaro Ohmori and Sahand Seifnashri suggested to the authors that a construction of the full defect TFT data may be possible through a generalization of [115].

[7] Such phases generalize the notion of (group-like) symmetry-protected gapless phases [51] and perfect metals [144].

modular invariance. Section 5.6 solves the constraints to construct a topological field theory with Haagerup symmetry. Section 5.7 examines the expectation that the boundary conditions furnish a non-negative integer matrix representation (NIM-rep) of the fusion ring. Section 5.8 discusses the relations among topological field theories by gauging algebra objects. Section 5.9 ends with some prospective questions. Appendix D.1 contains the F-symbols for the Haagerup $\mathcal{H}_3$ fusion category. Appendix D.2 analyzes the general crossing symmetry of defect operators.

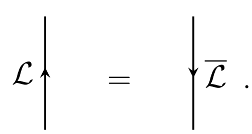
5.2 Topological field theory extended by defects

This section introduces the axiomatic data of a topological field theory (TFT) extended by defects, and the consistency conditions they must satisfy.

Fusion category of topological defect lines

The nontrivial splitting and joining relations of a finite set of topological defect lines (TDLs) are captured by a fusion category. A classic introduction to fusion categories can be found in [72, 73], and expositions in the physics context can be found in [36, 50]. Here we follow the latter and present a lightening review of the key properties of TDLs.

Topological defect lines are (generally oriented) defect lines whose isotopic transformations leave physical observables invariant. We restrict ourselves to considering sets of TDLs with finitely many *simple* TDLs $\{\mathcal{L}_i\}$; the others, the *non-simple* TDLs, are direct sums of the simple ones.[8] Among the simple TDLs there is a trivial TDL $\mathcal{I}$ representing nothingness. Furthermore, every TDL $\mathcal{L}$ has an orientation reversal $\overline{\mathcal{L}}$, as depicted by the equivalence



(5.1)

Whenever a TDL is isomorphic to its own orientation reversal, $\mathcal{L} = \overline{\mathcal{L}}$, we omit the arrows on the lines.[9]

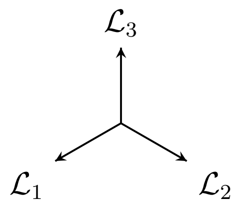
A general configuration of TDLs involves junctions built out of trivalent vertices. The allowed trivalent vertices are specified by the fusion ring

$$\mathcal{L}_i \mathcal{L}_j = N_{ij}^k \mathcal{L}_k \,, \tag{5.2}$$

[8] See [49] for progress in incorporating "non-compact" topological defect lines.

[9] The orientation cannot be completely ignored if the TDL has an orientation-reversal anomaly (nontrivial Frobenius-Schur indicator) [50]. This subtlety does not arise for the Haagerup and is therefore neglected.

where $N_{ij}^k \in \mathbb{Z}_{\geq 0}$ are the fusion coefficients. To simplify the discussion, it is assumed that (1) the fusion coefficients (dimensions of junction vector spaces) are zero or one, and (2) the trivalent vertices are cyclic-permutation invariant.[10] In conformity with [50, 102], we adopt the counter-clockwise convention for trivalent vertices, such that



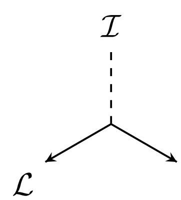
(5.3)

is allowed when $\mathcal{I} \in \mathcal{L}_1\mathcal{L}_2\mathcal{L}_3$. To completely specify a trivalent vertex, a junction vector must be chosen from the junction vector space $V_{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}$.[11] The collection of choices for all trivalent vertices formed by all simple TDLs constitutes a gauge.

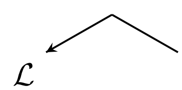
The fusion product of a simple TDL $\mathcal{L}$ with its orientation reversal contains the trivial TDL,

$$\mathcal{L}\,\overline{\mathcal{L}} = \mathcal{I} + \cdots \,, \tag{5.4}$$

because clearly



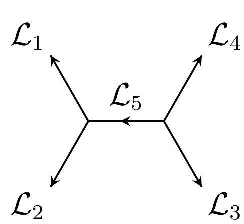
$=$



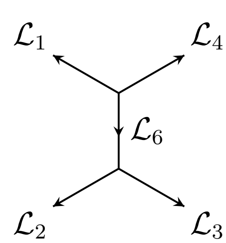
(5.5)

is allowed. Another important notion is *invertibility*. A TDL $\mathcal{L}$ is invertible if $\mathcal{L}\,\overline{\mathcal{L}} = \mathcal{I}$, and non-invertible otherwise. Invertible TDLs are equivalent to background gauge bundles for finite symmetry groups [36, 86].

The splitting and joining of TDLs can be decomposed into basic F-moves that are characterized by the F-symbols. In a given gauge, the F-symbols are $\mathbb{C}^\times$-numbers, and an F-move is the equivalence between the two configurations



$$= \sum_{\mathcal{L}_6} (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}_6}$$



. (5.6)

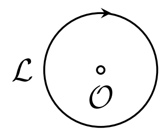
[10] Both assumptions are satisfied by the transparent Haagerup $\mathcal{H}_3$ fusion category. The reader is referred to [50] for a general discussion without these assumptions.

[11] In the path integral language, a junction vector specifies the boundary conditions of quantum fields at a trivalent vertex.

The F-symbols must satisfy the pentagon identity, which can only have finitely many solutions (up to gauge equivalence) for a given fusion ring due to Ocneanu rigidity [72, 169].

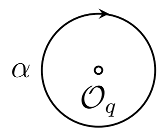
Local operators and commutative Frobenius algebra

Topological defect lines act on local operators by circling and shrinking. In conformity with [50, 102], we adopt the clockwise convention for action on local operators,



$$= \widehat{\mathcal{L}}(\mathcal{O})\,. \tag{5.7}$$

For instance, if $\mathcal{O}_q$ is a local operator with $\mathbb{Z}_3$-charge q, and if α is the TDL corresponding to the generator of $\mathbb{Z}_3$, then



$$= \omega^q\, \mathcal{O}_q\,. \tag{5.8}$$

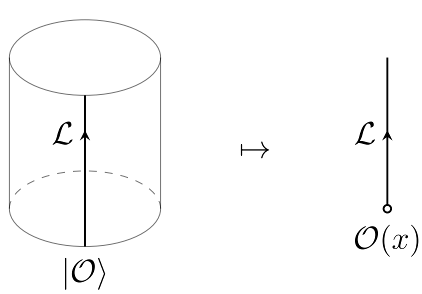
The data of local operators is captured by a commutative Frobenius algebra [3, 114]. Commutativity guarantees that a projector basis exists:

$$\{\pi_a,\, a = 1, \ldots, n_{\mathrm{V}} \mid \pi_a \pi_b = \delta_{ab} \pi_a\}\,, \tag{5.9}$$

where n_{V} denotes the number of vacua. In this basis, the nontrivial data is captured in the overlap of the projectors with the identity, *i.e.* the one-point functions $\langle \pi_a \rangle$. Most of this note does not work in the projector basis, because for us it is more convenient to work in a basis that simplifies the TDL actions as much as possible. However, the projector basis will figure in the discussion of boundary states in Section 5.7.

Defect operators, defect operator algebra, and lassos

Associated to every topological defect line $\mathcal{L}$ is a defect Hilbert space $\mathcal{H}_{\mathcal{L}}$, which contains states quantized on a spatial circle with twisted periodic boundary conditions. Via the state-operator map,



(5.10)

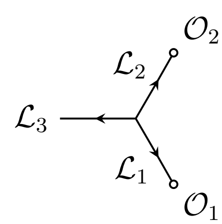
$\mathcal{H}_{\mathcal{L}}$ is also the Hilbert space of point-like *defect operators* on which $\mathcal{L}$ can end. Defect Hilbert spaces are equipped with a norm

$$\mathcal{H}_{\mathcal{L}} \otimes \mathcal{H}_{\overline{\mathcal{L}}} \to \mathbb{C}\,, \tag{5.11}$$

which defines a hermitian structure. The hermitian conjugate of $\mathcal{O}$ will be denoted by $\overline{\mathcal{O}}$.

The spectral data of a topological field theory extended by defects consists of the set of local operators, their representations under the fusion ring, and the set of defect operators. The dynamical data consists of the operator product

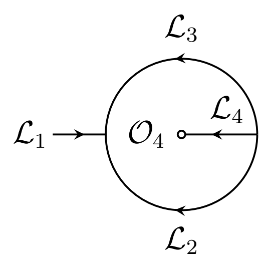
$$\mathcal{O}_1 \otimes \mathcal{O}_2 \in \mathcal{H}_{\overline{\mathcal{L}}_1} \otimes \mathcal{H}_{\overline{\mathcal{L}}_2} \quad \mapsto$$



$$\in \mathcal{H}_{\mathcal{L}_3} \tag{5.12}$$

and the lasso action

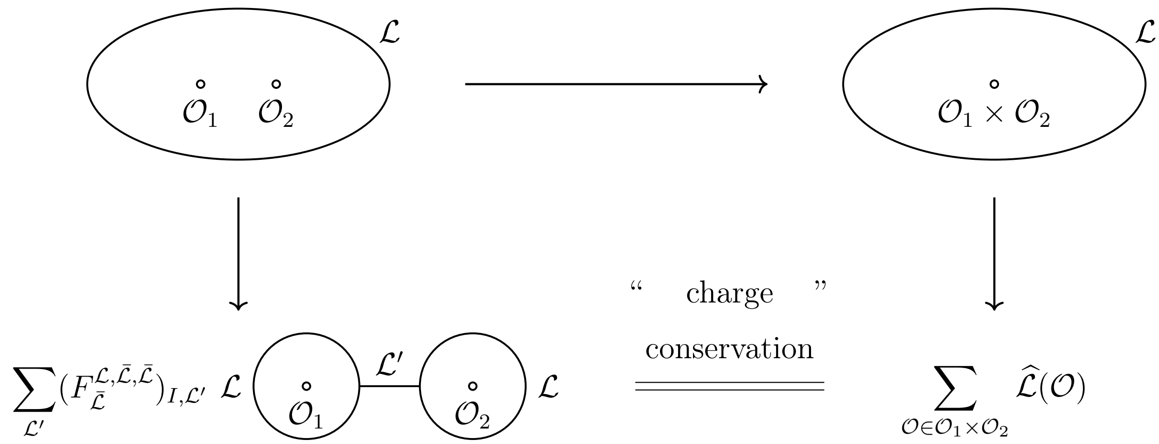
$$\mathcal{O}_4 \in \mathcal{H}_{\mathcal{L}_4} \quad \mapsto$$



$$\in \mathcal{H}_{\mathcal{L}_1}\,. \tag{5.13}$$

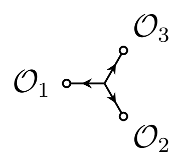
When $\mathcal{L}_1 = \mathcal{L}_4 = \mathcal{I}$ and $\mathcal{L}_2 = \overline{\mathcal{L}}_3$, the above diagram becomes (5.7), and the lasso action reduces to the TDL action $\widehat{\mathcal{L}}_2$ on local operators that maps $\mathcal{H}$ to $\mathcal{H}$. The lasso action is a generalization that maps a defect Hilbert space $\mathcal{H}_4$ to another defect Hilbert space $\mathcal{H}_1$. In the following, for TDLs ending on defect operators, the labeling of the former will be suppressed as it is implied by that of the latter.

The closest analog of charge conservation for a non-invertible TDL $\mathcal{L}$ is to circle a pair of local operators by $\mathcal{L}$, and impose the commutativity of (1) taking the local operator product and (2) performing an F-move and studying the defect operator product, as illustrated below:

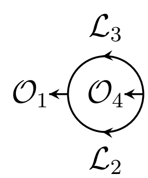


(5.14)

By the use of the norm, the operator product is equivalently encoded in the three-point coefficients

$c(\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3) \quad =$  $\in \mathbb{C}\,,$ (5.15)

and the lasso action is encoded in the lasso coefficients

 $\in \mathbb{C}\,.$ (5.16)

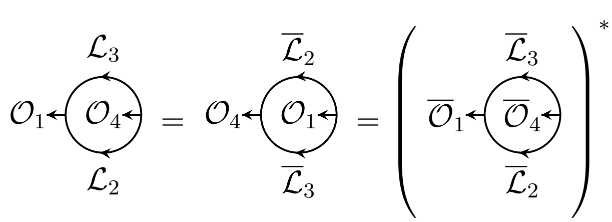
In the above, vacuum expectation values are implicitly taken. The three-point coefficients are invariant under cyclic permutations

$$c(\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3) = c(\mathcal{O}_2, \mathcal{O}_3, \mathcal{O}_1) = c(\mathcal{O}_3, \mathcal{O}_1, \mathcal{O}_2)\,, \tag{5.17}$$

and complex conjugate under reflections

$$c(\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3) = c(\overline{\mathcal{O}}_1, \overline{\mathcal{O}}_3, \overline{\mathcal{O}}_2)^*\,. \tag{5.18}$$

The lasso coefficients also enjoy the symmetries

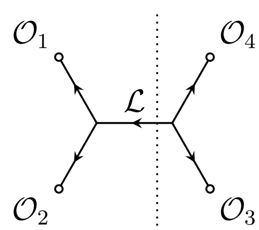
 . (5.19)

General observables, crossing symmetry, and modular invariance

A general observable in a topological field theory extended by defects is the vacuum expectation value of a graph—a configuration of topological defect lines with junctions and endpoints—on a Riemann surface.[12] On the sphere, any graph can be expanded into a sum of local operators, and taking the vacuum expectation value amounts to computing the overlap with the identity. The basic building blocks for this computation are the three-point and lasso coefficients introduced earlier, and the computation also involves basic manipulations of TDLs such as F-moves. Observables on general Riemann surfaces can be reduced to those on the sphere by a pair-of-pants decomposition. The equivalence of the various ways of building the same observable on a general Riemann surface is guaranteed by the four-point crossing symmetry and torus one-point modular invariance [50], generalizing the situation without defects argued by Sonoda [156, 157] and by Moore and Seiberg [130, 131]. In the following, all local and defect operators are taken to be canonically normalized under the hermitian structure,

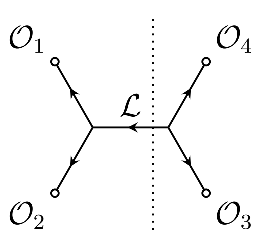
$$\langle\, \mathcal{O} \overset{\mathcal{L}}{\longleftarrow} \overline{\mathcal{O}} \,\rangle = 1\,. \tag{5.20}$$

On the sphere, the four-point correlator of local and defect operators $\mathcal{O}_i \in \mathcal{H}_{\mathcal{L}_i}$ bridged by an internal $\mathcal{L} \in \mathcal{L}_1\mathcal{L}_2 \cap \overline{\mathcal{L}}_4\overline{\mathcal{L}}_3$ can be decomposed into three-point coefficients by cutting across $\mathcal{L}$ (with the cut shown by the dotted lines),

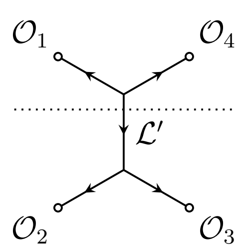


$$= \sum_{\mathcal{O}\in\mathcal{H}_{\mathcal{L}}} c(\mathcal{O}_1,\mathcal{O}_2,\mathcal{O})\, c(\mathcal{O}_3,\mathcal{O}_4,\overline{\mathcal{O}})\,. \tag{5.21}$$

Under an F-move,



$$= \sum_{\mathcal{L}'}$$



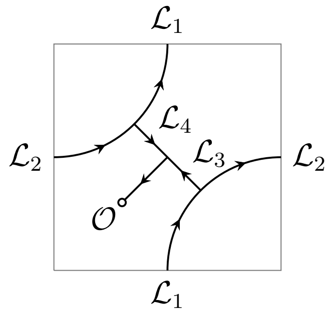
$$(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L},\mathcal{L}'}\,, \tag{5.22}$$

[12]Each observable can be interpreted as a transition amplitude over some time function, with non-trivial topology changes and defect dressing. See [36] for an exposition from this perspective.

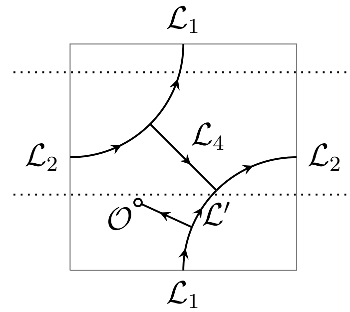
where each graph appearing on the right can be decomposed into three-point coefficients by cutting across $\mathcal{L}'$. Crossing symmetry is the equivalence of

$$\sum_{\mathcal{O}\in\mathcal{H}_{\mathcal{L}}} c(\mathcal{O}_1,\mathcal{O}_2,\mathcal{O})\, c(\mathcal{O}_3,\mathcal{O}_4,\overline{\mathcal{O}}) = \sum_{\mathcal{L}'} (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L},\mathcal{L}'} \sum_{\mathcal{O}'\in\mathcal{H}_{\mathcal{L}'}} c(\mathcal{O}_2,\mathcal{O}_3,\mathcal{O}')\, c(\mathcal{O}_4,\mathcal{O}_1,\overline{\mathcal{O}}')\,. \tag{5.23}$$

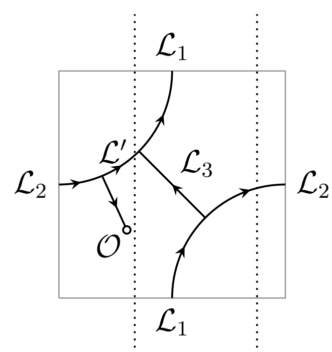
The modular invariance of the torus one-point function begins with performing F-moves on the configuration



$$= \sum_{\mathcal{L}'\in\mathcal{L}_{\mathcal{O}}\overline{\mathcal{L}}_1} (F^{\overline{\mathcal{L}}_4,\mathcal{L}_{\mathcal{O}},\overline{\mathcal{L}}_1}_{\overline{\mathcal{L}}_2})_{\mathcal{L}_3,\mathcal{L}'}$$



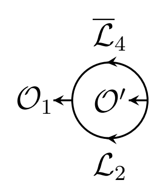
$$= \sum_{\mathcal{L}'\in\overline{\mathcal{L}}_2\mathcal{L}_{\mathcal{O}}} (F^{\mathcal{L}_1,\overline{\mathcal{L}}_2,\mathcal{L}_{\mathcal{O}}}_{\mathcal{L}_3})_{\overline{\mathcal{L}}_4,\mathcal{L}'}$$



(5.24)

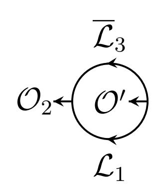
and demanding the equivalence of the two cuts shown by the dotted lines:

$$\sum_{\mathcal{L}'\in\mathcal{L}_{\mathcal{O}}\overline{\mathcal{L}}_1} \sum_{\mathcal{O}_1\in\mathcal{H}_{\mathcal{L}_1}} \sum_{\mathcal{O}'\in\mathcal{H}_{\mathcal{L}'}} (F^{\overline{\mathcal{L}}_4,\mathcal{L}_{\mathcal{O}},\overline{\mathcal{L}}_1}_{\overline{\mathcal{L}}_2})_{\mathcal{L}_3,\mathcal{L}'}\, c(\mathcal{O},\overline{\mathcal{O}}_1,\overline{\mathcal{O}}')$$



(5.25)

$$= \sum_{\mathcal{L}'\in\overline{\mathcal{L}}_2\mathcal{L}_{\mathcal{O}}} \sum_{\mathcal{O}_2\in\mathcal{H}_{\mathcal{L}_2}} \sum_{\mathcal{O}'\in\mathcal{H}_{\mathcal{L}'}} (F^{\mathcal{L}_1,\overline{\mathcal{L}}_2,\mathcal{L}_{\mathcal{O}}}_{\mathcal{L}_3})_{\overline{\mathcal{L}}_4,\mathcal{L}'}\, c(\mathcal{O},\overline{\mathcal{O}}_2,\overline{\mathcal{O}}')$$



.

5.3 Spectral constraints by Haagerup symmetry

This section studies the modular constraints on the spectral data—the set of local operators, their representations under the fusion ring, and the set of defect operators—when the theory is known to contain topological defect lines (TDLs) realizing the Haagerup $\mathcal{H}_3$ fusion category.

The Haagerup fusion ring with six simple objects

The Haagerup $\mathcal{H}_3$ fusion category was constructed by Grossman and Snyder [93] as a variant (Grothendieck equivalent) of the $\mathcal{H}_2$ fusion category that directly came from the Haagerup subfactor [14, 96]. There are six simple objects/TDLs, which we denote by

$$\mathcal{I}, \quad \alpha, \quad \alpha^2, \quad \rho, \quad \alpha\rho, \quad \alpha^2\rho\,. \tag{5.26}$$

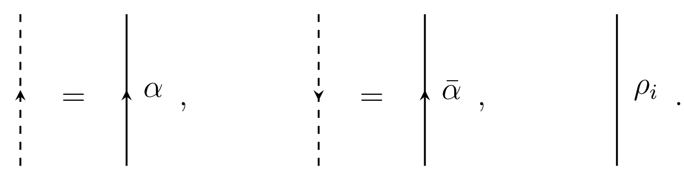
The fusion ring is fully specified by the relations

$$\alpha^3 = 1\,, \quad \alpha\rho = \rho\,\alpha^2\,, \quad \rho^2 = \mathcal{I} + \mathcal{Z}\,, \quad \mathcal{Z} \equiv \sum_{i=0}^{2} \alpha^i \rho\,. \tag{5.27}$$

For shorthand,

$$\rho_i \equiv \alpha^i \rho\,. \tag{5.28}$$

In the rest of this note, we use unoriented solid lines to denote the non-invertible self-dual simple TDLs ρ_i, and oriented dashed lines to denote the invertible ones:



(5.29)

There are two gauge-inequivalent unitary fusion categories realizing the above fusion ring, denoted $\mathcal{H}_2$ and $\mathcal{H}_3$ by Grossman and Snyder [93]. Whereas the Haagerup $\mathcal{H}_2$ fusion category descended directly from the Haagerup subfactor [14, 96], the Haagerup $\mathcal{H}_3$ fusion category was constructed by Grossman and Snyder [93] based on $\mathcal{H}_2$. It turns out to be easier to work with $\mathcal{H}_3$, but the analysis in this section applies to both $\mathcal{H}_2$ and $\mathcal{H}_3$. The F-symbols for $\mathcal{H}_3$ were implicit in the work of Grossman and Snyder [93] (using a generalization of the approach by Izumi [105] for $\mathcal{H}_2$), and also explicitly obtained by Titsworth [163], Osborne, Stiegemann, and Wolf [139]. In [102], the present authors recast the F-symbols in a gauge that manifests the transparent property, a notion we introduce in Section 5.4. The transparent F-symbols are given in Appendix D.1.

Action on local operators and representation theory

To describe how topological defect lines forming the Haagerup $\mathcal{H}_3$ fusion category act on local operators, we should first study the complex representation theory of its fusion ring. Since the fusion ring is non-commutative, the action of TDLs cannot be simultaneously diagonalized. We work in a basis in which the action of $\mathbb{Z}_3$ is diagonal.

- For a state $|\phi\rangle$ neutral under $\mathbb{Z}_3$,

$$\rho|\phi\rangle = \alpha\rho|\phi\rangle = \alpha^2\rho|\phi\rangle\,, \quad \mathcal{Z}|\phi\rangle = 3\rho|\phi\rangle\,, \tag{5.30}$$

  hence there are two one-dimensional representations,

$$\rho|\phi\rangle = \frac{3 \pm \sqrt{13}}{2}|\phi\rangle\,. \tag{5.31}$$

- For a state $|\phi\rangle$ with unit $\mathbb{Z}_3$-charge,

$$\alpha|\phi\rangle = \omega|\phi\rangle\,, \quad \alpha\rho|\phi\rangle = \rho\alpha^2|\phi\rangle = \omega^2\rho|\phi\rangle\,, \quad \alpha^2\rho|\phi\rangle = \rho\alpha|\phi\rangle = \omega\rho|\phi\rangle\,. \tag{5.32}$$

  It follows that $\mathcal{Z}|\phi\rangle = 0$, and hence

$$\rho^2|\phi\rangle = |\phi\rangle\,. \tag{5.33}$$

  If $\rho|\phi\rangle$ and $|\phi\rangle$ were equal up to a phase, then there would be two possible one-dimensional representations with

$$\rho|\phi\rangle = \pm|\phi\rangle\,, \tag{5.34}$$

  which is in conflict with $\alpha\rho = \rho\alpha^2$. Hence $\rho|\phi\rangle$ and $|\phi\rangle$ must be independent, and the representation is two-dimensional. In the $(|\phi\rangle, \rho|\phi\rangle)$ basis,

$$\alpha = \begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}\,, \quad \rho = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,. \tag{5.35}$$

The above classification of irreducible representations is summarized in Table 5.1. In a reflection-positive quantum field theory, the identity operator transforms in a one-dimensional representation with positive charges. Here, under the reflection-positive assumption, the identity operator must transform in the + representation.

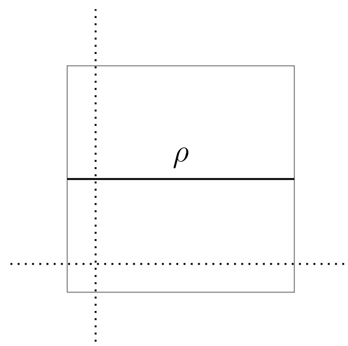
r	α	ρ
$+$	1	$\frac{3+\sqrt{13}}{2}$
$-$	1	$\frac{3-\sqrt{13}}{2}$
2	$\begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Table 5.1: Irreducible representations of the Haagerup fusion ring with six simple objects/TDLs.

Modular invariance and vacuum degeneracy

Let $n_{\rm V}$ denote the number of vacua (local operators), and $n_\pm$ and $n_{\boldsymbol{2}}$ be their multiplicities of representations (in the notation of Table 5.1). Clearly, $n_{\rm V} = n_+ + n_- + 2n_{\boldsymbol{2}}$.

Consider the modular invariance of the torus partition function with the non-invertible TDL ρ wrapped around a one-cycle



(5.36)

The horizontal cut computes the trace over the action of $\widehat{\rho}$ in the Hilbert space $\mathcal{H}$ of local operators, and the vertical cut simply counts the dimensionality of the defect Hilbert space $\mathcal{H}_\rho$. Modular invariance requires

$$\mathrm{Tr}_{\mathcal{H}}\, \widehat{\rho} = \mathrm{Tr}_{\mathcal{H}_\rho} 1 \in \mathbb{Z}_{\geq 0}\,. \tag{5.37}$$

Given the representation content of the Haagerup fusion ring, summarized in Table 5.1, we immediately conclude that $n_+ = n_-$, and the number of vacua must be even. Let us write

$$n_{\boldsymbol{1}} \equiv n_+ = n_- \tag{5.38}$$

to denote the multiplicity of each one-dimensional representation. Using the $\mathrm{U}(n_{\boldsymbol{2}})$ freedom, we can choose a basis of local operators to represent $\widehat{\rho}$ in block diagonal form

$$\widehat{\rho} = \bigoplus_{p=1}^{n_+} \left(\frac{3+\sqrt{13}}{2}\right) \oplus \bigoplus_{q=1}^{n_-} \left(\frac{3-\sqrt{13}}{2}\right) \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,. \tag{5.39}$$

Modular invariance (5.37) also implies that the defect Hilbert space $\mathcal{H}_\rho$ is $3n_{\boldsymbol{1}}$-dimensional, *i.e.* the TDL ρ can end on

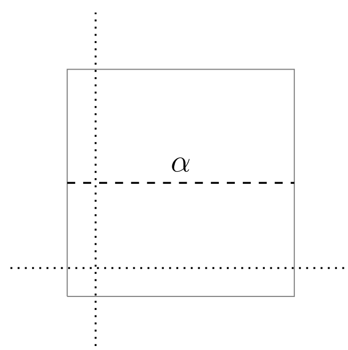
$$n_\rho = 3n_{\boldsymbol{1}} \tag{5.40}$$

independent defect operators, and similarly for each of the other ρ_i.

n_{V}	$n_{\mathbf{1}} = n_+ = n_-$	$n_{\mathbf{2}}$	$n_\alpha = n_{\bar{\alpha}}$	$n_\rho = n_{\alpha\rho} = n_{\alpha^2\rho}$	n_{P}
2	1	0	2	3	15
4	1	1	1	3	15
4	2	0	4	6	30
6	1	2	0	3	15
6	2	1	3	6	30
6	3	0	6	9	45

Table 5.2: Possible numbers of point-like operators that satisfy the torus one-point modular invariance (5.36) and (5.41). Here n_{V} denotes the total number of vacua (local operators), comprised of $n_{\mathbf{r}}$ copies of representation $\mathbf{r}$, where $\mathbf{r} = +, -, \mathbf{2}$; $n_{\mathcal{L}}$ denotes the number of defect operators in *each* $\mathcal{L}$, for $\mathcal{L} = \alpha, \bar{\alpha}, \rho, \alpha\rho, \alpha^2\rho$; n_{P} denotes the total number of point-like (local and defect) operators. Only the highlighted cases with $n_{\mathbf{1}} = 1$, $n_\rho = 3$, and $n_{\mathrm{P}} = 15$ are considered in this note.

Consider the modular invariance of the torus partition function with the invertible TDL α wrapped around a one-cycle



(5.41)

Modular invariance requires

$$\mathrm{Tr}_{\mathcal{H}}\, \widehat{\alpha} = \mathrm{Tr}_{\mathcal{H}_\alpha} 1 \in \mathbb{Z}_{\geq 0}\,. \tag{5.42}$$

Hence the α TDL hosts

$$n_\alpha = 2n_{\mathbf{1}} - n_{\mathbf{2}} \tag{5.43}$$

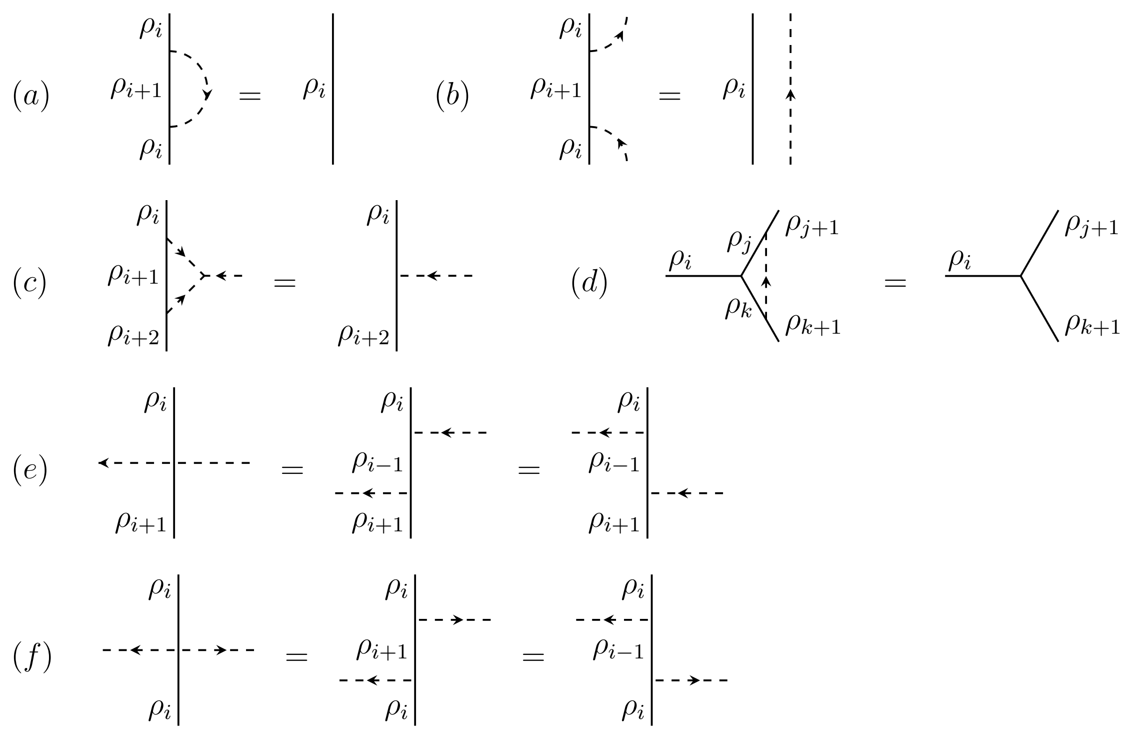
defect operators. The total number of point-like operators is

$$n_{\mathrm{P}} \equiv g + 2n_\alpha + 3n_\rho = (2n_{\mathbf{1}} + 2n_{\mathbf{2}}) + 2(2n_{\mathbf{1}} - n_{\mathbf{2}}) + 9n_{\mathbf{1}} = 15n_{\mathbf{1}}\,. \tag{5.44}$$

The first few possibilities are listed in Table 5.2 in the order of increasing n_{V}. Whenever $n_{\mathbf{2}} = 0$, the $\mathbb{Z}_3$ symmetry is not faithfully realized on the vacua. In the following, we consider the three minimal cases totaling $n_{\mathrm{P}} = 15$ point-like operators, highlighted in Table 5.2; each case has $n_{\mathbf{1}} = 1$ and $n_\rho = 3$. Eventually we will succeed in constructing a TFT realizing $n_{\mathrm{V}} = 6$, but along the way we also derive various constraints on $n_{\mathrm{V}} = 2, 4$.

5.4 Transparency and $\mathbb{Z}_3$ symmetry

This note works in a gauge of the $\mathcal{H}_3$ fusion category that manifests its "transparent" property [102]—the associator involving any invertible topological defect line (TDL) is the identity morphism. In terms of the F-symbols, it means that every F-symbol with an external invertible TDL takes value one. Hence invertible TDLs can be attached or detached "freely", changing the isomorphism classes of other involved TDLs but without generating extra F-symbols. Several diagrammatic identities are illustrated below:



(5.45)

Importantly, the four-way junctions in (e) and (f) are unambiguously defined.

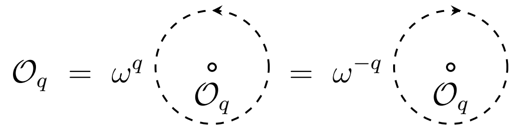
In [102], transparency and the $\mathbb{Z}_3$ symmetry were exploited to reduce the pentagon identity so that the F-symbols could be efficiently solved. Below, in attempting to construct a topological field theory, the utilization of the $\mathbb{Z}_3$ symmetry is also essential in reducing the amount of independent data.

$\mathbb{Z}_3$ relations for lassos and dumbbells

Let $\mathcal{O}_q$ be a local operator with $\mathbb{Z}_3$-charge $q \in \{0, \pm 1\}$, and consider the lasso

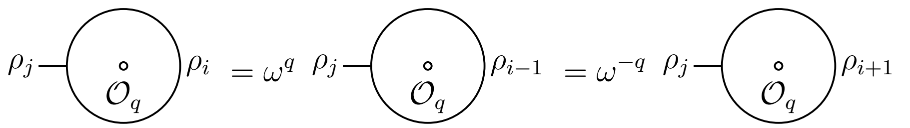
$$\rho_j \mathcal{O}_q \rho_i \ . \tag{5.46}$$

The $\mathbb{Z}_3$ symmetry relates lassos with different triples (q, i, j) as follows: replace $\mathcal{O}_q$ using the equalities



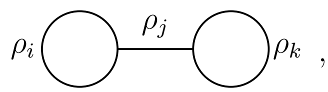
(5.47)

and fuse the $\mathbb{Z}_3$ symmetry line with ρ_i (apply (5.45)(b) and then (d)) to obtain the relations



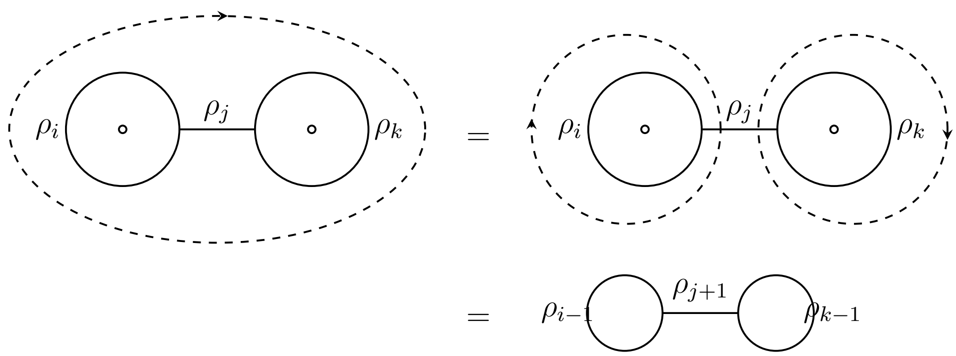
(5.48)

Next consider the dumbbell



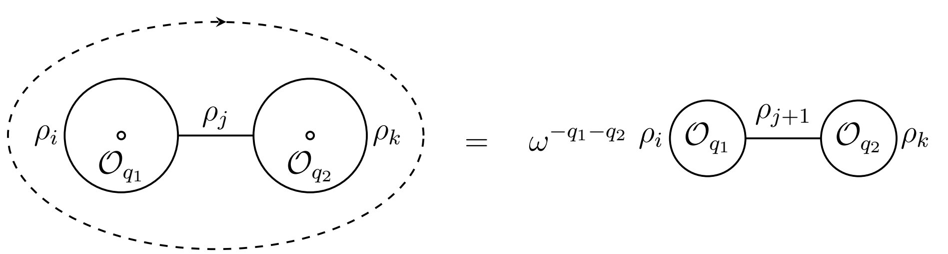
(5.49)

where each empty dot denotes an arbitrary local operator insertion. The $\mathbb{Z}_3$ action on the dumbbell (circling it with a clockwise $\mathbb{Z}_3$ loop) gives (see (5.45)(e) for the meaning of the four-way junction)



(5.50)

Combining (5.48) and (5.50), we obtain an identity that leaves the side loops intact and only changes the handle



(5.51)

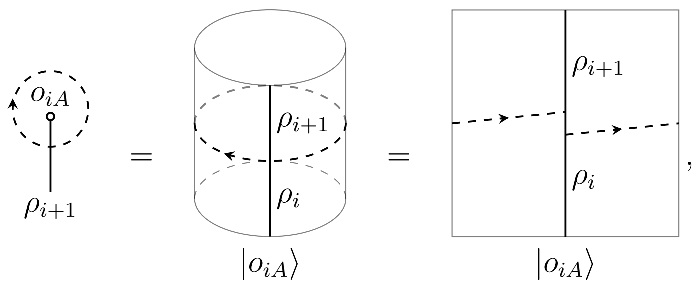
which will prove useful in Section 5.5. A mnemonic is that the $\mathbb{Z}_3$ symmetry line measures the *opposite* $\mathbb{Z}_3$-charge of the local operators $\mathcal{O}_{q_1}$ and $\mathcal{O}_{q_2}$ placed inside a dumbbell, because the $\mathbb{Z}_3$ symmetry line changes orientation when it crosses a ρ_i TDL, as illustrated in (5.45)(f).

$\mathbb{Z}_3$ action on defect operators

Recall that each ρ_i TDL hosts three independent defect operators. We work in an orthonormal basis and denote them by

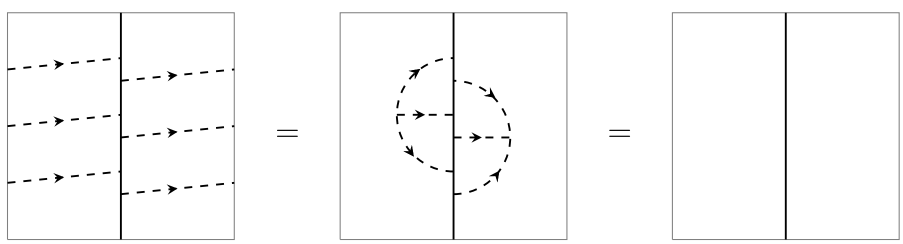
$$o_{iA}\,, \quad i = 0, 1, 2\,, \quad A = 1, 2, 3\,, \qquad \text{with} \quad \langle o_{iA} o_{jB} \rangle = \delta_{ij}\delta_{AB}\,. \tag{5.52}$$

Note that there is still an $\mathrm{O}(3)^3$ basis freedom. The $\mathbb{Z}_3$ action on a defect operator $o_i \in \mathcal{H}_{\rho_i}$ is defined by the lasso (see (5.45)(e) for the meaning of the four-way junction)



, (5.53)

where in the last diagram the left and right edges of the square are identified to represent a cylinder. Performing the $\mathbb{Z}_3$ action three times on $\mathcal{H}_{\rho_i}$ becomes a trivial action, as illustrated by the sequence of F-moves

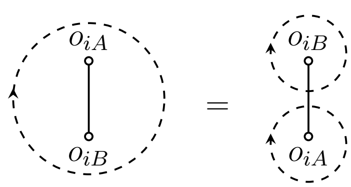


. (5.54)

We make use of the $\mathrm{O}(3)^2 \subset \mathrm{O}(3)^3$ basis freedom such that the lasso (5.53) representing the $\mathbb{Z}_3$ action takes

$$\mathbb{Z}_3: \quad o_{1A} \to o_{2A} \to o_{3A} \to o_{1A}\,. \tag{5.55}$$

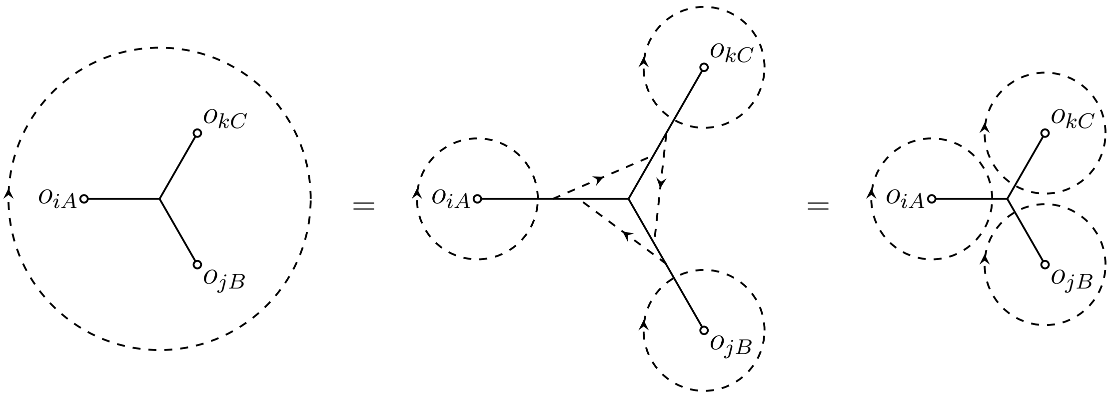
The $\mathbb{Z}_3$ action also gives rise to relations among the dynamical data. For instance, consider the $\mathbb{Z}_3$ action on the operator product of o_{iA} and o_{iB}



. (5.56)

If the vacuum expectation value is taken, possibly in the presence of other local operators, the $\mathbb{Z}_3$ symmetry line can be deformed to shrink in some other patch while picking up the

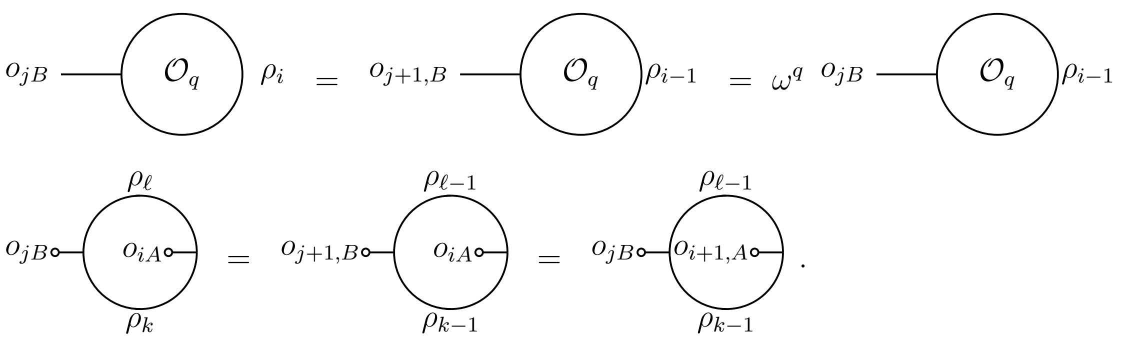
$\mathbb{Z}_3$-charges of other local operators. This process gives rise to identities among correlators. Similarly, the $\mathbb{Z}_3$ action



(5.57)

implies identities among different three-point coefficients, when the sphere vacuum expectation value is taken.

We can nucleate $\mathbb{Z}_3$ loops inside or outside a lasso to change the species of the ρ_i TDLs, resulting in the relations



(5.58)

5.5 Bootstrap constraints

Given the spectral constraints derived in Section 5.3, our goal now is to solve for a minimal defect topological field theory (TFT) with a total of $n_{\mathrm{P}} = 15$ point-like operators, and the number of vacua (local operators) can be $n_{\mathrm{V}} = 2, 4, 6$. For each case, there is one nontrivial $\mathbb{Z}_3$-neutral local operator v and three defect operators o_{iA} on each ρ_i line. The remaining four point-like operators can be two pairs of $\mathbb{Z}_3$-charged local operators u_a, $\bar{u}_a$, two pairs of $\mathbb{Z}_3$ defect operators $w_a \in \mathcal{H}_\alpha$, $\bar{w}_a \in \mathcal{H}_{\alpha^2}$, or a pair of each.

In this section, we delineate constraints of crossing symmetry and modular invariance that were formulated in generality in Section 5.2. For simplicity, we ignore the constraints involving $\mathbb{Z}_3$ defect operators w_a, $\bar{w}_a$, and only consider the part of crossing symmetry that is equivalent to the associativity involving at least one local operator. More general crossing symmetry is deferred to Appendix D.2.

We reserve the $i = 0, 1, 2$ index for the species of the ρ_i line, the $A = 1, 2, 3$ index for the species of the defect operators of each ρ_i line, and the $a = 1, \ldots, n_{\mathbf{2}}$ index for the species of $\mathbb{Z}_3$-charged local operators. Note that the $\mathbb{Z}_3$-charged operators u_a, $\bar{u}_a$ have a $\mathrm{U}(n_{\mathbf{2}})$ basis freedom.

Local operator algebra and associativity

The most general local operator algebra consistent with the $\mathbb{Z}_3$ symmetry is

$$\begin{aligned} v \times v = 1 + \beta v \,, \quad v \times u_a = \sum_b \xi_{ab} u_b \,, \\ u_a \times \bar{u}_b = \delta_{ab} + \xi_{ab} v \,, \quad u_a \times u_b = \sum_c \sigma_{abc} \bar{u}_c \,. \end{aligned} \tag{5.59}$$

The following are the constraints from associativity.

- $\underline{u_a u_b u_c}$

$$\begin{aligned} \sigma_{abc} = \sigma_{bca} \,, \quad \sum_d \sigma_{abd}\, \xi_{ed} = \sum_d \sigma_{bcd}\, \xi_{ad} \,, \\ \sum_e \sigma_{abe}\sigma_{cde} = \sum_e \sigma_{ade}\sigma_{bce} = \sum_e \sigma_{ace}\sigma_{bde} \,. \end{aligned} \tag{5.60}$$

  Hence σ_{abc} is totally symmetric.

- $\underline{u_a \bar{u}_b v}$

$$\xi_{ab} = \bar{\xi}_{ba} \,, \quad \delta_{ab} + \beta \xi_{ab} = \sum_c \xi_{ac}\, \bar{\xi}_{bc} = \sum_c \bar{\xi}_{bc}\, \xi_{ac} \,, \tag{5.61}$$

  The first condition says that ξ_{ab} is Hermitian, which allows us to use the $\mathrm{U}(n_{\mathbf{2}})$ basis freedom to diagonalize ξ_{ab}. Then the second condition, which also encompasses the associativity of $\underline{u_a v v}$, is solved by

$$\xi_{ab} = \xi_a \delta_{ab} \,, \quad \xi_a = \frac{\beta \pm \sqrt{\beta^2 + 4}}{2} \,. \tag{5.62}$$

- $\underline{u_a u_b v}$

$$\sum_c \sigma_{abc} \bar{\xi}^{cd} = \sum_c \xi_{bc} \sigma_{acd} = \sum_c \xi_{ac} \sigma_{bcd} \,. \tag{5.63}$$

- $\underline{u_a u_b \bar{u}_c}$

$$\sum_d \sigma_{abd} \bar{\sigma}_{dce} = \delta_{bc}\delta_{ae} + \xi_{bc}\xi_{ae} \,. \tag{5.64}$$

  In the special case of $a = e$ and $b = c$,

$$\sum_d \sigma_{abd} \bar{\sigma}_{dba} = 1 + \xi_a \xi_b \,. \tag{5.65}$$

Mixed local and ρ defect operators

The most general operator algebra involving mixed local and ρ defect operators is

$$
\begin{gathered}
o_{iA}\ \overset{\rho_i}{\circ\!\!-\!\!\!-\!\!\!-\!\!\circ}\ o_{iB} = \delta_{AB} + \kappa^i_{AB} v + \sum_a \left(\bar{\lambda}^i_{AB;a} u_a + \lambda^i_{AB;a} \bar{u}_a\right) \, , \\
\overset{\rho_i}{-\!\!\!-\!\!\!-\!\!\circ}\ o_{iA}\ v = \sum_B \kappa^i_{AB}\ \overset{\rho_i}{-\!\!\!-\!\!\!-\!\!\circ}\ o_{iB} \, , \qquad \overset{\rho_i}{-\!\!\!-\!\!\!-\!\!\circ}\ o_{iA}\ u_a = \sum_B \lambda^i_{AB;a}\ \overset{\rho_i}{-\!\!\!-\!\!\!-\!\!\circ}\ o_{iB} \, ,
\end{gathered} \tag{5.66}
$$

where κ^i_{AB} and $\lambda^i_{AB;a}$ are both symmetric in A, B, and the $\mathbb{Z}_3$ action (5.57) implies that

$$
\kappa^{i+1}_{AB} = \kappa^i_{AB} \, , \quad \lambda^{i+1}_{AB;a} = \omega^{-1} \lambda^i_{AB;a} \, . \tag{5.67}
$$

The following are the constraints from associativity.

- $\underline{o_{iA} o_{iB} v}$

$$
\begin{aligned}
& o_{iA}\ \overset{\rho_i}{\circ\!\!-\!\!\!-\!\!\!-\!\!\circ}\ o_{iB}\ v \\
&= \delta_{AB} v + \kappa^i_{AB}(1 + \beta v) + \sum_{a,b} \left(\bar{\lambda}^i_{AB;b} \xi_{ba} u_a + \lambda^i_{AB;b} \bar{\xi}_{ba} \bar{u}_a\right) \\
&= \kappa^i_{AB} + \left(\sum_C \kappa^i_{AC} \kappa^i_{BC}\right) v + \sum_C \kappa^i_{AC} \sum_a (\bar{\lambda}^i_{BC;a} u_a + \lambda^i_{BC;a} \bar{u}_a) \, .
\end{aligned} \tag{5.68}
$$

Hence,

$$
\sum_C \kappa^i_{AC} \kappa^i_{BC} = \delta_{AB} + \beta \kappa^i_{AB} \, , \tag{5.69}
$$

which also encompasses the associativity of $\underline{o_{iA} v v}$, and

$$
\sum_C \kappa^i_{AC} \lambda^i_{BC;a} = \sum_b \lambda^i_{AB;b} \bar{\xi}_{ba} \, . \tag{5.70}
$$

- $\underline{o_{iA} o_{iB} u_a}$

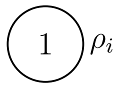
$$
\begin{aligned}
& o_{iA}\ \overset{\rho_i}{\circ\!\!-\!\!\!-\!\!\!-\!\!\circ}\ o_{iB}\ u_a \\
&= \delta_{AB} u_a + \sum_b \kappa^i_{AB} \xi_{ab} u_b + \sum_{b,c} \bar{\lambda}^i_{AB;b} \sigma_{abc} \bar{u}_c + \lambda^i_{AB;a} + \sum_b \lambda^i_{AB;b}\, \xi_{ab} v \\
&= \lambda^i_{AB;a} + \left(\sum_C \lambda^i_{AC;a} \kappa^i_{BC}\right) v + \sum_C \lambda^i_{AC;a} \sum_b (\bar{\lambda}^i_{BC;b} u_b + \lambda^i_{BC;b} \bar{u}_b) \, .
\end{aligned} \tag{5.71}
$$

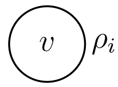
Hence,

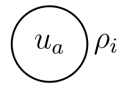
$$
\sum_C \lambda^i_{AC;a} \bar{\lambda}^i_{BC;b} = \delta_{AB} \delta_{ab} + \kappa^i_{AB} \xi_{ab} \, , \quad \sum_C \lambda^i_{AC;a} \lambda^i_{BC;b} = \sum_c \bar{\lambda}^i_{AB;c} \sigma_{abc} \, . \tag{5.72}
$$

ρ action on local operators

Let us study the analog of charge conservation (5.14) for the non-invertible TDLs ρ_i. We will constrain the ρ_i action on local operators,

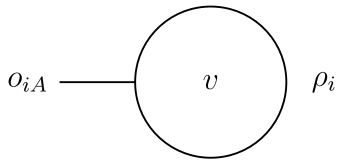


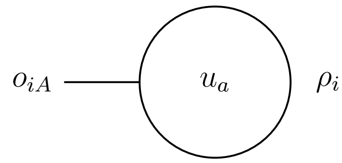
$= \zeta\,,$


$= -\zeta^{-1} v\,,$


$= \omega^i \sum_b R_{ab} \bar{u}_b\,,$ (5.73)

and the lassos on local operators,

$\varepsilon^i_A \equiv$


$,\quad \gamma^i_{aA} \equiv$


. (5.74)

The $\mathbb{Z}_3$ action relations (5.58) imply that

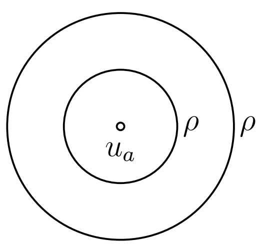
$$\varepsilon^{i+1}_A = \omega^{-i} \varepsilon^i_A\,, \quad \gamma^{i+1}_{aA} = \omega^{-i} \gamma^i_{aA}\,. \tag{5.75}$$

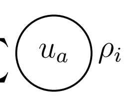
Note that in writing (5.62), we already used the $\mathrm{U}(n_{\mathbf{2}})$ freedom to diagonalize the operator product $u_a\, \bar{u}_b$, so we can no longer use it to simplify R_{ab}. In the following, we make frequent use of the explicit values of the F-symbols

$$(F^{\rho_i,\rho_i,\rho_i}_{\rho_i})_{\mathcal{I},\mathcal{I}} = \zeta^{-1}\,, \quad (F^{\rho_i,\rho_i,\rho_i}_{\rho_i})_{\mathcal{I},\rho_j} = \zeta^{-1}\,, \quad \zeta \equiv \frac{3+\sqrt{13}}{2} \tag{5.76}$$

from (D.7).

First, let us revisit the requirement that u_a transforms as a representation of the fusion ring.[13] The consideration of



$= \quad u_a + \sum_i$


(5.77)

leads to a constraint

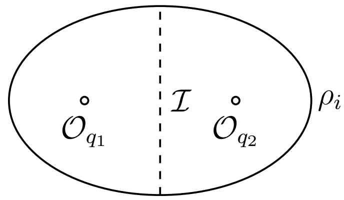
$$\sum_c R_{ac} \overline{R}_{cb} = \delta_{ab} + \sum_i \omega^i R_{ab} = \delta_{ab}\,, \tag{5.78}$$

where the left side comes from shrinking the inner and outer ρ loops consecutively, and the right side from fusing them before shrinking.[14]

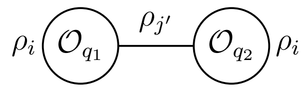
[13] The representation given in (5.35) was specialized to a particular basis for u_a. Here we prioritize the use of the $\mathrm{U}(n_{\mathbf{2}})$ basis freedom to diagonalize ξ_{ab} in (5.62), so the requirement that u_a transforms as a representation needs to be rewritten in a basis-independent fashion.

[14] The fusion of the two ρ TDLs can be understood as an F-move followed by the shrinking of the ρ loop.

Now, following the $\downarrow$ direction in (5.14), we circle ρ_i on the operator product of local operators, and apply the F-move to obtain



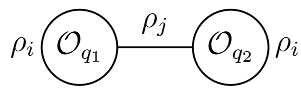
$$= \sum_{j'}$$



$$. \tag{5.79}$$

Using the $\mathbb{Z}_3$ action (5.51), we can simplify the sum of dumbbells to

$$3$$

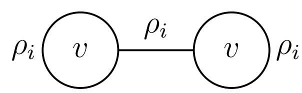


$$\Bigg|_{\mathbb{Z}_3\text{-charge } -(q_1+q_2)} , \tag{5.80}$$

where j is arbitrary. We might as well set $j = i$. In the following, we equate the above to the $\rightarrow\downarrow$ direction of (5.14) where the local operator product is taken first.

- <u>$v \times v = 1 + \beta v$</u>

$$\zeta^{-3}\,(1+\beta v) + 3\,\zeta^{-1}$$



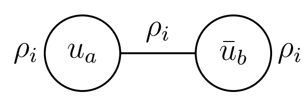
$$\Bigg|_{\mathbb{Z}_3\text{-neutral}} = \zeta - \zeta^{-1}\beta v\,. \tag{5.81}$$

Hence,

$$\sum_A (\varepsilon^i_A)^2 = \sqrt{13}\,, \quad \sum_{A,B} \varepsilon^i_A \kappa^i_{AB} \varepsilon^i_B = -\frac{\sqrt{13}}{3}\zeta^{-1}\beta\,. \tag{5.82}$$

- <u>$u_a \times \bar{u}_b = \delta_{ab} + \xi_{ab} v$</u>

$$\zeta^{-1}\left(\delta_{ab} + \sum_{c,d} R_{ac}\overline{R}_{bd}\xi_{dc} v\right) + 3\,\zeta^{-1}$$



$$\Bigg|_{\mathbb{Z}_3\text{-neutral}} \\ = \zeta\delta_{ab} - \zeta^{-1}\xi_{ab}v\,. \tag{5.83}$$

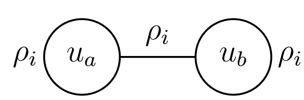
Hence,

$$\sum_A \gamma^i_{aA}\bar{\gamma}^i_{bA} = \zeta\delta_{ab}\,, \tag{5.84}$$

$$\sum_{A,B} \gamma^i_{aA}\bar{\gamma}^i_{bB}\kappa^i_{AB} = -\frac{1}{3}\left(\xi_{ab} + \sum_{c,d} R_{ac}\overline{R}_{bd}\xi_{dc}\right)\,. \tag{5.85}$$

- <u>$u_a \times u_b = \sum_c \sigma_{abc}\bar{u}_c$</u>

$$\zeta^{-1}\omega^{-i}\sum_{d,e,f} R_{ad}R_{be}\bar{\sigma}_{def}u_f + 3\,\zeta^{-1}$$



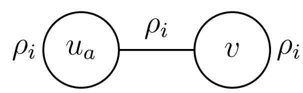
$$\Bigg|_{\mathbb{Z}_3\text{-charge } 1} \\ = \omega^{-i}\sum_{c,f}\sigma_{abc}\overline{R}_{cf}u_f\,. \tag{5.86}$$

Hence,

$$\sum_{A,B}\gamma^i_{aA}\gamma^i_{bB}\bar{\lambda}^i_{AB;f} = \frac{1}{3}\omega^{-i}\left(\zeta\sum_c \sigma_{abc}\overline{R}_{cf} - \sum_{d,e}R_{ad}\rho_{be}\bar{\sigma}_{def}\right)\,. \tag{5.87}$$

- $\underline{u_a \times v = \sum_b \xi_{ab}u_b}$

$$-\,\zeta^{-2}\omega^i\sum_{b,c}R_{ab}\bar{\xi}_{bc}\bar{u}_c + 3\,\zeta^{-1}$$



$$\Big|_{\mathbb{Z}_3\text{-charge } -1} \tag{5.88}$$

$$= \omega^i\sum_{b,c}\xi_{ab}R_{bc}\bar{u}_c\,.$$

Hence,

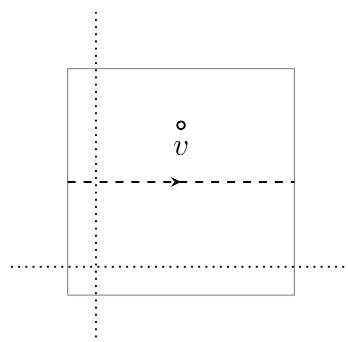
$$\sum_A \gamma^i_{aA}\lambda^i_{AB;c}\varepsilon^i_B = \frac{1}{3}\omega^i\sum_b\left(\zeta\xi_{ab}R_{bc} + \zeta^{-1}R_{ab}\bar{\xi}_{bc}\right)\,. \tag{5.89}$$

Torus one-point modular invariance

Consider the torus one-point modular invariance (5.25) in the special case of

$$\mathcal{L}_2 = \mathcal{L}_{\mathcal{O}} = \mathcal{I}\,,\quad \mathcal{L}_3 = \overline{\mathcal{L}}_4 = \mathcal{L}_1\,. \tag{5.90}$$

- $\underline{v \text{ with } \mathbb{Z}_3 \text{ symmetry line}}$



(5.91)

Let us denote the three-point coefficient of v with $\mathbb{Z}_3$ defect operators by

$$\tilde{\xi}_a = c(v, w_a, \bar{w}_a)\,. \tag{5.92}$$

Let us write down

$$\text{vertical cut} = \text{horizontal cut} \tag{5.93}$$

for different numbers of vacua.

(a) $n_{\mathrm{V}} = 6$

$$\begin{aligned} 0 &= c(v,v,v) + \omega \sum_a c(v, u_a, \bar{u}_a) + \omega^2 \sum_a c(v, \bar{u}_a, u_a) \\ &= c(v,v,v) - \sum_a c(v, u_a, \bar{u}_a) \\ &= \beta - \xi_1 - \xi_2 \,. \end{aligned} \tag{5.94}$$

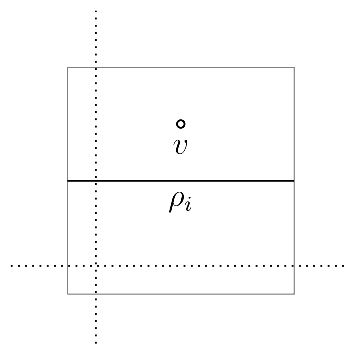
(b) $n_{\mathrm{V}} = 4$

$$\tilde{\xi} = \beta - \xi \,. \tag{5.95}$$

(c) $n_{\mathrm{V}} = 2$

$$\tilde{\xi}_1 + \tilde{\xi}_2 = \beta \,. \tag{5.96}$$

- v with ρ line



(5.97)

Under the vertical cut,

$$\sum_A c(v, o_{iA}, o_{iA}) = \sum_A \kappa^i_{AA} = \mathrm{tr}(\kappa^i) \,, \tag{5.98}$$

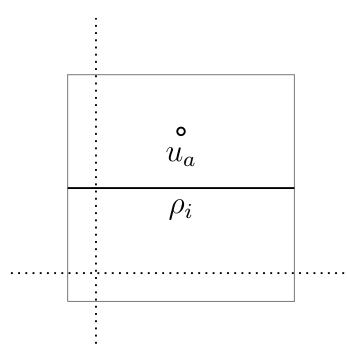
and under the horizontal cut.

$$-\zeta^{-1} c(v,v,v) = -\zeta^{-1} \beta \,, \tag{5.99}$$

Hence,

$$\mathrm{tr}(\kappa^i) = -\zeta^{-1} \beta \,. \tag{5.100}$$

- u_a with ρ line



(5.101)

Under the vertical cut

$$\sum_A c(u_a, o_{iA}, o_{iA}) = \sum_A \lambda^i_{AA;a}\,, \tag{5.102}$$

and under the horizontal cut,

$$\sum_{b,c} \overline{R}_{bc} c(u_a, u_b, u_c) = \sum_{b,c} \overline{R}_{bc} \sigma_{abc}\,. \tag{5.103}$$

Hence,

$$\sum_A \lambda^i_{AA;a} = \sum_{b,c} \overline{R}_{bc} \sigma_{abc}\,. \tag{5.104}$$

5.6 Topological field theory with Haagerup $\mathcal{H}_3$ symmetry

This section analyzes the bootstrap constraints delineated in the previous section. We first narrow down the local operator algebra to a handful of possibilities, and then proceed to construct a topological field theory with six vacua realizing the Haagerup $\mathcal{H}_3$ fusion category.

Local operator algebra

To solve for a defect topological field theory, we begin by examining the associativity of local operators detailed in Section 5.5. There we used the $\mathrm{U}(n_{\mathbf{2}})$ basis freedom for u_a and $\bar{u}_a$ to put ξ_{ab} into diagonal form, and used associativity to constrain the possible eigenvalues; the result was

$$\xi_{ab} = \xi_a \delta_{ab}\,, \quad \xi_a = \frac{\beta \pm \sqrt{\beta^2 + 4}}{2}\,. \tag{5.105}$$

In this basis, (5.63) becomes

$$\xi_a\, \sigma_{abc} = \xi_b\, \sigma_{abc} = \xi_c\, \sigma_{abc}\,. \tag{5.106}$$

Then for any pair (a, b) such that $\xi_a \neq \xi_b$, it follows that $\sigma_{abc} = 0$, *i.e.* the operator product $u_a u_b$ must vanish. We have the following scenarios:

(a) $n_{\mathrm{V}} = 2$. There is no $\mathbb{Z}_3$-charged operator.

(b) $n_{\mathrm{V}} = 4$. There is a single pair of $\mathbb{Z}_3$-charged operators. Then (5.65) reads

$$\sigma^2 = 1 + \xi^2\,. \tag{5.107}$$

(c) $n_{\mathrm{V}} = 6$, and there are two pairs of $\mathbb{Z}_3$-charged operators with *different* ξ_a. Because σ_{abc} with mixed indices vanish, (5.65) becomes

$$0 = 1 + \xi_1 \xi_2\,, \quad \sigma^2_{111} = 1 + \xi_1^2\,, \quad \sigma^2_{222} = 1 + \xi_2^2\,. \tag{5.108}$$

We can use the residual $\mathrm{U}(1)^2$ basis freedom to make σ_{aaa} real and non-negative. Without loss of generality,

$$\xi_1 = \frac{\beta - \sqrt{\beta^2+4}}{2}\,, \quad \xi_2 = \frac{\beta + \sqrt{\beta^2+4}}{2}\,,$$
$$\sigma_{111} = \sqrt{1+\xi_1^2}\,, \quad \sigma_{222} = \sqrt{1+\xi_2^2}\,. \tag{5.109}$$

(d) $n_{\mathrm{V}} = 6$, and there are two pairs of $\mathbb{Z}_3$-charged operators with *the same* ξ_a. It can be shown that the associativity of local operators admits a unique solution

$$\beta = 2i\,, \quad \xi_1 = \xi_2 = i\,, \quad \sigma_{abc} = 0\,. \tag{5.110}$$

This case will be ruled out momentarily.

To proceed, we examine the associativity of $\underline{o_{iA} o_{iB} v}$ detailed in Section 5.5. The first condition (5.69)

$$\sum_C \kappa^i_{AC} \kappa^i_{BC} = \delta_{AB} + \beta \kappa^i_{AB} \tag{5.111}$$

implies that κ^i_{AB} are 3×3 matrices with each eigenvalue taking one of two possible values

$$\text{each } eigval(\kappa^i) = \frac{\beta \pm \sqrt{\beta^2+4}}{2}\,. \tag{5.112}$$

And it follows from the torus one-point modular invariance condition (5.100) that

$$tr(\kappa^i) = -\zeta^{-1}\beta\,, \quad \zeta = \frac{3+\sqrt{13}}{2}\,. \tag{5.113}$$

We immediately see that (5.110) fails to satisfy this constraint, so (d) is ruled out. In the following, we analyze the two inequivalent possibilities for the eigenvalues $\underline{- - -}$ and $\underline{+ - -}$ as labeled by the signs taken in (5.112). The $\underline{+ + +}$ and $\underline{+ + -}$ cases are equivalent to $\underline{- - -}$ and $\underline{+ - -}$ by the redefinition $v \to -v$.

I. $\underline{- - -}$ The torus one-point modular invariance condition (5.100) becomes

$$3 \times \frac{\beta - \sqrt{\beta^2+4}}{2} = -\zeta^{-1}\beta \quad \Rightarrow \quad \beta = 3\,. \tag{5.114}$$

As all eigenvalues of κ^i_{AB} are the same, in any basis for o_{iA},

$$\kappa^i_{AB} = -\zeta^{-1}\delta_{AB}\,. \tag{5.115}$$

Besides the local operator algebra, the action of the ρ TDL on the $\mathbb{Z}_3$-charged operators is constrained as follows. First, recall from (5.78) that

$$\sum_c R_{ac}\overline{R}_{cb} = \delta_{ab} \,. \tag{5.116}$$

Second, by the use of (5.84) and (5.115), we can evaluate the left side of (5.85),

$$\sum_{A,B} \gamma^i_{aA}\bar{\gamma}^i_{bB}\kappa^i_{AB} = -\zeta^{-1}\sum_A \gamma^i_{aA}\bar{\gamma}^i_{bA} = -\delta_{ab} \,, \tag{5.117}$$

and then (5.85) becomes

$$-\delta_{ab} = -\frac{1}{3}\left(\xi_{ab} + \sum_{c,d} R_{ad}\overline{R}_{bc}\xi_{dc}\right) \,. \tag{5.118}$$

Let us examine the scenarios (a)(b)(c).

a) If $n_{\mathrm{V}} = 2$, then $\beta = 3$ completely specifies the local operator algebra.

b) If $n_{\mathrm{V}} = 4$, then (5.118) becomes a scalar equation reading

$$-1 = -\frac{1}{3}\left(\xi + R\overline{R}\xi\right) = \frac{2}{3}\xi \,, \qquad \xi \equiv \xi_{11} = \xi_1 \,, \quad R \equiv R_{11} \,, \tag{5.119}$$

which contradicts with the allowed ξ values (5.105) given $\beta = 3$. Hence this case is ruled out.

c) If $n_{\mathrm{V}} = 6$, then to be consistent with torus one-point modular invariance (5.94) and the allowed ξ_a values (5.105), we set without loss of generality

$$\xi_1 = -\zeta^{-1} \,, \quad \xi_2 = \zeta \,. \tag{5.120}$$

By (5.108), the non-vanishing three-point coefficients of $\mathbb{Z}_3$-charged operators are

$$\sigma_{111} = \sqrt{1+\zeta^{-2}} = \sqrt{\frac{13-3\sqrt{13}}{2}} \,, \quad \sigma_{222} = \sqrt{1+\zeta^2} = \sqrt{\frac{13+3\sqrt{13}}{2}} \,. \tag{5.121}$$

We have thus completely specified the operator product algebra. Together with (5.116) and (5.118), the action of the ρ TDL on $\mathbb{Z}_3$-charged local operators are restricted to be

$$\widehat{\rho}(u_a) = u_a \,\rho_i = \sum_b R_{ab}\bar{u}_b \,, \quad R = \theta \times \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \,, \quad \theta \in \mathbb{C} \,, \quad |\theta| = 1 \,. \tag{5.122}$$

II. $\underline{+--}$ The torus one-point modular invariance condition (5.100) becomes

$$\frac{3\beta - \sqrt{\beta^2+4}}{2} = -\zeta^{-1}\beta\,, \quad \Rightarrow \quad \beta = \frac{1}{\sqrt{3}}\,. \tag{5.123}$$

The values of ξ_{ab} and σ_{abc} are fixed by β through (5.105), (5.107), and (5.109). The bootstrap analysis of this possibility is more complicated than the $\underline{---}$ case, so we leave it for future work. However, some hints pointing towards the existence of a defect TFT of case II(b) with $n_{\rm V} = 4$, and arguments for the non-existence in cases II(a) with $n_{\rm V} = 2$ and II(c) with $n_{\rm V} = 6$ can be found in Section 5.7.

In the next section, we complete the construction of a TFT of case I(c) with $n_{\rm V} = 6$ and $\beta = 3$. The reader interested in boundary conditions can safely proceed to Section 5.7.

Topological field theory with six vacua

We now construct the rest of the defect TFT data in case I(c) with $n_{\rm V} = 6$ and $\beta = 3$, and solve **all** the consistency conditions outlined in Section 5.2.

It turns out that a good point of attack is the associativity of $\underline{o_{iA}o_{iB}v}$. The condition (5.70) in the basis diagonalizing ξ_{ab} (5.62) reads

$$\sum_C \kappa^i_{AC}\lambda^i_{BC;a} = \xi_a \lambda^i_{AB;a}\,, \tag{5.124}$$

which implies that for fixed A and a, $\lambda^i_{AB;a}$ must be an eigenvector of κ^i with eigenvalue ξ_a; otherwise $\lambda^i_{BC;a}$ vanishes. But because κ^i does not have ζ as an eigenvalue, it follows that

$$\lambda^i_{AB;2} = 0\,. \tag{5.125}$$

Note that the vanishing of $\lambda^i_{AB;2}$ is consistent with (5.72).

By considering the vanishing $\lambda^i_{AB;2}$, we can determine the ρ action, which we found to be parameterized by $\theta \in \mathbb{C}$ in (5.122). The nontrivial part of (5.87) with $f = 2$ becomes

$$0 = \zeta\sigma_{111}\bar{\theta} - \theta^2\sigma_{222}\,, \tag{5.126}$$

which by the use of (5.121) leads to

$$|\theta| = \zeta\frac{\sigma_{111}}{\sigma_{222}} = \zeta\sqrt{\frac{1+\zeta^{-2}}{1+\zeta^2}} = 1\,, \quad \theta^3 = 1\,. \tag{5.127}$$

Up to the relabeling of ρ_i,

$$\theta = 1\,. \tag{5.128}$$

For the non-vanishing $\lambda^i_{AB;1}$, it is convenient to define a normalized

$$\hat{\lambda}^i_{AB} \equiv \frac{\lambda^i_{AB;1}}{\sigma_{111}}\,, \tag{5.129}$$

and write the associativity of $\underline{o_{iA}o_{iB}u_a}$ (5.72) and the modular invariance condition (5.104) in matrix notation as (recall that $\hat{\lambda}^i$ is a symmetric matrix)

$$\hat{\lambda}^i\bar{\hat{\lambda}}^i = 1\,, \quad \hat{\lambda}^i\hat{\lambda}^i = \bar{\hat{\lambda}}^i\,, \quad \mathrm{T}r\,\lambda^i = 0\,. \tag{5.130}$$

The first equation says that $\hat{\lambda}^i$ is unitary, and combined with the second equation implies that $(\hat{\lambda}^i)^3 = 1$. The third equation then tells us that $\hat{\lambda}^i$ has eigenvalues

$$eigvals(\hat{\lambda}^i) = \{1,\, \omega,\, \omega^2\}\,. \tag{5.131}$$

We now prove that $\hat{\lambda}$ (suppressing superscript i) must be diagonalizable by an O(3) matrix. For convenience define $\Omega = \mathrm{diag}(1,\omega,\omega^2)$. Because $\hat{\lambda}$ is unitary, it can always be diagonalized by a unitary matrix Z, *i.e.* $\hat{\lambda} = Z^\dagger \Omega Z$. For $\hat{\lambda}$ to be symmetric, we must have

$$Z^\dagger W Z = Z^T W \bar{Z} \quad \Rightarrow \quad (ZZ^T)W\overline{(ZZ^T)} = W. \tag{5.132}$$

Let us define $A = ZZ^T$. The $(1,1)$-component of the matrix equation (5.132) reads

$$|A_{11}|^2 + \omega|A_{12}|^2 + \omega^2|A_{13}|^2 = 1\,, \tag{5.133}$$

where we used the fact that A is symmetric. Now for the above equation to have a solution, we must have $|A_{12}|^2 = |A_{13}|^2$, since otherwise the imaginary part cannot match. Let us call this value $x \equiv |A_{12}|^2 = |A_{13}|^2$. Then $|A_{11}|^2 = 1 + x$. Proceeding similarly, we arrive at the following matrix

$$A = \begin{pmatrix} e^{ia_{11}}\sqrt{1+x} & e^{ia_{12}}\sqrt{x} & e^{ia_{13}}\sqrt{x} \\ e^{ia_{12}}\sqrt{x} & e^{ia_{22}}\sqrt{1+x} & e^{ia_{23}}\sqrt{x} \\ e^{ia_{13}}\sqrt{x} & e^{ia_{23}}\sqrt{x} & e^{ia_{33}}\sqrt{1+x} \end{pmatrix}, \tag{5.134}$$

where a_{ij} are arbitrary phases. Finally, A must be unitary since

$$AA^\dagger = (ZZ^T)(ZZ^T)^\dagger = Z(Z^T\bar{Z})Z^\dagger = ZZ^\dagger = 1\,, \tag{5.135}$$

which means that

$$(AA^\dagger)_{11} = 1 + 3x = 1\,. \tag{5.136}$$

Hence $x = 0$ and we end up with

$$A = \begin{pmatrix} e^{ia_{11}} & & \\ & e^{ia_{22}} & \\ & & e^{ia_{33}} \end{pmatrix} . \tag{5.137}$$

The O(3) matrix of interest is given by[15]

$$O \equiv \sqrt{A^{-1}} Z \,. \tag{5.138}$$

We can therefore use the O(3) freedom to set

$$\hat{\lambda}^i_{AB;1} = \omega^{A-1-i} \delta_{AB} \quad \Rightarrow \quad \lambda^i_{AB;1} = \omega^{A-1-i} \sigma_{111} \delta_{AB} = \omega^{A-1-i} \sqrt{1+\zeta^{-2}}\, \delta_{AB} \,. \tag{5.139}$$

Let us summarize the solution we found so far into

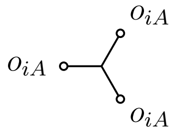
$$\begin{gathered} v \times v = 1 + 3v \,, \quad u_1 \times \bar{u}_1 = 1 - \zeta^{-1} v \,, \quad u_2 \times \bar{u}_2 = 1 + \zeta v \,, \\ u_1 \times \bar{u}_2 = 0 \,, \quad u_1 \times u_1 = \sqrt{1+\zeta^{-2}}\, \bar{u}_1 \,, \quad u_2 \times u_2 = \sqrt{1+\zeta^2}\, \bar{u}_2 \,, \\ o_{iA} \xrightarrow{\ \rho_i\ } o_{iB} = \begin{cases} 1 - \zeta^{-1} v + \sqrt{1+\zeta^{-2}} \left(\omega^{i+1-A} u_1 + \omega^{A-1-i} \bar{u}_1 \right) & A = B \,, \\ 0 & A \neq B \,. \end{cases} \end{gathered} \tag{5.140}$$

We proceed to solve the more general crossing symmetry involving four ρ_i defect operators. Some analytic progress is made in Appendix D.2, such as deriving a selection rule (D.20), but eventually we resort to computer numerics to find a solution.[16] Up to operator relabeling and sign redefinitions, the solution appears to be unique. The non-vanishing defect three-point coefficients are (vacuum expectation values are implicitly

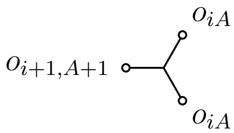
[15] The fact that O is an O(3) matrix follows from $OO^\dagger = OO^T = 1$. The first equality implies that $O^\dagger = O^T$, or equivalently that O is real, and the second equality is orthogonality.

[16] Up to this point in the main text, no assumption about reflection-positivity was needed. However, both Appendix D.2 and the computer numerics assume reflection-positivity.

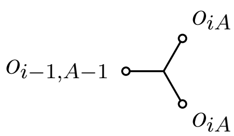
taken)



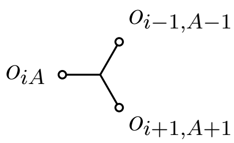
$$= \sqrt{\frac{17\sqrt{13}-52}{3}}\,,$$



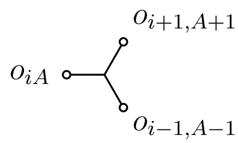
$$= -\frac{1}{2}\sqrt{\frac{1}{3}\left(-13+11\sqrt{13}+\sqrt{78\left(7\sqrt{13}-23\right)}\right)}\,,$$



$$= \frac{1}{2}\sqrt{\frac{1}{3}\left(-13+11\sqrt{13}-\sqrt{78\left(7\sqrt{13}-23\right)}\right)}\,, \tag{5.141}$$



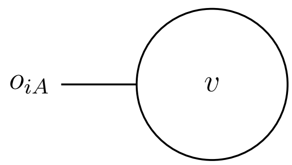
$$=$$



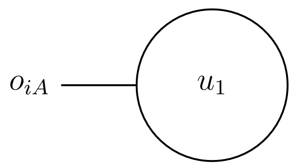
$$= -\sqrt{\frac{13+7\sqrt{13}}{6}}\,.$$

Note an interesting "superselection" rule: If we define three "sectors" labeled by $i - A \mod 3$, then all non-vanishing three-point coefficients are those that involve defect operators in a single sector.

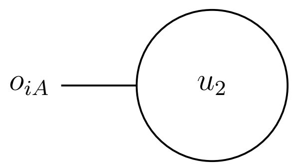
Finally, we can solve the full set of modular invariance constraints, which are linear in the lassos. We find a solution where some of the lassos are given by (vacuum expectation values are implicitly taken)



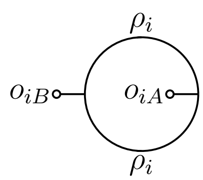
$$= \varepsilon^i_A = -\sqrt{\frac{\sqrt{13}}{3}}\,,$$



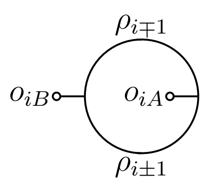
$$= \gamma^i_{1A} = \sqrt{\frac{\zeta}{3}}\,\omega^{A-1-i}e^{-i\phi_\gamma}\,, \quad \phi_\gamma = \frac{\pi}{4} + \frac{1}{2}\arccos\frac{1}{3+\sqrt{13}}\,,$$



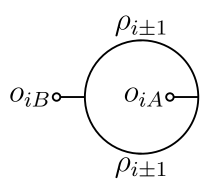
$$= \gamma^i_{2A} = -\sqrt{\frac{\zeta}{3}}\,\omega^{1-A-i}\,,$$



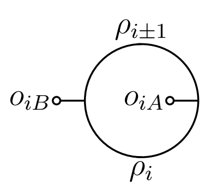
$$= \begin{cases} \dfrac{-1+5\sqrt{13}}{18} & A = B\,, \\[2em] \dfrac{-7-\sqrt{13}}{18} & A \neq B\,, \end{cases}$$



$$= \begin{cases} \dfrac{17+5\sqrt{13}}{18} & A = B\,, \\[2em] \dfrac{-7-\sqrt{13}}{18} & A \neq B\,, \end{cases}$$



$$= \begin{cases} \dfrac{5+2\sqrt{13}\pm\sqrt{15+6\sqrt{13}}}{18} & A = B\,, \\[2em] \dfrac{-11+\sqrt{13}\pm\sqrt{6\left(61+19\sqrt{13}\right)}}{36} & A \neq B\,, \end{cases}$$



$$= \begin{cases} \dfrac{5+2\sqrt{13}\pm\sqrt{15+6\sqrt{13}}}{18} & A = B\,, \\[2em] \dfrac{7+\sqrt{13}\mp\sqrt{798+222\sqrt{13}}}{36} & A \neq B\,, \end{cases} \tag{5.142}$$

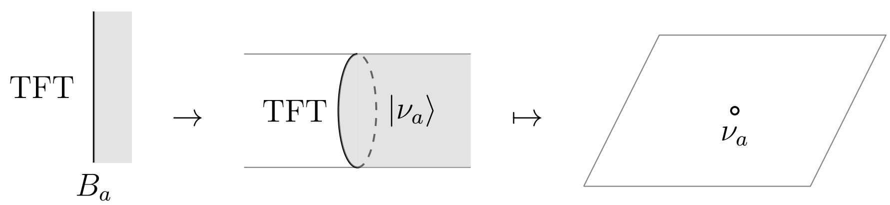
and the rest are related to the above via (5.58).

We have thus completed the construction of a defect topological field theory whose defining data solve **all** the consistency conditions outlined in Section 5.2.

5.7 Boundary conditions and NIM-reps

We can extend our topological field theory (TFT) further by considering boundaries. Given the bijection between $\mathcal{C}$-symmetric TFTs and $\mathcal{C}$-module categories argued by Thorngren and Wang [161], the fact that the Haagerup $\mathcal{H}_3$ fusion category has exactly three indecomposable module categories [93], with two, four, and six simple objects, is strongly suggestive of a connection to the three "minimal" possible TFTs (with $n_{\mathrm{P}} = 15$ point-like operators) in Table 5.2, with two, four, and six vacua. However, except in special cases, it is not known how to extract the axiomatic defect TFT data from the module category. In this section, without assuming any prior knowledge of module categories, we use the bootstrap results on the axiomatic defect TFT data to construct boundaries and examine their fusion with TDLs. This section can be read independently of Section 5.6.

We first review some nontrivial results in open/closed TFTs. The admissible boundary conditions of a TFT are direct sums of a set of elementary boundary conditions B_a, which are related to the so-called Cardy states ν_a by folding the boundary into a circle and invoking the state-operator map,



(5.143)

By solving the consistency conditions of open/closed TFT, Moore and Segal [129] established an explicit formula for the Cardy states ν_a in terms of the projectors π_a introduced in Section 5.2,

$$\nu_a = \frac{\pi_a}{\sqrt{\langle \pi_a \rangle}} \,. \tag{5.144}$$

In particular, the number of Cardy states is the same as the number of vacua n_{V}.

Let $\{\mathcal{L}_i \mid i = 1, \ldots, r\}$ be the set of simple TDLs, N_{ij}^k the fusion coefficients, and $\{\nu_a \mid a = 1, \ldots, n_{\mathrm{V}}\}$ the set of Cardy states. The fusion of any TDL with an admissible boundary must give another admissible boundary. Therefore, the action of TDLs on the Cardy states must furnish a non-negative integral matrix representation (NIM-rep): a set of $n_{\mathrm{V}} \times n_{\mathrm{V}}$ non-negative integral matrices $(\mathcal{N}_i)_a{}^b$, one for each line i, such that

$$\sum_c (\mathcal{N}_i)_a{}^c (\mathcal{N}_j)_c{}^b = N_{ij}^k \, (\mathcal{N}_k)_a{}^b \,. \tag{5.145}$$

Given that we have narrowed down the full local operator algebra to a few possibilities in Section 5.6, it is straightforward to compute the projector basis and examine the action of the TDLs on the projectors. To condense the discussion, we present formulae that apply to the entire family of local operator algebras, parameterized by β, that solve the associativity of local operators. Of course, we have seen that the associativity of $\underline{o_{iA}o_{iB}v}$ requires $\beta = 3$ or $\beta = \frac{1}{\sqrt{3}}$.

(a) Consider $n_{\mathrm{V}} = 2$. The projector basis for

$$v \times v = 1 + \beta v \tag{5.146}$$

is given by

$$\pi_1 = \frac{\zeta - v}{\sqrt{4+\beta^2}}\,, \quad \pi_2 = \frac{\zeta^{-1} + v}{\sqrt{4+\beta^2}}\,. \tag{5.147}$$

According to (5.144), the Cardy states are

$$\nu_1 = \frac{\sqrt[4]{4+\beta^2}}{\sqrt{\zeta}}\,\pi_1\,, \quad \nu_2 = \sqrt[4]{4+\beta^2}\sqrt{\zeta}\,\pi_2\,. \tag{5.148}$$

I. When $\beta = 3$, they furnish a NIM-rep

$$\mathcal{N}_\alpha = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \quad \mathcal{N}_\rho = \begin{pmatrix} 3 & 1 \\ 1 & \end{pmatrix}. \tag{5.149}$$

II. When $\beta = \frac{1}{\sqrt{3}}$, the representation

$$\mathcal{N}_\alpha = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \quad \mathcal{N}_\rho = \begin{pmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} \tag{5.150}$$

is not NIM.

(b) For $n_{\mathrm{V}} = 4$, the local operator algebra is given by

$$v \times v = 1 + \beta v\,, \quad u \times \bar{u} = 1 + \xi v\,, \quad u \times u = \sqrt{1+\xi^2}\,\bar{u}\,, \quad \xi = \frac{\beta \pm \sqrt{\beta^2+4}}{2}\,. \tag{5.151}$$

There are two possible choices for ξ, and we can construct the projector basis

$$\pi_a = \begin{cases} \dfrac{\epsilon\,\xi^{-1} + \epsilon\, v + \sqrt[4]{4+\beta^2}\,\sqrt{\epsilon\,\xi^{-1}}\,(\omega^{a-1}u + \omega^{1-a}\bar{u})}{3\sqrt{4+\beta^2}}\,, & a = 1,2,3\,, \\ \dfrac{\epsilon\,\xi - \epsilon\, v}{\sqrt{4+\beta^2}}\,, & a = 4\,, \end{cases} \tag{5.152}$$

where $\epsilon \equiv sign(\xi)$. The Cardy states (5.144) are then

$$\nu_a = \begin{cases} \left(\dfrac{\epsilon\,\xi^{-1}}{3\sqrt{4+\beta^2}}\right)^{-\frac{1}{2}} \pi_a\,, & a = 1,2,3\,, \\ \left(\dfrac{\epsilon\,\xi}{\sqrt{4+\beta^2}}\right)^{-\frac{1}{2}} \pi_a\,, & a = 4\,. \end{cases} \tag{5.153}$$

Whether they furnish a NIM-rep depends on how the ρ TDL acts, that is, on R.

I. When $\beta = \frac{1}{\sqrt{3}}$, we find that for ξ taking either value, that is $\epsilon = \pm$, there is exactly one value of R that gives rise to a NIM-rep:

$$\begin{aligned} &\epsilon = +\,, \quad R = 1: \quad \mathcal{N}_\alpha = \begin{pmatrix} & 1 & & \\ & & 1 & \\ 1 & & & \\ & & & 1 \end{pmatrix}, \quad \mathcal{N}_\rho = \begin{pmatrix} 1 & & & 1 \\ & & 1 & 1 \\ & 1 & & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \\ &\epsilon = -\,, \quad R = -1: \quad \mathcal{N}_\alpha = \begin{pmatrix} & 1 & & \\ & & 1 & \\ 1 & & & \\ & & & 1 \end{pmatrix}, \quad \mathcal{N}_\rho = \begin{pmatrix} & 1 & 1 & 1 \\ 1 & 1 & & 1 \\ 1 & & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}. \end{aligned} \tag{5.154}$$

II. When $\beta = 3$, we instead have

$$\begin{aligned} \epsilon = +: \ \mathcal{N}_\rho &= \begin{pmatrix} \frac{2\cos\phi}{3} & -\frac{\cos\phi}{3} + \frac{\sin\phi}{\sqrt{3}} & -\frac{\cos\phi}{3} - \frac{\sin\phi}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{\cos\phi}{3} + \frac{\sin\phi}{\sqrt{3}} & -\frac{\cos\phi}{3} - \frac{\sin\phi}{\sqrt{3}} & \frac{2\cos\phi}{3} & \frac{1}{\sqrt{3}} \\ -\frac{\cos\phi}{3} - \frac{\sin\phi}{\sqrt{3}} & \frac{2\cos\phi}{3} & -\frac{\cos\phi}{3} + \frac{\sin\phi}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 3 \end{pmatrix}, \\ \epsilon = -: \ \mathcal{N}_\rho &= \begin{pmatrix} 1 + \frac{2\cos\phi}{3} & 1 - \frac{\cos\phi}{3} + \frac{\sin\phi}{\sqrt{3}} & 1 - \frac{\cos\phi}{3} - \frac{\sin\phi}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 1 - \frac{\cos\phi}{3} + \frac{\sin\phi}{\sqrt{3}} & 1 - \frac{\cos\phi}{3} - \frac{\sin\phi}{\sqrt{3}} & 1 + \frac{2\cos\phi}{3} & \frac{1}{\sqrt{3}} \\ 1 - \frac{\cos\phi}{3} - \frac{\sin\phi}{\sqrt{3}} & 1 + \frac{2\cos\phi}{3} & 1 - \frac{\cos\phi}{3} + \frac{\sin\phi}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \end{pmatrix}, \end{aligned} \tag{5.155}$$

where $R = e^{i\phi}$. As we can see the representation is not NIM.

(c) Finally, consider $n_{\mathrm{V}} = 6$. The local operator algebra is

$$\begin{aligned} &v \times v = 1 + \beta v\,, \quad u_1 \times \bar{u}_1 = 1 + \xi_1 v\,, \quad u_2 \times \bar{u}_2 = 1 + \xi_2 v\,, \\ &u_1 \times \bar{u}_2 = 0\,, \quad u_1 \times u_1 = \sqrt{1+\xi_1^2}\,\bar{u}_1\,, \quad u_2 \times u_2 = \sqrt{1+\xi_2^2}\,\bar{u}_2\,, \\ &\xi_{1,2} = \frac{\beta \mp \sqrt{\beta^2+4}}{2}\,, \end{aligned} \tag{5.156}$$

and the projector basis is given by

$$\pi_a = \begin{cases} \dfrac{\xi_2 - v + \sqrt[4]{4+\beta^2}\,\sqrt{\xi_2}\,(\omega^{a-1}u_1 + \omega^{1-a}\bar{u}_1)}{3\sqrt{4+\beta^2}}\,, & a = 1,2,3\,, \\ \dfrac{-\xi_1 + v + \sqrt[4]{4+\beta^2}\,\sqrt{-\xi_1}\,(\omega^{a-1}u_2 + \omega^{1-a}\bar{u}_2)}{3\sqrt{4+\beta^2}}\,, & a = 4,5,6\,. \end{cases} \tag{5.157}$$

The Cardy states (5.144) are

$$\nu_a = \begin{cases} \left(\dfrac{\xi_2}{3\sqrt{4+\beta^2}}\right)^{-\frac{1}{2}} \pi_a\,, & a = 1,2,3\,, \\ \left(\dfrac{-\xi_1}{3\sqrt{4+\beta^2}}\right)^{-\frac{1}{2}} \pi_a\,, & a = 4,5,6\,. \end{cases} \tag{5.158}$$

Clearly, the triples (ν_1, ν_2, ν_3) and (ν_4, ν_5, ν_6) each transforms as a three-dimensional permutation representation under $\mathbb{Z}_3$.

I. When $\beta = 3$, with the ρ TDL action R_{ab} given by (5.122) and (5.128), one can check that the Cardy states furnish a NIM-rep

$$\mathcal{N}_\alpha = \begin{pmatrix} & 1 & & & & \\ & & 1 & & & \\ 1 & & & & & \\ & & & & 1 & \\ & & & & & 1 \\ & & & 1 & & \end{pmatrix}, \quad \mathcal{N}_\rho = \begin{pmatrix} 1 & 1 & 1 & 1 & & \\ 1 & 1 & 1 & & 1 & \\ 1 & 1 & 1 & & & 1 \\ 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \end{pmatrix}. \tag{5.159}$$

II. For $\beta = \frac{1}{\sqrt{3}}$, without assuming anything about the matrix $R_{ab} = x_{ab} + iy_{ab}$, we get the following representation for ρ action:

$$\mathcal{N}_\rho = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad B = \begin{pmatrix} \frac{1}{\sqrt{3}} + \frac{2x_{12}}{3} & -\frac{x_{12}}{3} + \frac{1+y_{12}}{\sqrt{3}} & -\frac{x_{12}}{3} + \frac{1-y_{12}}{\sqrt{3}} \\ -\frac{x_{12}}{3} + \frac{1+y_{12}}{\sqrt{3}} & -\frac{x_{12}}{3} + \frac{1-y_{12}}{\sqrt{3}} & \frac{1}{\sqrt{3}} + \frac{2x_{12}}{3} \\ -\frac{x_{12}}{3} + \frac{1-y_{12}}{\sqrt{3}} & \frac{1}{\sqrt{3}} + \frac{2x_{12}}{3} & -\frac{x_{12}}{3} + \frac{1+y_{12}}{\sqrt{3}} \end{pmatrix}, \tag{5.160}$$

and A, C, D are other 3×3 matrices whose explicit form we do not need. Suppose B is NIM. Because $B_{12} - B_{13} = \frac{2y_{12}}{\sqrt{3}}$, it follows that y_{12} must be a multiple of $\frac{\sqrt{3}}{2}$, and we can write $y_{12} = \frac{n\sqrt{3}}{2}$ with $n \in \mathbb{Z}$. But then $B_{11} + 2B_{12} = \sqrt{3} + n$. Hence no NIM-rep exists.

The results of the above analysis are summarized in Table 5.3. The defect TFT constructed in Section 5.6 passed the NIM-rep test, and notably the NIM-rep requirement in case II(b) allowed us to determine the action of the ρ TDL on the $\mathbb{Z}_3$-charged operators.

	(a) $n_{\mathrm{V}} = 2$	(b) $n_{\mathrm{V}} = 4$	(c) $n_{\mathrm{V}} = 6$
I. $\beta = 3$	$\circ$	$\times$	$\circ$
II. $\beta = \frac{1}{\sqrt{3}}$	$\times$	$\circ$	$\times$

Table 5.3: Existence of (1+1)d topological field theories realizing the Haagerup $\mathcal{H}_3$ fusion category from analyzing the fusion of topological defect lines with the admissible boundary conditions. We restrict to theories with exactly two $\mathbb{Z}_3$-neutral vacua 1 and v. Here n_{V} denotes the total number of vacua, and β is the coefficient in the fusion rule $v \times v = 1 + \beta v$. The $\circ$ marks the cases that pass the NIM-rep condition, and the $\times$ marks those ruled out. The theory constructed in Section 5.6 is highlighted.

5.8 Realization of Haagerup $\mathcal{H}_1$ and $\mathcal{H}_2$ via gauging

Given (a (1+1)d quantum field theory with) a finite symmetry group G that contains a non-anomalous subgroup H, gauging $H < G$ gives rise to (a quantum field theory with) a fusion category symmetry F' that contains a $\mathrm{R}ep(H)$ sub-category. This process can be reversed by gauging $\mathrm{R}ep(H) < F'$. In this sense, the pairs (G, H) and $(F, \mathrm{R}ep(H))$ are dual to each other. A generalization of the above statement is the following: given a fusion category $\mathcal{C}$ that contains an algebra object (a non-simple topological defect line satisfying certain conditions) A, gauging $A < \mathcal{C}$ gives rise to a fusion category $\mathcal{C}' = \mathrm{B}imod_{\mathcal{C}}(A, A)$ (category of (A, A) bimodules within $\mathcal{C}$) that contains a dual algebra object A', and this process can be reversed by gauging $A' < \mathcal{C}'$. Thus, the pairs $(\mathcal{C}, A)$ and $(\mathcal{C}', A')$ are dual to each other.[17] The reader is referred to [36] for a much more refined discussion, and to [47, 81, 83] for the original idea of generalized gauging.

The relations among the Haagerup $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ fusion categories can be understood this way. Up to automorphism, there are two nontrivial algebra objects in $\mathcal{H}_2$, one corresponding to the non-anomalous $\mathbb{Z}_3$ symmetry $\mathcal{I} + \alpha + \alpha^2$, and the other to $\mathcal{I} + \rho$. There are also two nontrivial algebra objects in $\mathcal{H}_3$, one again corresponding to the non-anomalous $\mathbb{Z}_3$ symmetry $\mathcal{I} + \alpha + \alpha^2$, and the other to $\mathcal{I} + \rho + \alpha\rho$. Gauging the $\mathbb{Z}_3$ symmetry exchanges $\mathcal{H}_2$ and $\mathcal{H}_3$, and gauging the other nontrivial algebra object in either $\mathcal{H}_2$ or $\mathcal{H}_3$ gives $\mathcal{H}_1$. These relations are summarized in Figure 5.1.

Thus, to construct topological field theories realizing the Haagerup $\mathcal{H}_1$ or $\mathcal{H}_2$ fusion

[17] Gauging by different algebra objects A_1 and A_2 with the same module category $\mathrm{M}od_{\mathcal{C}}(A_1) = \mathrm{M}od_{\mathcal{C}}(A_2)$ gives rise to the same gauged theory, so A_1 and A_2 are equivalent in the context of gauging. The duality pairing of $(\mathcal{C}, A)$ and $(\mathcal{C}', A')$ is up to this equivalence.

category, one can simply take a topological field theory realizing $\mathcal{H}_3$, such as the one we constructed in Section 5.6, and gauge $\mathcal{I}+\rho+\alpha\rho$ or $\mathcal{I}+\alpha+\alpha^2$ (the $\mathbb{Z}_3$ symmetry), respectively. A discussion on the gauging of algebra objects in (1+1)d topological field theory can be found in [36]. In particular, gauging the theory we constructed, which has $n_{\mathrm{V}}=6$ vacua and realizes the Haagerup $\mathcal{H}_3$ fusion category, by $\mathbb{Z}_3$ gives rise to a theory that has $n_{\mathrm{V}}=2$ vacua and realizes the Haagerup $\mathcal{H}_2$ fusion category.

$$\begin{tikzcd}
 & \mathcal{H}_1 & \\
\mathcal{H}_2 \arrow[ur, "\text{gauge } \mathcal{I}+\rho"] \arrow[rr, shift left] & & \mathcal{H}_3 \arrow[ul, "\text{gauge } \mathcal{I}+\rho+\alpha\rho"'] \arrow[ll, shift left, "\substack{\text{gauge } \mathbb{Z}_3 \\ (\text{gauge } \mathcal{I}+\alpha+\alpha^2)}"]
\end{tikzcd}$$

Figure 5.1: Gauging relations among theories realizing the three Haagerup fusion categories.

5.9 Prospective questions

- What is the full axiomatic data, when boundaries are included, of the defect topological field theory that we constructed in Section 5.6?

- Is there an explanation for the "superselection" rule noted below (5.141)?

- The construction of topological field theories realizing cases I(a), with $n_{\mathrm{V}}=2$ vacua and $\beta=\frac{1}{\sqrt{3}}$, and II(b), with $n_{\mathrm{V}}=4$ vacua and $\beta=\frac{1}{\sqrt{3}}$, is left for future work. For these cases, we showed in Section 5.7 that the Cardy states obtained from bootstrap furnish non-negative integer matrix representations under fusion with topological defect lines.

- Is there a conformal field theory realizing Haagerup or its quantum double? Despite nontrivial positive evidence from the work of Evans and Gannon [75], and recent attempts by Wolf [176], the question remains open. The defect modular bootstrap approach of [121, 122] may put universal constraints on such conformal field theories.

- Is Haagerup truly *exotic* (whatever exotic means)? Evans and Gannon [75] suggested not, as it sits inside a hypothetically infinite family of Haagerup-Izumi subfactors/fusion categories [105]. The transparent F-symbols for some higher members of this family have been recently computed by the present authors [102], and may allow for the construction of the corresponding defect topological field theories.

- Finally, the broader questions **Q1**, **Q2**, and **Q4** of Section 5.1 motivating this work remain open.

Acknowledgements

We thank Yuji Tachikawa for suggesting this interesting project, for helpful advice, and for comments on the first draft. We also benefited from enlightening discussions with Chi-Ming Chang, Kantaro Ohmori, Sahand Seifnashri, Shu-Heng Shao, and Yifan Wang. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632. YL is supported by the Simons Collaboration Grant on the Non-Perturbative Bootstrap and by the Sherman Fairchild Foundation.

BIBLIOGRAPHY

[1] Gerard 't Hooft et al., eds. *Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979.* Vol. 59. 1980, pp.1–438. DOI: 10.1007/978-1-4684-7571-5.

[2] David Aasen, Paul Fendley, and Roger S. K. Mong. "Topological Defects on the Lattice: Dualities and Degeneracies". In: (Aug. 2020). arXiv: 2008.08598 [cond-mat.stat-mech].

[3] Leonard S. Abrams. "Two-dimensional topological quantum field theories and Frobenius algebras". In: *J. Knot Theor. Ramifications* 5 (1996), pp. 569–587. DOI: 10.1142/S0218216596000333.

[4] Allan Adams et al. "Causality, analyticity and an IR obstruction to UV completion". In: *JHEP* 10 (2006), p. 014. DOI: 10.1088/1126-6708/2006/10/014. arXiv: hep-th/0602178.

[5] Y. Aharonov, A. Komar, and Leonard Susskind. "Superluminal behavior, causality, and instability". In: *Phys. Rev.* 182 (1969), pp. 1400–1403. DOI: 10.1103/PhysRev.182.1400.

[6] A. Alexeevski and S. Natanzon. "Noncommutative extensions of two-dimensional topological field theories and Hurwitz numbers for real algebraic curves". In: (Feb. 2002). arXiv: math/0202164.

[7] Nima Arkani-Hamed, Freddy Cachazo, and Jared Kaplan. "What is the Simplest Quantum Field Theory?" In: *JHEP* 09 (2010), p. 016. DOI: 10.1007/JHEP09(2010)016. arXiv: 0808.1446 [hep-th].

[8] Nima Arkani-Hamed, Yu-Tin Huang, and Shu-Heng Shao. "On the Positive Geometry of Conformal Field Theory". In: *JHEP* 06 (2019), p. 124. DOI: 10.1007/JHEP06(2019)124. arXiv: 1812.07739 [hep-th].

[9] Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-Tin Huang. "The EFT-Hedron". In: *JHEP* 05 (2021), p. 259. DOI: 10.1007/JHEP05(2021)259. arXiv: 2012.15849 [hep-th].

[10] Nima Arkani-Hamed, Tzu-Chen Huang, and Yu-tin Huang. "Scattering amplitudes for all masses and spins". In: *JHEP* 11 (2021), p. 070. DOI: 10.1007/JHEP11(2021)070. arXiv: 1709.04891 [hep-th].

[11] Nima Arkani-Hamed and Jaroslav Trnka. "The Amplituhedron". In: *JHEP* 10 (2014), p. 030. DOI: 10.1007/JHEP10(2014)030. arXiv: 1312.2007 [hep-th].

[12] Nima Arkani-Hamed et al. "Celestial amplitudes from UV to IR". In: *JHEP* 08 (2021), p. 062. DOI: 10.1007/JHEP08(2021)062. arXiv: 2012.04208 [hep-th].

[13] Nima Arkani-Hamed et al. *Grassmannian Geometry of Scattering Amplitudes*. Cambridge University Press, Apr. 2016. ISBN: 978-1-107-08658-6, 978-1-316-57296-2. DOI: 10.1017/CBO9781316091548. arXiv: 1212.5605 [hep-th].

[14] Marta Asaeda and Uffe Haagerup. "Exotic subfactors of finite depth with Jones indices $(5+\sqrt{13})/2$ and $(5+\sqrt{17})/2$". In: *Communications in mathematical physics* 202.1 (1999), pp. 1–63.

[15] Michael Atiyah. "Topological quantum field theories". In: *Publications Mathématiques de l'Institut des Hautes Études Scientifiques* 68.1 (Jan. 1988), pp. 175–186.

[16] S. D. Badger. "Direct Extraction Of One Loop Rational Terms". In: *JHEP* 01 (2009), p. 049. DOI: 10.1088/1126-6708/2009/01/049. arXiv: 0806.4600 [hep-ph].

[17] S. D. Badger, E. W. Nigel Glover, and Valentin V. Khoze. "Recursion relations for gauge theory amplitudes with massive vector bosons and fermions". In: *JHEP* 01 (2006), p. 066. DOI: 10.1088/1126-6708/2006/01/066. arXiv: hep-th/0507161.

[18] S. D. Badger et al. "Recursion relations for gauge theory amplitudes with massive particles". In: *JHEP* 07 (2005), p. 025. DOI: 10.1088/1126-6708/2005/07/025. arXiv: hep-th/0504159.

[19] Simon Badger et al. "A unitarity compatible approach to one-loop amplitudes with massive fermions". In: *JHEP* 06 (2017), p. 141. DOI: 10.1007/JHEP06(2017)141. arXiv: 1703.05734 [hep-ph].

[20] V. Bargmann and Eugene P. Wigner. "Group Theoretical Discussion of Relativistic Wave Equations". In: *Proc. Nat. Acad. Sci.* 34 (1948), p. 211. DOI: 10.1073/pnas.34.5.211.

[21] John W. Barrett and Bruce W. Westbury. "Invariants of piecewise linear three manifolds". In: *Trans. Am. Math. Soc.* 348 (1996), pp. 3997–4022. DOI: 10.1090/S0002-9947-96-01660-1. arXiv: hep-th/9311155.

[22] Brando Bellazzini. "Softness and amplitudes' positivity for spinning particles". In: *JHEP* 02 (2017), p. 034. DOI: 10.1007/JHEP02(2017)034. arXiv: 1605.06111 [hep-th].

[23] Brando Bellazzini et al. "Beyond Positivity Bounds and the Fate of Massive Gravity". In: *Phys. Rev. Lett.* 120.16 (2018), p. 161101. DOI: 10.1103/PhysRevLett.120.161101. arXiv: 1710.02539 [hep-th].

[24] Brando Bellazzini et al. "Positive moments for scattering amplitudes". In: *Phys. Rev. D* 104.3 (2021), p. 036006. DOI: 10.1103/PhysRevD.104.036006. arXiv: 2011.00037 [hep-th].

[25] Paolo Benincasa and Freddy Cachazo. "Consistency Conditions on the S-Matrix of Massless Particles". In: (May 2007). arXiv: 0705.4305 [hep-th].

[26] Nathan Benjamin et al. "Light-cone modular bootstrap and pure gravity". In: *Phys. Rev. D* 100.6 (2019), p. 066029. DOI: 10.1103/PhysRevD.100.066029. arXiv: 1906.04184 [hep-th].

[27] Nathan Benjamin et al. "Twist gap and global symmetry in two dimensions". In: *Phys. Rev. D* 101.10 (2020), p. 106026. DOI: 10.1103/PhysRevD.101.106026. arXiv: 2003.02844 [hep-th].

[28] Frits A. Berends et al. "Single Bremsstrahlung Processes in Gauge Theories". In: *Phys. Lett. B* 103 (1981), pp. 124–128. DOI: 10.1016/0370-2693(81)90685-7.

[29] Z. Bern and A. G. Morgan. "Massive loop amplitudes from unitarity". In: *Nucl. Phys. B* 467 (1996), pp. 479–509. DOI: 10.1016/0550-3213(96)00078-8. arXiv: hep-ph/9511336.

[30] Zvi Bern, Lance J. Dixon, and David A. Kosower. "Progress in one loop QCD computations". In: *Ann. Rev. Nucl. Part. Sci.* 46 (1996), pp. 109–148. DOI: 10.1146/annurev.nucl.46.1.109. arXiv: hep-ph/9602280.

[31] Zvi Bern, Lance J. Dixon, and David A. Kosower. "Two-loop g —> gg splitting amplitudes in QCD". In: *JHEP* 08 (2004), p. 012. DOI: 10.1088/1126-6708/2004/08/012. arXiv: hep-ph/0404293.

[32] Zvi Bern et al. "Fusing gauge theory tree amplitudes into loop amplitudes". In: *Nucl. Phys. B* 435 (1995), pp. 59–101. DOI: 10.1016/0550-3213(94)00488-Z. arXiv: hep-ph/9409265.

[33] Zvi Bern et al. "Generalized Unitarity and Six-Dimensional Helicity". In: *Phys. Rev. D* 83 (2011), p. 085022. DOI: 10.1103/PhysRevD.83.085022. arXiv: 1010.0494 [hep-th].

[34] Zvi Bern et al. "One loop n point gauge theory amplitudes, unitarity and collinear limits". In: *Nucl. Phys. B* 425 (1994), pp. 217–260. DOI: 10.1016/0550-3213(94)90179-1. arXiv: hep-ph/9403226.

[35] Zvi Bern et al. "One loop selfdual and N=4 superYang-Mills". In: *Phys. Lett. B* 394 (1997), pp. 105–115. DOI: 10.1016/S0370-2693(96)01676-0. arXiv: hep-th/9611127.

[36] Lakshya Bhardwaj and Yuji Tachikawa. "On finite symmetries and their gauging in two dimensions". In: *JHEP* 03 (2018), p. 189. DOI: 10.1007/JHEP03(2018)189. arXiv: 1704.02330 [hep-th].

[37] Rutger H. Boels and Tobias Hansen. "String theory in target space". In: *JHEP* 06 (2014), p. 054. DOI: 10.1007/JHEP06(2014)054. arXiv: 1402.6356 [hep-th].

[38] Wieb Bosma, John Cannon, and Catherine Playoust. "The Magma algebra system. I. The user language". In: *J. Symbolic Comput.* 24.3-4 (1997). Computational algebra and number theory (London, 1993), pp. 235–265. ISSN: 0747-7171. DOI: 10.1006/jsco.1996.0125. URL: http://dx.doi.org/10.1006/jsco.1996.0125.

[39] Ruth Britto, Freddy Cachazo, and Bo Feng. "Generalized unitarity and one-loop amplitudes in N=4 super-Yang-Mills". In: *Nucl. Phys. B* 725 (2005), pp. 275–305. DOI: 10.1016/j.nuclphysb.2005.07.014. arXiv: hep-th/0412103.

[40] Ruth Britto, Freddy Cachazo, and Bo Feng. "New recursion relations for tree amplitudes of gluons". In: *Nucl. Phys. B* 715 (2005), pp. 499–522. DOI: 10.1016/j.nuclphysb.2005.02.030. arXiv: hep-th/0412308.

[41] Ruth Britto and Edoardo Mirabella. "External leg corrections in the unitarity method". In: *JHEP* 01 (2012), p. 045. DOI: 10.1007/JHEP01(2012)045. arXiv: 1109.5106 [hep-ph].

[42] Ruth Britto et al. "Direct proof of tree-level recursion relation in Yang-Mills theory". In: *Phys. Rev. Lett.* 94 (2005), p. 181602. DOI: 10.1103/PhysRevLett.94.181602. arXiv: hep-th/0501052.

[43] Paul Bruillard. "Rank 4 premodular categories". In: *New York J. Math* 22 (2016), pp. 775–800.

[44] Freddy Cachazo, Peter Svrcek, and Edward Witten. "MHV vertices and tree amplitudes in gauge theory". In: *JHEP* 09 (2004), p. 006. DOI: 10.1088/1126-6708/2004/09/006. arXiv: hep-th/0403047.

[45] Xian O. Camanho et al. "Causality Constraints on Corrections to the Graviton Three-Point Coupling". In: *JHEP* 02 (2016), p. 020. DOI: 10.1007/JHEP02(2016)020. arXiv: 1407.5597 [hep-th].

[46] Simon Caron-Huot and Vincent Van Duong. "Extremal Effective Field Theories". In: *JHEP* 05 (2021), p. 280. DOI: 10.1007/JHEP05(2021)280. arXiv: 2011.02957 [hep-th].

[47] Nils Carqueville and Ingo Runkel. "Orbifold completion of defect bicategories". In: *Quantum Topol.* 7 (2016), p. 203. DOI: 10.4171/QT/76. arXiv: 1210.6363 [math.QA].

[48] Gordon Chalmers and Warren Siegel. "Simplifying algebra in Feynman graphs. 3. Massive vectors". In: *Phys. Rev. D* 63 (2001), p. 125027. DOI: 10.1103/PhysRevD.63.125027. arXiv: hep-th/0101025.

[49] Chi-Ming Chang and Ying-Hsuan Lin. "Lorentzian Dynamics and Factorization Beyond Rationality". In: (Dec. 2020). arXiv: 2012.01429 [hep-th].

[50] Chi-Ming Chang et al. "Topological Defect Lines and Renormalization Group Flows in Two Dimensions". In: *JHEP* 01 (2019), p. 026. DOI: 10.1007/JHEP01(2019)026. arXiv: 1802.04445 [hep-th].

[51] Xie Chen, Zheng-Xin Liu, and Xiao-Gang Wen. "Two-dimensional symmetry-protected topological orders and their protected gapless edge excitations". In: *Phys. Rev. B* 84.23 (2011), p. 235141. DOI: 10.1103/PhysRevB.84.235141. arXiv: 1106.4752 [cond-mat.str-el].

[52] Clifford Cheung and Donal O'Connell. "Amplitudes and Spinor-Helicity in Six Dimensions". In: *JHEP* 07 (2009), p. 075. DOI: 10.1088/1126-6708/2009/07/075. arXiv: 0902.0981 [hep-th].

[53] Sidney R. Coleman and J. Mandula. "All Possible Symmetries of the S Matrix". In: *Phys. Rev.* 159 (1967). Ed. by A. Zichichi, pp. 1251–1256. DOI: 10.1103/PhysRev.159.1251.

[54] Scott Collier, Ying-Hsuan Lin, and Xi Yin. "Modular Bootstrap Revisited". In: *JHEP* 09 (2018), p. 061. DOI: 10.1007/JHEP09(2018)061. arXiv: 1608.06241 [hep-th].

[55] Eduardo Conde, Euihun Joung, and Karapet Mkrtchyan. "Spinor-Helicity Three-Point Amplitudes from Local Cubic Interactions". In: *JHEP* 08 (2016), p. 040. DOI: 10.1007/JHEP08(2016)040. arXiv: 1605.07402 [hep-th].

[56] Eduardo Conde and Andrea Marzolla. "Lorentz Constraints on Massive Three-Point Amplitudes". In: *JHEP* 09 (2016), p. 041. DOI: 10.1007/JHEP09(2016)041. arXiv: 1601.08113 [hep-th].

[57] Nathaniel Craig et al. "Massive amplitudes on the Coulomb branch of N=4 SYM". In: *JHEP* 12 (2011), p. 097. DOI: 10.1007/JHEP12(2011)097. arXiv: 1104.2050 [hep-th].

[58] Alexei Davydov, Liang Kong, and Ingo Runkel. "Field theories with defects and the centre functor". In: (July 2011). Ed. by Hisham Sati and Urs Schreiber, pp. 71–128. arXiv: 1107.0495 [math.QA].

[59] Andrei I. Davydychev. "Standard and hypergeometric representations for loop diagrams and the photon-photon scattering". In: *7th International Seminar on High-energy Physics*. May 1993. arXiv: hep-ph/9307323.

[60] P. De Causmaecker et al. "Multiple Bremsstrahlung in Gauge Theories at High-Energies. 1. General Formalism for Quantum Electrodynamics". In: *Nucl. Phys. B* 206 (1982), pp. 53–60. DOI: 10.1016/0550-3213(82)90488-6.

[61] Robertus Henricus Dijkgraaf. "A geometrical approach to two-dimensional conformal field theory". In: *PhDT* (1989).

[62] Lance J. Dixon. "Calculating scattering amplitudes efficiently". In: *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 95): QCD and Beyond*. Jan. 1996, pp. 539–584. arXiv: hep-ph/9601359.

[63] Lance J. Dixon et al. "Strings on Orbifolds". In: *Nucl. Phys. B* 261 (1985). Ed. by B. Schellekens, pp. 678–686. DOI: 10.1016/0550-3213(85)90593-0.

[64] Lance J. Dixon et al. "Strings on Orbifolds. 2." In: *Nucl. Phys. B* 274 (1986), pp. 285–314. DOI: 10.1016/0550-3213(86)90287-7.

[65] Chongying Dong and Geoffrey Mason. "Coset constructions and dual pairs for vertex operator algebras". In: *arXiv preprint math/9904155* (1999).

[66] Vladimir Dotsenko et al. "Coupled Potts models: Self-duality and fixed point structure". In: *Nucl. Phys. B* 546 (1999), pp. 505–557. DOI: 10.1016/S0550-3213(99)00097-8. arXiv: cond-mat/9812227.

[67] Nadav Drukker, Davide Gaiotto, and Jaume Gomis. "The Virtue of Defects in 4D Gauge Theories and 2D CFTs". In: *JHEP* 06 (2011), p. 025. DOI: 10.1007/JHEP06(2011)025. arXiv: 1003.1112 [hep-th].

[68] R. Keith Ellis et al. "Masses, fermions and generalized D-dimensional unitarity". In: *Nucl. Phys. B* 822 (2009), pp. 270–282. DOI: 10.1016/j.nuclphysb.2009.07.023. arXiv: 0806.3467 [hep-ph].

[69] Henriette Elvang and Yu-tin Huang. "Scattering Amplitudes". In: (Aug. 2013). arXiv: 1308.1697 [hep-th].

[70] Henriette Elvang et al. "On renormalization group flows and the a-theorem in 6d". In: *JHEP* 10 (2012), p. 011. DOI: 10.1007/JHEP10(2012)011. arXiv: 1205.3994 [hep-th].

[71] Oluf Tang Engelund and Radu Roiban. "Correlation functions of local composite operators from generalized unitarity". In: *JHEP* 03 (2013), p. 172. DOI: 10.1007/JHEP03(2013)172. arXiv: 1209.0227 [hep-th].

[72] Pavel Etingof, Dmitri Nikshych, and Viktor Ostrik. "On fusion categories". In: (). eprint: math/0203060. URL: https://arxiv.org/abs/math/0203060.

[73] Pavel Etingof et al. *Tensor categories*. Vol. 205. American Mathematical Soc., 2016.

[74] David E. Evans and Terry Gannon. "Non-unitary fusion categories and their doubles via endomorphisms". In: *Adv. Math.* 310 (2017), pp. 1–43. DOI: 10.1016/j.aim.2017.01.015. arXiv: 1506.03546 [math.QA].

[75] David E. Evans and Terry Gannon. "The exoticness and realisability of twisted Haagerup-Izumi modular data". In: *Commun. Math. Phys.* 307 (2011), pp. 463–512. DOI: 10.1007/s00220-011-1329-3. arXiv: 1006.1326 [math.QA].

[76] Raffaele A. Fazio et al. "On the Four-Dimensional Formulation of Dimensionally Regulated Amplitudes". In: *Eur. Phys. J. C* 74.12 (2014), p. 3197. DOI: 10.1140/epjc/s10052-014-3197-4. arXiv: 1404.4783 [hep-ph].

[77] Adrian Feiguin et al. "Interacting anyons in topological quantum liquids: The golden chain". In: *Phys. Rev. Lett.* 98.16 (2007), p. 160409. DOI: 10.1103/PhysRevLett.98.160409. arXiv: cond-mat/0612341.

[78] Darren Forde. "Direct extraction of one-loop integral coefficients". In: *Phys. Rev. D* 75 (2007), p. 125019. DOI: 10.1103/PhysRevD.75.125019. arXiv: 0704.1835 [hep-ph].

[79] A. Fotopoulos and Mirian Tsulaia. “On the Tensionless Limit of String theory, Off - Shell Higher Spin Interaction Vertices and BCFW Recursion Relations”. In: *JHEP* 11 (2010), p. 086. DOI: 10.1007/JHEP11(2010)086. arXiv: 1009.0727 [hep-th].

[80] Daniel Friedan, Zong-an Qiu, and Stephen H. Shenker. “Conformal Invariance, Unitarity and Two-Dimensional Critical Exponents”. In: *Phys. Rev. Lett.* 52 (1984), pp. 1575–1578. DOI: 10.1103/PhysRevLett.52.1575.

[81] Jurg Frohlich et al. “Defect lines, dualities, and generalised orbifolds”. In: *16th International Congress on Mathematical Physics*. Sept. 2009. DOI: 10.1142/9789814304634_0056. arXiv: 0909.5013 [math-ph].

[82] Marcel Froissart. “Asymptotic behavior and subtractions in the Mandelstam representation”. In: *Phys. Rev.* 123 (1961), pp. 1053–1057. DOI: 10.1103/PhysRev.123.1053.

[83] Jurgen Fuchs, Ingo Runkel, and Christoph Schweigert. “TFT construction of RCFT correlators 1. Partition functions”. In: *Nucl. Phys. B* 646 (2002), pp. 353–497. DOI: 10.1016/S0550-3213(02)00744-7. arXiv: hep-th/0204148.

[84] Matthias R. Gaberdiel, Harsha R. Hampapura, and Sunil Mukhi. “Cosets of Meromorphic CFTs and Modular Differential Equations”. In: *JHEP* 04 (2016), p. 156. DOI: 10.1007/JHEP04(2016)156. arXiv: 1602.01022 [hep-th].

[85] Davide Gaiotto. “Open Verlinde line operators”. In: (Apr. 2014). arXiv: 1404.0332 [hep-th].

[86] Davide Gaiotto et al. “Generalized Global Symmetries”. In: *JHEP* 02 (2015), p. 172. DOI: 10.1007/JHEP02(2015)172. arXiv: 1412.5148 [hep-th].

[87] S Gelfand and D Kazhdan. “Invariants of three-dimensional manifolds”. In: *Geometric & Functional Analysis GAFA* 6.2 (1996), pp. 268–300.

[88] P. Goddard, A. Kent, and David I. Olive. “Unitary Representations of the Virasoro and Supervirasoro Algebras”. In: *Commun. Math. Phys.* 103 (1986), pp. 105–119. DOI: 10.1007/BF01464283.

[89] Michael B. Green and Congkao Wen. “Superstring amplitudes, unitarily, and Hankel determinants of multiple zeta values”. In: *JHEP* 11 (2019), p. 079. DOI: 10.1007/JHEP11(2019)079. arXiv: 1908.08426 [hep-th].

[90] Pinhas Grossman and Masaki Izumi. “Drinfeld centers of fusion categories arising from generalized Haagerup subfactors”. In: *arXiv preprint arXiv:1501.07679* (2015).

[91] Pinhas Grossman and Masaki Izumi. “Infinite families of potential modular data related to quadratic categories”. In: *arXiv preprint arXiv:1906.07397* (2019).

[92] Pinhas Grossman, Masaki Izumi, and Noah Snyder. “The Asaeda–Haagerup fusion categories”. In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 2018.743 (2018), pp. 261–305.

[93] Pinhas Grossman and Noah Snyder. "Quantum Subgroups of the Haagerup Fusion Categories". In: *Communications in Mathematical Physics* 311.3 (Mar. 2012), pp. 617–643. DOI: 10.1007/s00220-012-1427-x. URL: https://doi.org/10.1007%2Fs00220-012-1427-x.

[94] Alfredo Guevara, Alexander Ochirov, and Justin Vines. "Scattering of Spinning Black Holes from Exponentiated Soft Factors". In: *JHEP* 09 (2019), p. 056. DOI: 10.1007/JHEP09(2019)056. arXiv: 1812.06895 [hep-th].

[95] J. F. Gunion and Z. Kunszt. "Improved Analytic Techniques for Tree Graph Calculations and the G g q anti-q Lepton anti-Lepton Subprocess". In: *Phys. Lett. B* 161 (1985), p. 333. DOI: 10.1016/0370-2693(85)90774-9.

[96] Uffe Haagerup. "Principal graphs of subfactors in the index range $4 < [M : N] < 3 + \sqrt{2}$". In: *Subfactors (Kyuzeso, 1993)* (1994), pp. 1–38.

[97] Kaoru Hagiwara and D. Zeppenfeld. "Helicity Amplitudes for Heavy Lepton Production in e+ e- Annihilation". In: *Nucl. Phys. B* 274 (1986), pp. 1–32. DOI: 10.1016/0550-3213(86)90615-2.

[98] W. Heisenberg. "Production of Meson Showers". In: *Nature* 164 (1949), pp. 65–66. DOI: 10.1038/164065c0.

[99] Johannes M. Henn and Jan C. Plefka. *Scattering Amplitudes in Gauge Theories*. Vol. 883. Berlin: Springer, 2014. ISBN: 978-3-642-54021-9. DOI: 10.1007/978-3-642-54022-6.

[100] Yu-tin Huang and David McGady. "Consistency Conditions for Gauge Theory S Matrices from Requirements of Generalized Unitarity". In: *Phys. Rev. Lett.* 112.24 (2014), p. 241601. DOI: 10.1103/PhysRevLett.112.241601. arXiv: 1307.4065 [hep-th].

[101] Yu-tin Huang et al. "Carving out the Space of Open-String S-matrix". In: *JHEP* 04 (2021), p. 195. DOI: 10.1007/JHEP04(2021)195. arXiv: 2008.02293 [hep-th].

[102] Tzu-Chen Huang and Ying-Hsuan Lin. "The F-Symbols for Transparent Haagerup-Izumi Categories with $G = \mathbb{Z}_{2n+1}$". In: (July 2020). arXiv: 2007.00670 [math.CT].

[103] Tzu-Chen Huang and Ying-Hsuan Lin. "Topological Field Theory with Haagerup Symmetry". In: (Feb. 2021). DOI: 10.1063/5.0079062. arXiv: 2102.05664 [hep-th].

[104] Masaki Izumi. "The classification of 3^n subfactors and related fusion categories". In: *Quantum Topology* 9.3 (July 2018), pp. 473–562. ISSN: 1663-487X. DOI: 10.4171/qt/113. URL: http://dx.doi.org/10.4171/qt/113.

[105] Masaki Izumi. "The structure of sectors associated with Longo–Rehren inclusions II: examples". In: *Reviews in Mathematical Physics* 13.05 (2001), pp. 603–674.

[106] Henrik Johansson and Alexander Ochirov. "Color-Kinematics Duality for QCD Amplitudes". In: *JHEP* 01 (2016), p. 170. DOI: 10.1007/JHEP01(2016)170. arXiv: 1507.00332 [hep-ph].

[107] V.F.R. Jones. "Index for subfactors". In: *Invent. Math.* 72 (1983), pp. 1–25. DOI: 10.1007/BF01389127.

[108] Vaughan Jones and Viakalathur Shankar Sunder. *Introduction to subfactors*. Vol. 234. Cambridge University Press, 1997.

[109] Vaughan FR Jones. *von Neumann algebras in mathematics and physics*. American Mathematical Society, 1990.

[110] Anton Kapustin and Ryan Thorngren. "Anomalies of discrete symmetries in various dimensions and group cohomology". In: (Apr. 2014). arXiv: 1404.3230 [hep-th].

[111] Samuel Karlin and James McGregor. "Coincidence properties of birth and death processes." In: *Pacific Journal of Mathematics* 9.4 (1959), pp. 1109–1140. DOI: pjm/1103038888. URL: https://doi.org/.

[112] R. Kleiss and W. James Stirling. "Spinor Techniques for Calculating p anti-p —> W+- / Z0 + Jets". In: *Nucl. Phys. B* 262 (1985), pp. 235–262. DOI: 10.1016/0550-3213(85)90285-8.

[113] R. Kleiss and W. James Stirling. "TOP QUARK PRODUCTION AT HADRON COLLIDERS: SOME USEFUL FORMULAE". In: *Z. Phys. C* 40 (1988), pp. 419–423. DOI: 10.1007/BF01548856.

[114] Joachim Kock. *Frobenius algebras and 2-d topological quantum field theories*. Vol. 59. Cambridge University Press, 2004.

[115] Zohar Komargodski et al. "Symmetries and Strings of Adjoint QCD_2". In: (Aug. 2020). arXiv: 2008.07567 [hep-th].

[116] David A. Kosower. "Next-to-maximal helicity violating amplitudes in gauge theory". In: *Phys. Rev. D* 71 (2005), p. 045007. DOI: 10.1103/PhysRevD.71.045007. arXiv: hep-th/0406175.

[117] Aaron D. Lauda and Hendryk Pfeiffer. "Open-closed strings: Two-dimensional extended TQFTs and Frobenius algebras". In: (Oct. 2005). arXiv: math/0510664.

[118] C.I. Lazaroiu. "On the structure of open - closed topological field theory in two-dimensions". In: *Nucl. Phys. B* 603 (2001), pp. 497–530. DOI: 10.1016/S0550-3213(01)00135-3. arXiv: hep-th/0010269.

[119] Michael A. Levin and Xiao-Gang Wen. "String net condensation: A Physical mechanism for topological phases". In: *Phys. Rev. B* 71 (2005), p. 045110. DOI: 10.1103/PhysRevB.71.045110. arXiv: cond-mat/0404617.

[120] Haisheng Li. "On Abelian coset generalized vertex algebras". In: (Aug. 2000). arXiv: math/0008062.

[121] Ying-Hsuan Lin and Shu-Heng Shao. "$\mathbb{Z}_N$ Symmetries, Anomalies, and the Modular Bootstrap". In: (Jan. 2021). arXiv: 2101.08343 [hep-th].

[122] Ying-Hsuan Lin and Shu-Heng Shao. "Anomalies and Bounds on Charged Operators". In: *Phys. Rev. D* 100.2 (2019), p. 025013. DOI: 10.1103/PhysRevD.100.025013. arXiv: 1904.04833 [hep-th].

[123] F. E. Low. “Bremsstrahlung of very low-energy quanta in elementary particle collisions”. In: *Phys. Rev.* 110 (1958), pp. 974–977. DOI: 10.1103/PhysRev.110.974.

[124] Jacob Lurie. “On the Classification of Topological Field Theories”. In: (May 2009). arXiv: 0905.0465 [math.CT].

[125] Saunders MacLane. “Natural associativity and commutativity”. In: *Rice Institute Pamphlet-Rice University Studies* 49.4 (1963).

[126] A. Martin. “Unitarity and high-energy behavior of scattering amplitudes”. In: *Phys. Rev.* 129 (1963), pp. 1432–1436. DOI: 10.1103/PhysRev.129.1432.

[127] Andre Martin. “Extension of the axiomatic analyticity domain of scattering amplitudes by unitarity. 1.” In: *Nuovo Cim. A* 42 (1965), pp. 930–953. DOI: 10.1007/BF02720568.

[128] David A. McGady and Laurentiu Rodina. “Higher-spin massless S-matrices in four-dimensions”. In: *Phys. Rev. D* 90.8 (2014), p. 084048. DOI: 10.1103/PhysRevD.90.084048. arXiv: 1311.2938 [hep-th].

[129] Gregory W. Moore and Graeme Segal. “D-branes and K-theory in 2D topological field theory”. In: (Aug. 2006). arXiv: hep-th/0609042.

[130] Gregory W. Moore and Nathan Seiberg. “Classical and Quantum Conformal Field Theory”. In: *Commun. Math. Phys.* 123 (1989), p. 177. DOI: 10.1007/BF01238857.

[131] Gregory W. Moore and Nathan Seiberg. “LECTURES ON RCFT”. In: *1989 Banff NATO ASI: Physics, Geometry and Topology.* Sept. 1989, pp. 1–129.

[132] Gregory W. Moore and Nathan Seiberg. “Naturality in Conformal Field Theory”. In: *Nucl. Phys. B* 313 (1989), pp. 16–40. DOI: 10.1016/0550-3213(89)90511-7.

[133] Gregory W. Moore and Nathan Seiberg. “Polynomial Equations for Rational Conformal Field Theories”. In: *Phys. Lett. B* 212 (1988), pp. 451–460. DOI: 10.1016/0370-2693(88)91796-0.

[134] Gregory W. Moore and Nathan Seiberg. “Taming the Conformal Zoo”. In: *Phys. Lett. B* 220 (1989), pp. 422–430. DOI: 10.1016/0370-2693(89)90897-6.

[135] Michael Müger. “From subfactors to categories and topology I: Frobenius algebras in and Morita equivalence of tensor categories”. In: *Journal of Pure and Applied Algebra* 180.1-2 (2003), pp. 81–157.

[136] Yu Nakayama. “Perturbative search for dead-end CFTs”. In: *JHEP* 05 (2015), p. 046. DOI: 10.1007/JHEP05(2015)046. arXiv: 1501.02280 [hep-th].

[137] Alberto Nicolis, Riccardo Rattazzi, and Enrico Trincherini. “Energy’s and amplitudes’ positivity”. In: *JHEP* 05 (2010). [Erratum: JHEP 11, 128 (2011)], p. 095. DOI: 10.1007/JHEP05(2010)095. arXiv: 0912.4258 [hep-th].

[138] Adrian Ocneanu. “Quantized groups, string algebras and Galois theory for algebras”. In: *Operator algebras and applications* 2 (1988), pp. 119–172.

[139] Tobias J. Osborne, Deniz E. Stiegemann, and Ramona Wolf. "The F-Symbols for the $\mathcal{H}_3$ Fusion Category". In: (2019). arXiv: 1906.01322 [math.CT].

[140] Giovanni Ossola, Costas G. Papadopoulos, and Roberto Pittau. "On the Rational Terms of the one-loop amplitudes". In: *JHEP* 05 (2008), p. 004. DOI: 10.1088/1126-6708/2008/05/004. arXiv: 0802.1876 [hep-ph].

[141] K. J. Ozeren and W. J. Stirling. "Scattering amplitudes with massive fermions using BCFW recursion". In: *Eur. Phys. J. C* 48 (2006), pp. 159–168. DOI: 10.1140/epjc/s10052-006-0007-7. arXiv: hep-ph/0603071.

[142] V.B. Petkova and J.B. Zuber. "Generalized twisted partition functions". In: *Phys. Lett. B* 504 (2001), pp. 157–164. DOI: 10.1016/S0370-2693(01)00276-3. arXiv: hep-th/0011021.

[143] T. N. Pham and Tran N. Truong. "Evaluation of the Derivative Quartic Terms of the Meson Chiral Lagrangian From Forward Dispersion Relation". In: *Phys. Rev. D* 31 (1985), p. 3027. DOI: 10.1103/PhysRevD.31.3027.

[144] Eugeniu Plamadeala, Michael Mulligan, and Chetan Nayak. "Perfect Metal Phases of One-Dimensional and Anisotropic Higher-Dimensional Systems". In: *Phys. Rev. B* 90.24 (2014), p. 241101. DOI: 10.1103/PhysRevB.90.241101. arXiv: 1404.4367 [cond-mat.str-el].

[145] Sorin Popa et al. "Classification of amenable subfactors of type II". In: *Acta Mathematica* 172.2 (1994), pp. 163–255.

[146] Claudia de Rham et al. "Positivity bounds for scalar field theories". In: *Phys. Rev. D* 96.8 (2017), p. 081702. DOI: 10.1103/PhysRevD.96.081702. arXiv: 1702.06134 [hep-th].

[147] Claudia de Rham et al. "UV complete me: Positivity Bounds for Particles with Spin". In: *JHEP* 03 (2018), p. 011. DOI: 10.1007/JHEP03(2018)011. arXiv: 1706.02712 [hep-th].

[148] Ingo Runkel and Rafal R. Suszek. "Gerbe-holonomy for surfaces with defect networks". In: *Adv. Theor. Math. Phys.* 13.4 (2009), pp. 1137–1219. DOI: 10.4310/ATMP.2009.v13.n4.a5. arXiv: 0808.1419 [hep-th].

[149] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. "ORBITOPES". In: *Mathematika* 57.2 (June 2011), pp. 275–314. DOI: 10.1112/s002557931100132x. URL: https://doi.org/10.1112%2Fs002557931100132x.

[150] Stephen Sawin. "Direct sum decompositions and indecomposable TQFTs". In: *J. Math. Phys.* 36 (1995), pp. 6673–6680. DOI: 10.1063/1.531180. arXiv: q-alg/9505026.

[151] Albert S. Schwarz. "On Regular Solutions of Euclidean Yang-Mills Equations". In: *Phys. Lett. B* 67 (1977), pp. 172–174. DOI: 10.1016/0370-2693(77)90095-8.

[152] Christian Schwinn and Stefan Weinzierl. "Scalar diagrammatic rules for Born amplitudes in QCD". In: *JHEP* 05 (2005), p. 006. DOI: 10.1088/1126-6708/2005/05/006. arXiv: hep-th/0503015.

[153] Christian Schwinn and Stefan Weinzierl. "SUSY ward identities for multi-gluon helicity amplitudes with massive quarks". In: *JHEP* 03 (2006), p. 030. DOI: 10.1088/1126-6708/2006/03/030. arXiv: hep-th/0602012.

[154] David Simmons-Duffin. "The Conformal Bootstrap". In: *Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings*. 2017, pp. 1–74. DOI: 10.1142/9789813149441_0001. arXiv: 1602.07982 [hep-th].

[155] Aninda Sinha and Ahmadullah Zahed. "Crossing Symmetric Dispersion Relations in Quantum Field Theories". In: *Phys. Rev. Lett.* 126.18 (2021), p. 181601. DOI: 10.1103/PhysRevLett.126.181601. arXiv: 2012.04877 [hep-th].

[156] Hidenori Sonoda. "SEWING CONFORMAL FIELD THEORIES". In: *Nucl. Phys. B* 311 (1988), pp. 401–416. DOI: 10.1016/0550-3213(88)90066-1.

[157] Hidenori Sonoda. "SEWING CONFORMAL FIELD THEORIES. 2." In: *Nucl. Phys. B* 311 (1988), pp. 417–432. DOI: 10.1016/0550-3213(88)90067-3.

[158] "Summationsmethoden und Momentfolgen I." de. In: *Mathematische Zeitschrift* 10 (1921), pp. 74–109. URL: http://dml.mathdoc.fr/item/GDZPPN002365804.

[159] "Summationsmethoden und Momentfolgen II." de. In: *Mathematische Zeitschrift* 10 (1921), pp. 280–299. URL: http://dml.mathdoc.fr/item/GDZPPN002365944.

[160] Yuji Tachikawa. "On gauging finite subgroups". In: *SciPost Phys.* 8.1 (2020), p. 015. DOI: 10.21468/SciPostPhys.8.1.015. arXiv: 1712.09542 [hep-th].

[161] Ryan Thorngren and Yifan Wang. "Fusion Category Symmetry I: Anomaly In-Flow and Gapped Phases". In: (Dec. 2019). arXiv: 1912.02817 [hep-th].

[162] Josiah Thornton. "Generalized near-group categories". In: (2012).

[163] Matthew Titsworth. Private communication: to appear.

[164] Andrew J. Tolley, Zi-Yue Wang, and Shuang-Yong Zhou. "New positivity bounds from full crossing symmetry". In: *JHEP* 05 (2021), p. 255. DOI: 10.1007/JHEP05(2021)255. arXiv: 2011.02400 [hep-th].

[165] V.G. Turaev. *Quantum invariants of knots and three manifolds*. Vol. 18. 1994.

[166] V.G. Turaev and O.Y. Viro. "State sum invariants of 3 manifolds and quantum 6j symbols". In: *Topology* 31 (1992), pp. 865–902. DOI: 10.1016/0040-9383(92)90015-A.

[167] Erik P. Verlinde. "Fusion Rules and Modular Transformations in 2D Conformal Field Theory". In: *Nucl. Phys. B* 300 (1988), pp. 360–376. DOI: 10.1016/0550-3213(88)90603-7.

[168] Juven Wang, Xiao-Gang Wen, and Edward Witten. "Symmetric Gapped Interfaces of SPT and SET States: Systematic Constructions". In: *Phys. Rev. X* 8.3 (2018), p. 031048. DOI: 10.1103/PhysRevX.8.031048. arXiv: 1705.06728 [cond-mat.str-el].

[169] A Wassermann. "Quantum subgroups and vertex algebras". In: *Lectures given at MSRI in Dec* (2000).

[170] Steven Weinberg. "Infrared photons and gravitons". In: *Phys. Rev.* 140 (1965), B516–B524. DOI: 10.1103/PhysRev.140.B516.

[171] Steven Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, June 2005. ISBN: 978-0-521-67053-1, 978-0-511-25204-4.

[172] Steven Weinberg and Edward Witten. "Limits on Massless Particles". In: *Phys. Lett. B* 96 (1980), pp. 59–62. DOI: 10.1016/0370-2693(80)90212-9.

[173] Eugene P. Wigner. "On Unitary Representations of the Inhomogeneous Lorentz Group". In: *Annals Math.* 40 (1939). Ed. by Y. S. Kim and W. W. Zachary, pp. 149–204. DOI: 10.2307/1968551.

[174] Edward Witten. "Quantum Field Theory and the Jones Polynomial". In: *Commun. Math. Phys.* 121 (1989). Ed. by Asoke N. Mitra, pp. 351–399. DOI: 10.1007/BF01217730.

[175] Edward Witten. "Topological quantum field theory". In: *Commun. Math. Phys.* 117.3 (Sept. 1988), pp. 353–386.

[176] Ramona Wolf. "Microscopic Models for Fusion Categories". PhD Thesis. Jan. 2021. DOI: 10.15488/10324. arXiv: 2101.04154 [math-ph].

[177] Zhan Xu, Da-Hua Zhang, and Lee Chang. "Helicity Amplitudes for Multiple Bremsstrahlung in Massless Nonabelian Gauge Theories". In: *Nucl. Phys. B* 291 (1987), pp. 392–428. DOI: 10.1016/0550-3213(87)90479-2.

Appendix A

APPENDICE TO CHAPTER 2

A.1 Conventions

In this paper we follow the mostly minus convention $(+,-,-,-)$, and an on-shell momentum satisfies $p^2 = m^2$. The SL(2,C) and SU$_L$(2) indices are raised and lowered as

$$\psi_\alpha = \psi^\beta \varepsilon_{\alpha\beta}, \quad \psi^\alpha = \varepsilon^{\alpha\beta}\psi_\beta, \quad \varepsilon^{\alpha\beta}\varepsilon_{\beta\gamma} = \delta^\alpha_\gamma \tag{A.1}$$

where we use $\varepsilon_{\alpha\beta} = -\varepsilon^{\alpha\beta} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The spinor contraction can be converted to vector contraction following

$$p_1^{\alpha\dot{\alpha}} p_{2\alpha\dot{\alpha}} = 2p_1^\mu p_{2\mu}\,, \tag{A.2}$$

and hence for massive momenta, $p^{\alpha\dot{\alpha}}p_{\alpha\dot{\alpha}} = 2m^2$. The vector indices are converted to spinorial ones as:

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu_{\alpha\dot{\beta}} \\ \bar{\sigma}^{\mu\dot{\alpha}\beta} & 0 \end{pmatrix} \to (\not{p} + m) = \begin{pmatrix} m\delta_\alpha^\beta & p_{\alpha\dot{\beta}} \\ p^{\dot{\alpha}\beta} & m\delta^{\dot{\alpha}}_{\dot{\beta}} \end{pmatrix} \tag{A.3}$$

A.2 SU(2) Irreps as Symmetric Tensors

In this appendix we review, mostly to set notation, the elementary treatment of representations of $SU(2)$ as symmetric tensors, and briefly discuss some of its standard applictions, such as a transparent determination of spherical harmonics. The standard treatment of representations of $SU(2)$ is the one encountered by most undergraduates in beginning quantum mechanics courses. Since we can mutually diagonalize $\vec{J}^2$ and J_z, eigenstates of these operators are labeled by $|s, j_{\hat{z}}\rangle$, where the $\hat{z}$ reminds us that we have chosen to diagonalize the operator J_z, and we have $\vec{J}^2|s, j_{\hat{z}}\rangle = s(s+1)|s, j_{\hat{z}}\rangle$, $J_z|s, j_{\hat{z}}\rangle = m|s, j_{\hat{z}}\rangle$. The irrep is $(2s+1)$ dimensional with $j_{\hat{z}}$ taking all the values $-s \leq j_{\hat{z}} \leq +s$. The spin information in a general state $|\psi\rangle$ is then entirely contained in specifying $\langle s, j_{\hat{z}}|\psi\rangle$.

But for our purposes it is more convenient to describe an irrep of $SU(2)$ as a completely symmetric $SU(2)$ tensor with $2j$ indices:

$$\psi_{i_1\cdots i_{2s}} \tag{A.4}$$

where i is the $SU(2)$ index. The inner product $\langle\chi|\psi\rangle$ between two states is given by

$$\langle\chi|\psi\rangle = \varepsilon^{i_1 j_1}\cdots\varepsilon^{i_{2s}j_{2s}}(\chi_{i_1\cdots i_{2s}})^*\psi_{j_1\cdots j_{2s}} \tag{A.5}$$

Saying that ψ is an $SU(2)$ tensor is just the statement that the rotation generators $\vec{J}$ act as

$$(\vec{J}\psi)_{i_1\cdots i_{2s}} = (\frac{1}{2}\vec{\sigma})_{i_1}^{j_1}\psi_{j_1\cdots i_{2s}} + \cdots + (\frac{1}{2}\vec{\sigma})_{i_{2s}}^{j_{2s}}\psi_{i_1\cdots j_{2s}} \tag{A.6}$$

Note that the dimensionality of ths space is precisely $2\times3\times\cdots\times(2j+1)/(1\times2\times\cdots\times2j) = (2j+1)$ as desired. Using that $\vec{\sigma}_i^j \cdot \vec{\sigma}_k^l = 2\delta_k^j\delta_i^l - \delta_i^j\delta_k^l$, we trivially see that $(\vec{J}^2\psi)_{i_1\cdots i_{2s}} = s(s+1)\psi_{i_1\cdots i_{2s}}$. If we choose to diagonalize σ_z with eigenstates $(\sigma_z)_i^j\zeta_j^{\hat{z},\pm} = \pm\zeta_i^{\hat{z},\pm}$, then the spin s tensor that is an eigenstate of J_z with eigenvalue $j_{\hat{z}}$ is

$$\psi^{s,j_{\hat{z}}} = (\zeta^{\hat{z},+})^{s+j_{\hat{z}}}(\zeta^{\hat{z},-})^{j-j_{\hat{z}}} \tag{A.7}$$

where here and in what follows, since the tensor indices on ψ are always symmetrized there is no need to write them explicitly when no confusion can arise. We can also express the same fact in a different way, telling us how to extract $\langle s, j_{\hat{z}}|\psi\rangle$ from the tensor $\psi_{i_1,\cdots,i_{2s}}$:

$$\zeta_i \equiv \alpha_+\zeta_i^{\hat{z},+} + \alpha_-\zeta_i^{\hat{z},-};\ \zeta^{i_1}\cdots\zeta^{i_{2s}}\psi_{i_1\cdots i_{2s}} = \sum_{j_{\hat{z}}} \alpha_+^{s+j_{\hat{z}}}\alpha_-^{s-j_{\hat{z}}}\langle s, j_{\hat{z}}|\psi\rangle \tag{A.8}$$

The tensor representation makes it trivial to give explicit expressions for finite rotations, and expand the eigenstate $\psi^{s,j_{\hat{n}}}$ for a general direction $\hat{n}$ pointing in the usual (θ,ϕ) direction, as a linear combination of $\psi^{s,j_{\hat{z}}}$'s. We only need to know the relation for spin 1/2:

$$\begin{pmatrix} \zeta^{\hat{n},+} \\ \zeta^{\hat{n},-} \end{pmatrix} = \begin{pmatrix} c & -s^* \\ s & c \end{pmatrix} \begin{pmatrix} \zeta^{\hat{z},+} \\ \zeta^{\hat{z},-} \end{pmatrix}\ where\, c \equiv cos\frac{\theta}{2},\ s \equiv sin\frac{\theta}{2}e^{i\phi} \tag{A.9}$$

We can then look at

$$\begin{aligned} \psi^{s,j_{\hat{n}}} &= \left(\zeta^{\hat{n},+}\right)^{s+j_{\hat{n}}}\left(\zeta^{\hat{n},-}\right)^{s-j_{\hat{n}}} \\ &= \left(c\zeta^{\hat{z},+} - s\zeta^{\hat{z},-}\right)^{s+j_{\hat{n}}}\left(s^*\zeta^{\hat{z},+} + c\zeta^{\hat{z},-}\right)^{s-j_{\hat{n}}} \\ &= \sum_{j_{\hat{z}}} R^s_{j_{\hat{n}},j_{\hat{z}}}(\theta,\phi)\psi^{s,j_{\hat{z}}} \end{aligned} \tag{A.10}$$

with

$$R^s_{j_{\hat{n}},j_{\hat{z}}}(\theta,\phi) = \sum_{m_\pm, m_+ + m_- = s+j_{\hat{z}}} \begin{pmatrix} s+j_{\hat{n}} \\ m_+ \end{pmatrix}\begin{pmatrix} s-j_{\hat{n}} \\ m_- \end{pmatrix}(c)^{m_+}(-s)^{s+j_{\hat{n}}-m_+}(c)^{s-j_{\hat{n}}-m_-}(s^*)^{m_-} \tag{A.11}$$

The tensor formalism also makes it trivial to construct spherical harmonics, which naturally arise in building irreps of $SU(2)$ which are polynomials in a 3-vector $\vec{x}$. Of course we are used to converting $\vec{x}$ to $SU(2)$ indices by dotting with the σ matrices, but this gives us an object $\vec{\sigma}_i^j \cdot \vec{x}$ with an upstairs and downstairs index, while for the purposes of building irreps we would like to work with symmetric tensors and all downstairs indices. So it is natural to look instead at $x_{ij} = \epsilon_{ik}x_j^k$; explicitly we have

$$x_i^j = \begin{pmatrix} z & x-iy \\ x+iy & -z \end{pmatrix},\ x_{ij} = \begin{pmatrix} -(x-iy) & z \\ z & (x+iy) \end{pmatrix} \tag{A.12}$$

We would like to make symmetric rank $2s$ tensors from a product of s x_{ij}'s. But we don't need to do the symmetrizations explicitly; again because of the symmetrization all the information is contained in

$$\zeta^{i_1}\zeta_{j_1}\cdots\zeta^{i_s}\zeta_{j_s}x_{i_1j_1}\cdots x_{i_sj_s} = (\zeta\zeta x)^s \tag{A.13}$$

Putting $\zeta_i = (\alpha_+, \alpha_-)$ and so $\zeta^i = (-\alpha_-, \alpha_+)$, expanding the above gives us the generating function for spherical harmonics. Letting $\vec{x}$ be the unit vector with $(x+iy) = sin(\theta)e^{i\phi}$ and $z = cos(\theta)$, we have

$$(\zeta\zeta x)^s = \left(\alpha_+^2 sin(\theta)e^{i\phi} - 2\alpha_+\alpha_- cos(\theta) - \alpha_-^2 sin(\theta)e^{-i\phi}\right)^s \equiv \sum_{j_{\hat{z}}} \alpha_+^{s+j_{\hat{z}}}\alpha_-^{s-j_{\hat{z}}}Y_{s,j_{\hat{z}}}(\theta,\phi) \tag{A.14}$$

A.3 Explicit Kinematics

For massless particles, we have

$$\lambda_\alpha = \sqrt{2E}\begin{pmatrix} c \\ s \end{pmatrix}, \tilde{\lambda}_{\dot{\alpha}} = \sqrt{2E}\begin{pmatrix} c \\ s^* \end{pmatrix} \tag{A.15}$$

For massive particles, we can write

$$\lambda_\alpha^I = \begin{pmatrix} \sqrt{E+p}c & -\sqrt{E-p}s^* \\ \sqrt{E+p}s & \sqrt{E-p}c \end{pmatrix}, \tilde{\lambda}_{I\dot{\alpha}} = \begin{pmatrix} \sqrt{E+p}c & -\sqrt{E-p}s^* \\ \sqrt{E+p}s & \sqrt{E-p}c \end{pmatrix} \tag{A.16}$$

We can write this equivalently as

$$\begin{aligned} \lambda_\alpha^I &= \sqrt{E+p}\zeta_\alpha^+(p)\zeta^{-I}(k) + \sqrt{E-p}\zeta_\alpha^-(p)\zeta^{+I}(k) \\ \tilde{\lambda}_{\dot{\alpha}}^I &= \sqrt{E+p}\tilde{\zeta}_{\dot{\alpha}}^-(p)\zeta^{+I}(k) + \sqrt{E-p}\tilde{\zeta}_{\dot{\alpha}}^+(p)\zeta^{-I}(k) \end{aligned} \tag{A.17}$$

where

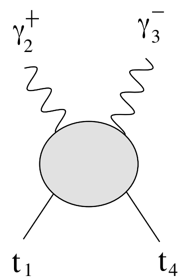
$$\zeta_\alpha^+ = \begin{pmatrix} c \\ s \end{pmatrix}, \tilde{\zeta}_{\dot{\alpha}}^- = \begin{pmatrix} c \\ s^* \end{pmatrix};\ \zeta_\alpha^- = \begin{pmatrix} -s^* \\ c \end{pmatrix}, \tilde{\zeta}_{\dot{\alpha}}^+ = \begin{pmatrix} -s \\ c \end{pmatrix} \tag{A.18}$$

We can read off the specific spin components as in the previous appendix, since by using the above expressions for $\lambda_\alpha^I, \tilde{\lambda}_{\dot{\alpha}}^I$ we can expand for any particle:

$$M^{\{I_1\cdots I_{2S}\}} = \sum_{j_z}\left((\zeta^+)^{S+j_z}(\zeta^-)^{S-j_z}\right)^{\{I_1\cdots I_{2S}\}} M(j_z) \tag{A.19}$$

A.4 Comparison with Feynman Diagrams for Compton Scattering

Here we directly construct Compton scattering from Feynman rules, and converting into our notations. We begin with



$$: \quad \epsilon_2^\mu \epsilon_3^\nu \bar{v}_1 \left(\frac{\gamma^\nu (\not{P}_{12} + m)\gamma^\mu}{s - m^2} + \frac{\gamma^\mu (\not{P}_{13} + m)\gamma^\nu}{u - m^2} \right) u_4 \,, \tag{A.20}$$

where $P_{ij} = p_i + p_j$. Peeling off u_4 and $\bar{v}_1$, we obtain two 4×4 numerator factor each given by:

$$n_s = \begin{pmatrix} m\epsilon_{3\alpha\dot{\gamma}}\epsilon_2^{\dot{\gamma}\delta} & \epsilon_{3\alpha\dot{\beta}}(P_{21})^{\dot{\beta}\gamma}\epsilon_{2\gamma\dot{\delta}} \\ \epsilon_3^{\dot{\alpha}\beta}(P_{12})_{\beta\dot{\gamma}}\epsilon_2^{\dot{\gamma}\delta} & m\epsilon_3^{\dot{\alpha}\gamma}\epsilon_{2\gamma\dot{\delta}} \end{pmatrix}, \quad n_u = \begin{pmatrix} m\epsilon_{2\alpha\dot{\gamma}}\epsilon_3^{\dot{\gamma}\delta} & \epsilon_{2\alpha\dot{\beta}}(P_{13})^{\dot{\beta}\gamma}\epsilon_{3\gamma\dot{\delta}} \\ \epsilon_2^{\dot{\alpha}\beta}(P_{13})_{\beta\dot{\gamma}}\epsilon_3^{\dot{\gamma}\delta} & m\epsilon_2^{\dot{\alpha}\gamma}\epsilon_{3\gamma\dot{\delta}} \end{pmatrix}. \tag{A.21}$$

Substituting the explicit polarization vectors one finds:

$$n_s = \frac{\begin{pmatrix} m\lambda_{3\alpha}[\tilde{q}2]q^\delta & \lambda_{4\alpha}[\tilde{q}|P_{12}|q\rangle\tilde{\lambda}_{2\dot{\delta}} \\ \tilde{q}^{\dot{\alpha}}\langle 3|p_1|2]q^\delta & m\tilde{q}^{\dot{\alpha}}\langle 1q\rangle\tilde{\lambda}_{4\dot{\delta}} \end{pmatrix}}{\langle 2q\rangle[3\tilde{q}]}, \quad n_u = \frac{\begin{pmatrix} mq_\alpha[2\tilde{q}]\lambda_3^\delta & q_\alpha[2|p_1|3\rangle\tilde{q}_{\dot{\delta}} \\ \tilde{\lambda}_2^{\dot{\alpha}}\langle q|P_{12}|\tilde{q}]\lambda_3^\delta & m\tilde{\lambda}_2^{\dot{\alpha}}\langle q3\rangle\tilde{q}_{\dot{\delta}} \end{pmatrix}}{\langle 2q\rangle[3\tilde{q}]} \tag{A.22}$$

where $q, \tilde{q}$ are the reference spinors for the polarization vectors. The elements in the 4×4 matrix is in different SL(2,C) representations. We again judicially multiply factors of p/m to convert it into our preferred basis, which has leg 1 in the undotted basis, and leg 4 in the dotted basis. That is, we multiply:

$$\begin{pmatrix} \frac{p_4^{\dot{\alpha}\alpha}}{m} & \delta^{\dot{\alpha}}_{\dot{\beta}} \end{pmatrix} \begin{pmatrix} \mathcal{O}_\alpha{}^\delta & \mathcal{O}_{\alpha\dot{\delta}} \\ \mathcal{O}^{\dot{\beta}\delta} & \mathcal{O}^{\dot{\beta}}{}_{\dot{\delta}} \end{pmatrix} \begin{pmatrix} \delta^\beta_\delta \\ \frac{p_1^{\dot{\delta}\beta}}{m} \end{pmatrix} \tag{A.23}$$

where the $\mathcal{O}$s are stand ins for matrix elements of n_s, n_u. Summing up the terms and choosing $q = \lambda_3$ and $\tilde{q} = \tilde{\lambda}_2$, one finds:

$$\frac{\langle 3|p_1|2](\tilde{\lambda}_2^{\dot{\alpha}}\lambda_3^\beta - p_{4\alpha}^{\dot{\alpha}}\lambda_3^\alpha p_1^\beta{}_{\dot{\delta}}\tilde{\lambda}_2^{\dot{\delta}}/m^2)}{(u - m^2)(s - m^2)} \,. \tag{A.24}$$

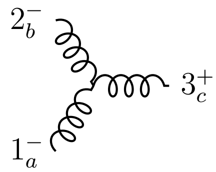
Contracting with external $\lambda, \tilde{\lambda}$s we recover eq.(2.99).

A.5 The High Energy Limit of Massive Three-Point Amplitude

Let us consider the HE limit of the three-point massive vector amplitude

$$\frac{gf^{abc}}{m_a m_b m_c}\left[\langle\mathbf{12}\rangle[\mathbf{12}]\langle\mathbf{3}|p_1{-}p_2|\mathbf{3}] + \text{cyc.}\right] \tag{A.25}$$

First consider the component amplitude $(1^-2^-3^+)$. Its high energy limit is given by:



$$\rightarrow \frac{1}{m_a m_b m_c} \tag{A.26}$$

$$\cdot\, g f^{abc} \left(\langle 12\rangle[\tilde\eta_1\tilde\eta_2]\langle\eta_3|p_1-p_2|3] + \langle 2\eta_3\rangle[\tilde\eta_2 3]\langle 1|p_2-p_3|\tilde\eta_1] + \langle\eta_3 1\rangle[3\tilde\eta_1]\langle 2|p_3-p_1|\tilde\eta_2]\right)$$

Since in the high energy limit we will be interested in the MHV configuration, we have:

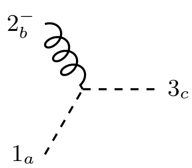
$$\tilde\lambda_1 = \langle 23\rangle\tilde\xi, \quad \tilde\lambda_2 = \langle 31\rangle\tilde\xi, \quad \tilde\lambda_3 = \langle 12\rangle\tilde\xi\,, \tag{A.27}$$

and eq.(A.26) simplifies to:

$$\begin{aligned}&\frac{g f^{abc}\left(\langle 2\eta_3\rangle[\tilde\eta_2 3]\langle 1|p_2-p_3|\tilde\eta_1] + \langle\eta_3 1\rangle[3\tilde\eta_1]\langle 2|p_3-p_1|\tilde\eta_2]\right)}{m_a m_b m_c}\\ =\;& 2\frac{g f^{abc}}{m_c}\left(\frac{\langle\eta_3 2\rangle\langle 12\rangle^2}{\langle 23\rangle} + \frac{\langle\eta_3 1\rangle\langle 12\rangle^2}{\langle 31\rangle}\right) = 2 g f^{abc}\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 31\rangle}.\end{aligned} \tag{A.28}$$

where we have repeatedly used identities such as $[\tilde\eta_1 3] = \frac{\langle 12\rangle}{\langle 23\rangle}[\tilde\eta_1 1] = m_a\frac{\langle 12\rangle}{\langle 23\rangle}$, which holds for MHV kinematics.

A more interesting component would be $(1^0 2^- 3^0)$. Keeping in mind that extracting the longitudinal term corresponds to choosing $\lambda^{\{I}\tilde\lambda^{J\}} \rightarrow \lambda\tilde\lambda - \eta\tilde\eta$, the relevant terms are:



$$\begin{aligned}\rightarrow\;& \frac{g f^{abc}}{m_a m_b m_c}\Big\{\langle 12\rangle[1\tilde\eta_2]\left(\langle 3\eta_1\rangle[\tilde\eta_1 3] - \langle 3\eta_2\rangle[\tilde\eta_2 3] - \langle 1\eta_3\rangle[\tilde\eta_3 1] + \langle 2\eta_3\rangle[\tilde\eta_3 2]\right)\\ +\;& \langle 23\rangle[\tilde\eta_2 3]\left(\langle 1\eta_2\rangle[\tilde\eta_2 1] - \langle 1\eta_3\rangle[\tilde\eta_3 1] - \langle\eta_1 2\rangle[2\tilde\eta_1] + \langle\eta_1 3\rangle[3\tilde\eta_1]\right)\\ -\;& \left(\frac{1}{2}\langle 23\rangle[3\tilde\eta_2]\left(\langle\eta_3 1\rangle[\tilde\eta_3 1] + \langle\eta_1 3\rangle[\tilde\eta_1 3]\right) - \frac{1}{2}\langle 21\rangle[1\tilde\eta_2]\left(\langle\eta_3 1\rangle[\tilde\eta_3 1] + \langle\eta_1 3\rangle[\tilde\eta_1 3]\right)\right)\Big\}\,.\end{aligned} \tag{A.29}$$

Substituting explicit representation for $[\tilde\eta_i j]$ for MHV kinematics, one finds:

$$\frac{g f^{abc}}{m_a m_c}\frac{\langle 12\rangle\langle 23\rangle}{\langle 31\rangle}\left(m_b^2 - m_c^2 - m_a^2\right)\,. \tag{A.30}$$

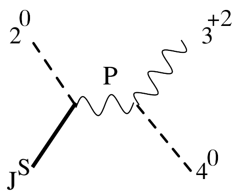
A.6 Examples for 1 Massive 3 Massless Amplitudes

For three-point amplitudes, since the all massless and one-massive two-massless amplitudes are unique, this tells us that the massless residue for the one-massive three-massless amplitude is unique. If the residue is non-local, then consistent factorization in the other channel may force the theory to have a particular one-massless two-massive interaction. Here we present some examples.

We consider the four-point amplitude of arbitrary higher spin-S, two massless scalars and a graviton:

$$M(\mathbf{1}^S 2^0 3^{+2} 4^0)\,. \tag{A.31}$$

We can now look at the massless residue for s-channel,

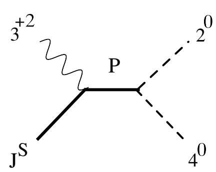


$$\frac{(\lambda_2)^S(\lambda_P)^S[2P]^S}{m^{2S-1}} \times \frac{[3P]^2[34]^2}{[4P]^2 M_{pl}} = \frac{(\lambda_2)^S([2|p_1)^{S-2}}{m^{2S-5}}\frac{[34]^2(\lambda_4)^2}{\langle 23\rangle^2 M_{pl}}\,, \tag{A.32}$$

where M_{pl} is the Plank mass. Note that we have double poles $1/\langle 23\rangle^2$, which is a general feature for couplings involving gravitons. The presence of double poles indicate that we have access to information in other channel. Let's start with $S=2$; dressing the residue with $1/s$ propagator, we find:

$$\frac{m}{M_{pl}}\frac{1}{s}(\lambda_2)^2(\lambda_4)^2\frac{[34]^2}{\langle 23\rangle^2} = \frac{m}{M_{pl}}\frac{(\lambda_2)^2(\lambda_4)^2[34][23]}{\langle 32\rangle\langle 43\rangle t} \quad \to \quad M(\mathbf{1}^2 2^0 3^{+2} 4^0) = \frac{m}{M_{pl}}\frac{(\lambda_2)^2(\lambda_2)^2[34][23]}{\langle 32\rangle\langle 34\rangle(u-m^2)}\,. \tag{A.33}$$

Note that the double pole has been converted into a t-channel massless and an u-channel massive pole $u-m^2$. The residue of the massive channel can be identified with $M_3(\mathbf{1}^{S=2}3^{+2}\mathbf{P}^{S=2}) \times M_3(\mathbf{P}^{S=2}2^0 4^0)$, where $M_3(\mathbf{1}^{S=2}3^{+2}\mathbf{P}^{S=2})$ is the minimally coupling between a graviton and massive spin-2 states. Indeed using minimal coupling in the u-channel, we find the following residue:



$$x_{13}^2 \times [24]^2(\lambda_2)^2(\lambda_4)^2 \sim (\lambda_2)^2(\lambda_4)^2\frac{[34][23]}{\langle 43\rangle\langle 23\rangle}\,, \tag{A.34}$$

which indeed matches that of eq.(A.33). This is a general feature for amplitudes of eq.(A.31), consistent factorization will require the presence of a three point minimal coupling for graviton to two massive states. Consider $S=3$, the s-channel residue can be represented in a way that it can readily be completed:

$$(\lambda_2)^S([2|p_1)^{S-2}\frac{[34]^2(\lambda_4)^2}{\langle 23\rangle^2}\bigg|_{\langle 34\rangle=0} = (\lambda_2)^3(\lambda_4)^3\left(\frac{[34]^2[32]}{\langle 23\rangle\langle 24\rangle} - \frac{[42][34]^2[23]}{\langle 23\rangle t}\right)\,, \tag{A.35}$$

Indeed putting back the s-channel propagator and writing $-t \to (u-m^2)$, we find the form factor given as:

$$M(\mathbf{1}^3 2^0 3^{+2} 4^0) = (\lambda_2)^3(\lambda_4)^3\left(\frac{[34][32]}{\langle 23\rangle\langle 24\rangle\langle 43\rangle} + \frac{[42][34][23]}{\langle 23\rangle\langle 43\rangle(u-m^2)}\right)\,. \tag{A.36}$$

It is not difficult to see that the massive residue of this amplitude contains the minimum coupling for the spin-3 states:

$$x_{13}^2(\lambda_2)^3(\lambda_4)^3[24]^3 \sim (\lambda_2)^3(\lambda_4)^3\frac{[23][34][24]}{\langle 23\rangle\langle 43\rangle}\,. \tag{A.37}$$

Appendix B

APPENDICE TO CHAPTER 3

B.1 Causality constraints on amplitudes

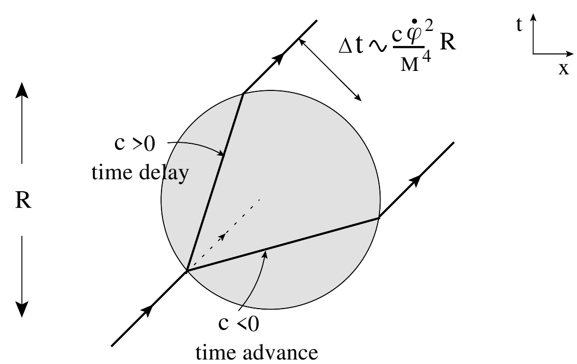
Time delay and positivity bounds

It is well-known that causality puts interesting positivity bounds on the amplitude in the low energy effective field theories. Perhaps the simplest example is the case of a single derivatively coupled scalar with lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 + \frac{c}{M^4}(\partial\phi)^4 + \cdots . \tag{B.1}$$

The claim is causality demands $c > 0$ [4]. This is slightly surprising at first sight: c reflects unknown physics in the UV, ordinarily we can only probe higher-dimension operators if they violate a symmetry of the low-energy theory, but that is not the case here. And indeed, there is nothing *obviously* wrong with this as a **Euclidean** EFT. However in the physical Lorentzian world, there *is* something "right-on-the-edge" in the 2-derivative theory: ϕ excitations propagate *exactly* on the light-cone. It can happen that in simple backgrounds the coefficient for the higher-dimensional operators push propagation *outside* the light-cone. We can consider for instance the spatially translationally invariant background $\phi = \phi_0 + \varphi$ where $\dot{\phi}_0 \neq 0$. We can make $(\dot{\phi}_0/M^2)$ as tiny as we like such that the background is trustworthy within the EFT. The background breaks Lorentz invariance and small fluctuations propagate with speed $v = (1 - \frac{c\dot{\phi}_0^2}{M^4})$, so we must have $c > 0$ to avoid superluminality.

Note that despite being associated with a higher-dimensional operator, the effect of the superluminality is not "small. Indeed if we turn on $\dot{\phi}_0 \neq 0$ inside some bubble of radius R, and throw in a φ excitation, we get a time advance/delay of φ propagation that is $\delta t = \delta_v R = \frac{c\dot{\phi}_0^2}{M^4} R$, which can be made arbitrarily large by increasing R.



This highlights the fundamental fact that the usual Wilsonian intuition about the decoupling of "short-distance" from "long-distance" physics is **fundamentally Euclidean**. In Euclidean signature, to probe a distance $(x-y)^2 \sim \frac{1}{\Lambda_{\mathrm{UV}}^2}$, one needs probes with wavelength near the UV scale Λ_{UV}. By contrast in Minkowski space, ultra-small spacetimes

$(x-y)^2 \sim \frac{1}{\Lambda_{\rm UV}^2}$ can be probed by very long-distance experiments since (x, y) can be separated by huge distances and time but be close to the light-cone, with advances/delays that can be made parametrically large.

As is also well-known, these positivity constraints can also be derived from unitarity plus dispersion relations, reflecting the historic origin of analytic properties of Green's functions and amplitudes in the investigation of causal propagation! We will recap this story, but instead of jumping from the classical picture of φ propagation around a background to dispersion relations for the forward $2 \to 2$ scattering amplitude, we will connect the two pictures directly, by repeating the above analysis, preformed in the language of classical field theory, in terms of particle propagation plus scattering. As we will see this will in fact give us more than simply the positivity of the $(\partial\phi)^4$ coefficient; we will see that

$$\frac{\partial}{\partial s}\frac{M(s)}{s} > 0\,, \tag{B.2}$$

where $M(s)$ is the four particle φ scattering amplitude in the forward limit as $t \to 0$.

As is ubiquitous in the quantum particle-classical field theory connection for bosons, we recover the classical field picture of time advance/delay for small fluctuations about the background, by considering the scattering of a single hard φ quanta, against a bose condensate of a large number N of soft ϕ_0 quanta, representing the blob. We begin by recalling familiar undergraduate basics on wave-packets and the connection between amplitude phase shifts and time delays: first, free propagation, where we have one particle states with momentum $\vec{p}$. From these we can build good approximation to particles moving with constant momentum trajectories. We can define the state $|\vec{x}_*, \vec{p}_*; t_*\rangle$ as

$$|\vec{x}_*, \vec{p}_*; t_*\rangle = \int d^d p \;\; e^{i(\vec{p}\cdot\vec{x}_* - E(\vec{p})t_*)}\Psi_{\Delta p}(\vec{p} - \vec{p}_*)\,, \tag{B.3}$$

where $\Psi_{\Delta p}(\vec{p} - \vec{p}_*)$ is sharply localized around $\vec{p} = \vec{p}_*$, for example $\Psi_{\Delta p}(\vec{p} - \vec{p}_*) \propto e^{-(\vec{p}-\vec{p}_*)^2/(\Delta p)^2}$. With this definition, we can compute $|\langle \vec{x}_2, \vec{p}, t_2 | \vec{x}_1, \vec{p}, t_1\rangle|^2$ via stationary phase approximation, giving

$$|\langle \vec{x}_2, \vec{p}, t_2 | \vec{x}_1, \vec{p}, t_1\rangle|^2 = e^{-\Delta p^2\left((\vec{x}_1 - \vec{x}_2) - \vec{V}(\vec{p})(t_1 - t_2)\right)^2}\,, \tag{B.4}$$

where $\vec{V}(\vec{p}) = \frac{\partial E(\vec{p})}{\partial \vec{p}}$; this peaked on the classical constant velocity trajectory $\Delta\vec{x} = \vec{V}\Delta t$ with the unavoidable quantum-mechanical uncertainty of order $\frac{1}{\Delta p}$.

Now let's instead imagine that we are propagating through our blob above. Now in computing the same overlap, we will need the S-matrix element for φ scattering off the blob,

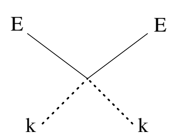
$$\langle B, \vec{p}|S|B, \vec{p}\rangle = e^{i\delta(E(\vec{p}))}. \tag{B.5}$$

Note that the momentum uncertainty/transfer associated with the blob is $k \sim \frac{1}{R}$, which we assume to be much smaller than $|\vec{p}|$, so the outgoing momentum is the same as the incoming one. We also assume no other particles were produced, so this amplitude is just a phase $e^{i\delta(E(\vec{p}))}$. Repeating the stationary phase analysis, we now find that

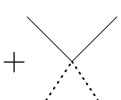
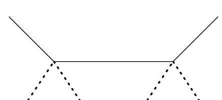
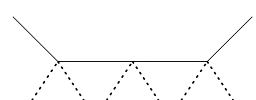
$$|\langle \vec{x}_2, \vec{p}, t_2 | \vec{x}_1, \vec{p}, t_1 \rangle|^2 = e^{-\Delta p^2 \left(\Delta x - \vec{V}(\vec{p})(\Delta t + \frac{\partial \delta(E)}{\partial E})\right)^2} . \tag{B.6}$$

Thus, the presence of the blob has given us a time delay/advance given by $\Delta^{blob} t = \frac{\partial \delta(E)}{\partial E}$. In order for this to be detectable above the quantum uncertainty $\Delta^{quantum} t \sim \frac{1}{\Delta p} \sim \frac{1}{E}$, clearly we must have that the phase $\delta(E) \gg 1$ is parametrically large.

Thus to find a situation where the delay/advance is reliably calculable, we must find a setting where $\delta(E) \gg 1$ is reliably calculable. Now when we consider few particle scattering in any situation with a weak coupling, where amplitudes are reliably calculable, essentially by definition the phase above will be perturbatively small. However, $\delta(E) \gg 1$ is exactly what happens when we scatter φ off the condensate "blob", which we can think of as a large number N of φ quanta with $k \sim \frac{1}{R}$. Note that the relation between N and the classical background field $(\partial \phi_0)$ is given by matching the energy of the blob in the two pictures, as $N \sim (\partial \phi_0)/k^4$. Now let's consider $M = \langle B, E|S|B, E\rangle$ computed in perturbation theory. We can take momentum of order k for the background. At lowest order, we have

$M = 1 +$  $+ \cdots = 1 + iA(s{=}kE) + \cdots .$

Again so long as we have weak coupling, $A(s = kE)$ is small. But since k is so small, the corrections from multi-particle scattering are *significantly enhanced* by the s-channel propagator $\frac{1}{s} \sim \frac{1}{kE}$. Thus the full amplitude is then the sum over all disconnected graphs, scattering of $0, 2, 4, \cdots, m$ soft particles

$M = 1+$  $+$  $+$  $+ \cdots$

$$iA^f(s) \quad + \left[iA^f(s)\right] \frac{1}{kE} \left[iA^f(s)\right] \quad + \left[iA^f(s)\right] \frac{1}{kE} \left[iA^f(s)\right] \frac{1}{kE} \left[iA^f(s)\right] . \tag{B.7}$$

where, since we imagine $k \sim \frac{1}{R}$ is tiny, $A^f(s)$ is the forward-limit amplitude. Note these are amplitudes with the conventional relativistic normalization of states: $M = \langle B, E|S|B, E\rangle$ is dimensionless and the units are made up for with powers of k. At large m number of scattering, we have

$$M = \sum_m \left(\frac{iA^f(s)k}{E}\right)^m \begin{pmatrix} N \\ m \end{pmatrix} = \left(1 + \frac{iA^f(s)k}{E}\right)^N . \tag{B.8}$$

Using $N \sim (\partial\phi_0)^2/k^4$ we have

$$M = exp\left[i\frac{A^f(s)}{Ek}\frac{(\partial\phi_0)^2}{k^2}\right], \tag{B.9}$$

and thus we can identify

$$\delta(E) = \left(\frac{A(s)}{s}\right)\frac{(\partial\phi_0)^2}{k^2}. \tag{B.10}$$

So the time delay is

$$\boxed{\Delta t = \frac{\partial}{\partial E}\delta(E) = \frac{\partial}{\partial s}\left(\frac{A(s)}{s}\right)(\partial\phi_0)^2 R}\,. \tag{B.11}$$

From here, we can reproduce the previous result for the $c(\partial\varphi)^4$ theory: there $A(s) = \frac{cs^2}{M^4}$, so $\Delta t = \frac{c}{M^4}(\partial\phi_0)^2 R$.

As another quick check, suppose we had turned on a $\lambda\varphi^4$ interaction. Then $A(s) \sim -\lambda$, and $\Delta t = \frac{\lambda}{s^2}\frac{(\varphi_0)^2}{R^2}R \sim \lambda\frac{(\varphi_0)^2}{E^2}R$. This is again as we'd expect: inside the blob the φ particle picks up a mass $m_0^2 \sim \lambda\varphi_0^2$. So if the velocity (for $E \gg m_0$) is reduced to $(1 - \frac{m_0^2}{E^2}) = (1 - \lambda\frac{\varphi_0^2}{E^2})$, this leads to a time delay of $\Delta = \lambda\frac{\varphi_0^2}{E^2}R$. Note however that if $\lambda < 0$ this does *not* mean we have superluminal propagation; and indeed it is possible to have consistent theories with $\lambda < 0$, with vacuum instability on exponentially long time scales $\propto exp(b/|\lambda|)$ as in the Higgs instability in the Standard Model. If $\lambda < 0$, turning on φ_0 destabilizes the vacuum inside the bubble, and so the perturbative assumption of this computation is violated. Strictly speaking then, our arguments says that $\frac{\partial}{\partial s}\left(\frac{A(s)}{s}\right) > 0$ so long as $A(s = 0) \leq 0$ (which allows of course for $A(s = 0)$ as for goldstones).

Thus from consideration of scattering off the blob, we conclude that $\frac{\partial}{\partial s}\left(\frac{A(s)}{s}\right) > 0$, a stronger statement than merely the positivity of the coefficient of (s^2) in the low energy expansion of $A(s)$.

We now switch gears to discuss the dispersive representation of the (forward) scattering amplitude, and show how analyticity and unitarity allow us to conclude that $\frac{\partial}{\partial s}\left(\frac{A(s)}{s}\right) > 0$ when $A(s = 0) \leq 0$. The non-trivial statement that makes this is possible is the Froissart bound, which we will review shortly, following from assumptions of analyticity and a reasonable polynomial boundedness of the forward amplitude. The bound tells us that $A^f(s) < s\log^2 s$ at large s, and so Cauchy's theorem allows us to express for a single scalar φ (with $s{-}u$ symmetry)

$$A(s) = A_0 + \int dM^2 \rho(M^2)\left[\frac{1}{M^2-s} + \frac{1}{M^2+s} - \frac{2}{M^2}\right] \tag{B.12}$$

where we've separated out the constant piece $A_0 = A(s=0)$, since these are not captured by contour integration, and the expression in the brackets vanishes at $s = 0$. Of course unitarity tells us that $\rho(M^2) \geq 0$. Now we simply note that

$$\frac{\partial}{\partial s}\frac{1}{s}\left[\frac{1}{M^2{-}s} + \frac{1}{M^2{+}s} {-} \frac{2}{M^2}\right] = \frac{2(M^4+s^2)}{M^2(M^4-s^2)^2} > 0\,, \tag{B.13}$$

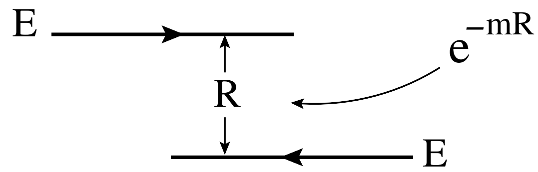
and thus if $A_0 <\leq 0$ so that $\frac{\partial}{\partial s}\frac{A_0}{s} > 0$, we have that $\frac{\partial}{\partial s}\left(\frac{A(s)}{s}\right) > 0$ as desired. This shows quite vividly how unitarity and analyticity in the UV guarantee a rather non-trivial condition needed for IR causality.

We have seen that reliable causality constraints on scattering amplitudes can arise if we can find a background in which small, perturbative amplitude phase-shifts can be calculably exponentiated to large phases, that allow us to look for the presence of a time advance or delay in the scattering process. We have discussed one such background:the "soft blob" of a scalar condensate, through which we shot a hard probe. Another limit of this kind arises when we have gravitational long-range forces, and consider the scattering in the Eikonal limit, or equivalently, shooting a probe particle through a gravitational shock wave [45]. In the impact parameter representation, where the impact parameter $\vec{b}$ is fourier-conjugate to the momentum transfer $\vec{q}$ with $t = -\vec{q}^2$, the amplitude again exponentiates to a phase $\delta(s,\vec{b})$ at small $\vec{b}$. If further we assume the UV theory has a weak coupling and so a scale of new physics beneath the Planck scale, as in string theory, at fixed t, the leading weak coupling amplitude at large s scales as $a(s)/t$, which maps to an Eikonal phase $\delta(s,\vec{b}) = \frac{a(s)}{s}\log b$. The center of mass energy $s = E_{probe}E_{shock}$; causality and unitarity demand that $|e^{i\delta(E_{probe})}| < 1$ everywhere in the upper-half E_{probe} plane, and this tells us that $\delta(E_{probe})$ itself must be bounded by E^1_{probe} at large E_{probe}. This in turn tells us that the fixed t amplitude is bounded by s^2 at large s. This is easily seen to be satisfied for gravity amplitudes in string theory, which has a Regge behavior at fixed t, large s given by $s^{2+\alpha' t}/t$, giving a power smaller than s^2 for physical $t < 0$.

It is amusing that, while the "small-phase exponentiating backgrounds" are different in these two examples, the final practical constraint on the high-energy behavior of amplitudes is the same. The usual Froissart bound (whose derivation we will review in a moment) tells us that the amplitude at fixed t can grow only logarithmically faster than s, while the shockwave arguments applicable for weakly coupled in the UV gravitational theories tells us that the amplitude can't grow as fast as s^2. In both cases, we learn that the amplitude is bounded by s^2 at fixed t.

Froissart bound

Let's recall first the intuition behind the Froissart bound, going back to an argument by Heisenberg [98]. Consider particles scattering at center of mass energy E, involving exchange of a particle with mass m. We can imagine the interaction strength grows as gE^n, but in position space we also expect the amplitude to behave as e^{-mR}:



Thus the relevant contributions are given by

$$(gE^n)e^{-mR} \sim 1 \quad \to R \leq \frac{n \log E}{m}\,, \tag{B.14}$$

so the total cross section should be bounded by

$$\sigma \sim R^2 \leq \frac{n^2 \log^2 E}{m^2}\,. \tag{B.15}$$

Now since $\sigma(s) = \frac{Im[M(s,t\to 0)]}{s}$, this also tells us that

$$Im[M(s,t \to 0)] \leq \frac{cs \log^2 s}{m^2}\,. \tag{B.16}$$

for some constant c at large s. Note that locality, seen in the finite range of the effective interaction, was crucial to this argument.

We'd like to see how to understand this intuitive result directly from properties of the amplitude. Very naively, one might think that an upper bound on the amplitude would come from unitarity, but this is not enough; as we've seen locality is also crucial, and thus some "good" analytic properties of the amplitude must also be needed. To begin with, let's write the partial wave expansion of the amplitude

$$M(s, \cos\theta) = \sum_\ell (2\ell{+}1)\, a_\ell\, P_\ell(\cos\theta) \quad \to \quad M(s,t) = \sum_\ell (2\ell{+}1)\, a_\ell\, P_\ell(1{+}2t/s)\,. \tag{B.17}$$

Unitarity tells us that $|1{+}ia_\ell|^2 \leq 1$ so $0 \leq |a_\ell|^2 \leq 2\mathrm{Im}\, a_\ell \leq 1$. Note the extremely naive intuition that "unitarity means $A(s)$ can't get too big" is wrong, since unitarity only tells us each a_ℓ individually can't get too big. Indeed if we keep all a_ℓ's to be $\mathcal{O}(1)$ up to some $\ell \sim \ell_{\max}$, we'd have that

$$M(s,0) = \sum_{\ell \leq \ell_{max}} (2\ell{+}1)\, a_\ell \, \sim \ell^2_{max}\,. \tag{B.18}$$

Going again to the Heisenberg picture, at the distance $R_{\max} \sim \frac{\log E}{m}$, the angular momentum is $\ell_{max} \sim ER_{\max} \sim \frac{E\log E}{m}$, so $M \leq \ell_{\max}^2 \sim s\frac{\log^2 s}{m^2}$ would agree with our Froissart intuition.

So unitarity is not enough, we need an extra argument to tell us that the partial waves above $\ell_{\max}(E) \sim E\log E$ are shut off. Let's imagine working at fixed t smaller than any of the thresholds. Importantly we assume that the amplitude at fixed s is analytic in t: in other words, we can continue from small negative t (i.e. the physical region) to small positive t smoothly. We will also have at fixed but small t, the amplitude is polynomial bounded at large s, $M < s^N$. We've already seen heuristic reasons for this from causality, though those are only applicable for physical (negative) t. It is our assumption of analyticity in t for small enough t that allows us to continue the bound to positive t, which is crucial for the following argument. Now the Legendre polynomials $P_\ell(x)$ are wildly oscillating for large ℓ when $|x| = \cos\theta < 1$, but for $x > 1$ they instead are exponentially growing:

$$P_\ell\left(1+\frac{2t}{s}\right) \sim \frac{1}{\sqrt{\ell}}e^{2\ell\sqrt{\frac{2t}{s}}} \tag{B.19}$$

for $t/s > 0$. Now consider $\text{Im}M(s,t) = \sum_\ell (2\ell+1)\text{Im}a_\ell\, P_\ell(1+\frac{2t}{s})$. If we want this to be bounded by s^N at large s, $\text{Im}a_\ell$ have to sharply die above some $\ell_{\max(s)}$, estimated as

$$e^{2\ell_{\max}(s)\sqrt{\frac{2t}{s}}} < s^N \quad \to \quad \ell_{\max}(s) \sim N\sqrt{\frac{s}{2t}}\log s\,. \tag{B.20}$$

Note this is in agreement with what we expect from the Heisenberg picture; taking $t \sim 1/R^2$, we have $\ell_{\max} \sim NRE\log E$ as expected. From here, we recover the Froissart bound.

Note we can also say slightly more, not just about the imaginary part of the amplitude, but the amplitude itself. We've already seen that $\text{Im}a_\ell \to 0$ for $\ell > \ell_{\max}(s)$. But since by unitarity we have $|a_\ell|^2 < 2\text{Im}\, a_\ell$, this means that $\text{Re}a_\ell \to 0$. Thus we learn that for small enough $|t|$

$$M(s,t) \leq s\log^2 s \tag{B.21}$$

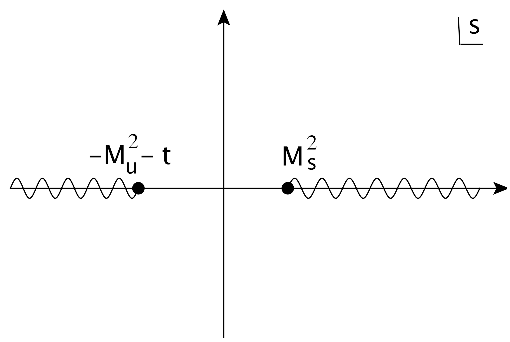
for large s. This is interesting: we began only by assuming $M(s,t) < s^N$ for **some** power N; but analyticity in t for small t, and unitarity, then forces upon us the much stronger statement that $M(s,t) < s\log^2 s$.

B.2 Dispersive representation of loop amplitudes

In this section, we will show that by integrating out massive states in loops, so long as $t \ll m^2$, the four point amplitude admits the following dispersive representation:

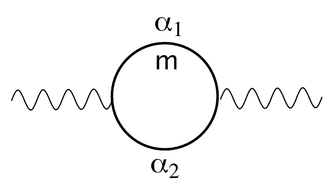
$$M(s,t)|_{t\ll m^2} = M^{\mathrm{Sub}} + \int_{M_s^2}^{\infty} dM^2\, \frac{\rho_s(M^2)}{s-M^2} + \int_{M_u^2}^{\infty} dM^2\, \frac{\rho_u(M^2)}{u-M^2} \tag{B.22}$$

where M^{Sub} is the subtraction terms reproducing with boundary behaviour of $M(s,t)$ as $s\to\infty$, and M_s^2, M_u^2 are the leading thresholds in the s and u- channel. In other words, near the forward limit, the analytic behaviour of the amplitude takes the form



Note that we can say that the loop integral can be represented as a (continuous) sum of tree-exchanges. We will see in generality that this representation follows directly from the Schwinger parameter representation.

We will illustrate the ideas of the general proof by working through the example of the 1-loop box in $D=4$. But just as an initial warm up, we can consider the bubble in $D=2$



,

where the parametric representation is

$$I(s) = \int \frac{d\alpha_1 d\alpha_2}{\mathrm{GL}(1)} \frac{1}{(-s)\alpha_1\alpha_2 + m^2(\alpha_1+\alpha_2)^2}\,. \tag{B.23}$$

The important point is that this is manifestly a (continuous) sum over simple poles in s - that is the dispersive representation! More formally, we can write:

$$I(s) = \int dM^2 \frac{\rho(M^2)}{-s+M^2} \tag{B.24}$$

where

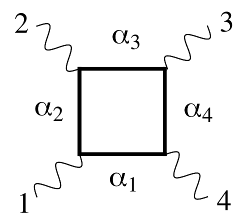
$$\rho(M^2) = \int \frac{d\alpha_1 d\alpha_2}{\mathrm{GL}(1)} \frac{1}{\alpha_1\alpha_2} \delta\left(M^2 - \frac{m^2(\alpha_1+\alpha_2)^2}{\alpha_1\alpha_2}\right)$$
$$= \int \frac{d\alpha_1}{\alpha_1} \delta\left(M^2 - \frac{m^2(\alpha_1+1)^2}{\alpha_1}\right). \tag{B.25}$$

Since the α_is are integrated over R^+, $min\ \frac{(1+\alpha_1)^2}{\alpha_1} = 4$, and thus $\rho(M^2) = 0$ when $M^2 < 4m^2$. For $M^2 > 4m^2$ the integral is localized by the delta function and one has:

$$\rho(M^2) = \frac{2}{\sqrt{M^2(M^2-4m^2)}}\Theta(M^2-4m^2)\,. \tag{B.26}$$

This manifests the position of the branch point at $s = 4m^2$.

We now turn to the $D = 4$



.

We will see that

$$I(s,t) = \int_{4m^2}^{\infty} dM^2\, \frac{\rho(M^2,t)}{-s+M^2} \tag{B.27}$$

where $\rho(M^2,t)$ is analytic in t around $t = 0$, with a cut at large positive $t \sim m^2$, but finite for $t < 0$. Note that importantly the starting point of the integral is at $4m^2$ which is independent of t. If this had then say $4m^2 - t$, then we would not have an analytic expression in t. Now let's look at the the box integral in Schwinger parameter space:

$$I(s,t) = \int \frac{d\alpha_1 \cdots d\alpha_4}{\mathrm{GL}(1)} \frac{1}{(\underbrace{(-s)\alpha_1\alpha_3 + (-t)\alpha_2\alpha_4 + m^2(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^2}_{\Delta})^2} \tag{B.28}$$

We begin in the Euclidean regime where $-s, -t > 0$, the denominator Δ is positive, and the integral is perfectly analytic. In fact, even if $(-s)$ and $(-t)$ are negative, as long as they are small with respect to m^2 we are fine, since Δ can be rewritten as

$$\Delta = (4m^2-s)\alpha_1\alpha_3 + (4m^2-t)\alpha_2\alpha_4 + m^2\left((\alpha_1-\alpha_3)^2 + (\alpha_2-\alpha_4)^2 + 2(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)\right). \tag{B.29}$$

Now let's keep t fixed and small but increase s. Clearly $\Delta > 0$ for any $s < 4m^2$. But note that for **any** positive ϵ, we can make $\Delta < 0$ at $s = 4m^2 + \epsilon$. Naively one might worry about $(-t)$ being positive, but simply by considering the limit $(\alpha_1, \alpha_3) \to \infty$ while

(α_2, α_4) held fixed, we can make $\Delta < 0$ for any value of positive ϵ. So, we see that we hit a branch point singularity at $s = 4m^2$ independent to the value of t.

Now let's first get the dispersive representation starting in the forward limit $t \to 0$. Fixing the GL(1) symmetry by setting $\alpha_1 = 1$, we have

$$\begin{aligned} I(s, t{=}0) &= \int d\alpha_2 d\alpha_3 d\alpha_4 \frac{1}{(\underbrace{(-s)\alpha_3 + m^2(1+\alpha_2+\alpha_3+\alpha_4)^2}_{\Delta})^2} \\ &= \int dM^2 \frac{\tilde{\rho}(M^2)}{(M^2 - s)^2} \end{aligned} \tag{B.30}$$

where

$$\tilde{\rho}(M^2) = \int d\alpha_2 d\alpha_3 d\alpha_4 \frac{1}{\alpha_3^2} \delta\left(M^2 - \frac{m^2(1+\alpha_2+\alpha_3+\alpha_4)^2}{\alpha_3}\right). \tag{B.31}$$

Note that since the minimum of $(1+\alpha_2+\alpha_3+\alpha_4)^2/\alpha_3$ is at 4, $\tilde{\rho}(M^2)$ will vanish when $M < 4m^2$ so we have $I(s, t{=}0) = \int_{4m^2}^{\infty} dM^2 \frac{\tilde{\rho}(M^2)}{(M^2-s)^2}$. Integrating by parts, we have

$$I(s, t{=}0) = -\int_{4m^2}^{\infty} dM^2 \frac{\partial}{\partial M^2} \frac{\tilde{\rho}(M^2)}{(M^2 - s)} + \int_{4m^2}^{\infty} dM^2 \frac{1}{(M^2 - s)} \frac{\partial}{\partial M^2} \tilde{\rho}(M^2) \tag{B.32}$$

The boundary term at $M^2 = \infty$ vanishes. Importantly, for $M^2 = 4m^2$, $\tilde{\rho}(4m^2)$ itself also vanishes. This can be explicitly confirmed, but it must be: if $\tilde{\rho}(M^2 \to 4m^2) = const.$, then the integral near

$$\int_{4m^2} dM^2 \frac{1}{(M^2 - s)^2} \sim \frac{1}{4m^2 - s} \tag{B.33}$$

gives a pole in $s = 4m^2$, while we can see easily that one can at most get a branch cut there. Let us explicitly compute $\tilde{\rho}(M^2)$:

$$\tilde{\rho}(M^2) = \int d\alpha_2 d\alpha_3 d\alpha_4 \, \frac{1}{m^2 \alpha_3} \delta\left((\alpha_3 - \alpha^+)(\alpha_3 - \alpha^-)\right) , \tag{B.34}$$

where $\alpha^{\pm} = -(1+\alpha_2+\alpha_4) + \frac{x}{2}(1 \pm \sqrt{1 - \frac{4}{x}(1+\alpha_2+\alpha_4)})$ and $x = \frac{M^2}{m^2}$. We use the delta functions to localize α_3, while the integration over α_2 and α_4 is bounded by $1+\alpha_2+\alpha_4 \leq \frac{x}{4}$ to ensure that $\alpha^{\pm}$ stays real. In the end we find:

$$\tilde{\rho}(M^2) = \frac{\log(1 + \sqrt{1 - \frac{4}{x}}) - \log(1 - \sqrt{1 - \frac{4}{x}}) - 2\sqrt{1 - \frac{4}{x}}}{m^2} , \tag{B.35}$$

which indeed vanishes when $M^2 = 4m^2$. Substituting the result back into eq.(B.32), we find

$$I(s, t{=}0) = \int_{4m^2}^{\infty} dM^2 \frac{\rho(M^2)}{(M^2 - s)} \quad \rho(M^2) = \frac{\sqrt{1 - \frac{4}{x}}}{M^2 m^2} . \tag{B.36}$$

We can proceed in the same way to compute the t-expansion. We simply Taylor expand eq.(B.28) where we have:

$$\sum_q (-t)^q \int dM^2 \, \frac{\tilde{\rho}^{(q)}(M^2)}{(M^2-s)^{q+2}} \tilde{\rho}^{(q)}(M^2) = \int d\alpha_2 d\alpha_3 d\alpha_4 \frac{(\alpha_2\alpha_4)^q}{\alpha_3^{2+q}} \delta\left(M^2 - \frac{m^2(1+\alpha_2+\alpha_3+\alpha_4)^2}{\alpha_3}\right). \tag{B.37}$$

Again the α_3 integral localizes and we are restricted to $\alpha_2+\alpha_4 < \frac{M^2}{4m^2} - 1$. Note that this shows that due to the $(\alpha_2\alpha_4)^q$ factor, $\tilde{\rho}^{(q)}(M^2)$ and all q of its derivatives with respect to M^2 vanishes at $M^2 \to 4m^2$. Thus we can write the coefficient of $(-t)^q$ as

$$\int_{4m^2}^{\infty} dM^2 \, \frac{\rho^{(q)}(M^2)}{(M^2-s)}, \quad \rho^{(q)}(M^2) = \frac{\partial^{q+1}}{\partial(M^2)^{q+1}} \tilde{\rho}^{(q)}(M^2). \tag{B.38}$$

This leads to the dispersive representation for the box integral around $t = 0$:

$$I(s,t) = \int_{4m^2}^{\infty} dM^2 \frac{\sum_q (-t)^q \rho^{(q)}(M^2)}{(M^2-s)}. \tag{B.39}$$

As an example we can explicitly compute $\rho^{(1)}(M^2)$. Starting with:

$$\tilde{\rho}^{(1)}(M^2) = \frac{3(M^2+6m^2)\left(\log(1+\sqrt{1-\frac{4}{x}}) - \log(1-\sqrt{1-\frac{4}{x}})\right) - (11M^2+16m^2)\sqrt{1-\frac{4}{x}}}{18m^4}, \tag{B.40}$$

we find:

$$\rho^{(1)}(M^2) = \frac{\partial^2}{\partial(M^2)^2} \tilde{\rho}^{(1)}(M^2) = \frac{(1-\frac{4}{x})\sqrt{1-\frac{4}{x}}}{6M^2m^4}. \tag{B.41}$$

Finally, we note that due to the increase in M^2 derivatives, $\rho^{(q)}(M^2)$ are increasingly suppressed for larger q as $M^2 \to \infty$. We will come back to this point when we study the partial wave expansion of the numerator in eq.(B.39).

Having seen all the relevant ideas in the 1-loop examples, let's now consider the general story. Consider any integral associated with a graph G, as far as the analytic structure is concerned we can just take scalar graphs with numerator = 1. The integral in general takes the form

$$I = \Gamma\left(E - LD/2\right) \int \frac{d^E\alpha}{GL(1)} \frac{1}{\mathcal{U}^{D/2}} \left(\frac{\mathcal{U}}{\mathcal{F}}\right)^{E-LD/2}, \tag{B.42}$$

where $\mathcal{U}$, $\mathcal{F}$ are the Symanzik polynomials given as

$$\mathcal{U} = \sum_{\substack{T \in spanning \\ tree}} \left(\prod_{i \notin T} \alpha_i\right), \quad \mathcal{F} = \mathcal{F}^0 + (\sum_i m_i^2 \alpha_i)\mathcal{U}$$

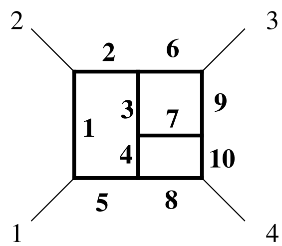
$$\mathcal{F}^0 = \sum_{\substack{T_2 \in spanning \\ 2-tree}} \left(\prod_{i \notin T_2} \alpha_i\right) (\sum_{j \in L} p_j)^2 \tag{B.43}$$

In particular all the dependence on the external Mandelstams is in the $\mathcal{F}$-polynomial. Specializing to four-points, we have that

$$\mathcal{F} = (-s)\mathcal{F}_s^0 + (-t)\mathcal{F}_t^0 + (-u)\mathcal{F}_u^0 + (\sum_i m_i^2 \alpha_i)\mathcal{U} \tag{B.44}$$

Note that every 2-tree that contributes to $\mathcal{F}^0$ must appear in $\mathcal{U}$, so every monomial in $\mathcal{F}^0$ also occurs in $(\sum_i m_i^2 \alpha_i)\mathcal{U}$; this makes it manifest that $\mathcal{F} > 0$, so long as $(-s)$, $(-t)$, $(-u)$ are small enough.

Now we'd like to show that, at fixed t, we have some branch point singularity at $s \to M_s^2$ (independent of t), and $u \to M_u^2$ (again independent of t). Of course at general loops there can be many "thresholds", but one of them will occur at smallest s; for example



we can have thresholds at $s = (m_2 + m_5)^2$ or $(m_6 + m_7 + m_8)^2$. We can systematically identify these as follows. Pick any monomial $m^{(s)}$ in $\mathcal{F}_s^0$, since these monomials do not appear in $\mathcal{F}_t^0$ or $\mathcal{F}_u^0$, they will dominate if we scale those αs $\to \infty$. So for each monomial we will have some threshold $M^2_{m^{(s)}}$. The minimum of those over all monomials $m^{(s)}$ is some $m^{*(s)}$, and the branch point is at $M_s^2 \equiv M^2_{m^{*(s)}}$. Similarly for M_u^2. Furthermore, for any $\epsilon > 0$, by scaling all of the αs in $m^{*(s)}$ to infinity, we see that we can always make $\mathcal{F} < 0$ for $s = M_s^2 + \epsilon$, so the branch point sits at $s = M_s^2$ independent of t, and similarly for M_u^2.

Now in general the four-point loop integral takes the form

$$\begin{aligned} I(s,t) &= \int \frac{d^E\alpha}{GL(1)} \frac{1}{\mathcal{U}^a} \frac{1}{((-s)\mathcal{F}_s^0 + (-t)\mathcal{F}_t^0 + (-u)\mathcal{F}_u^0 + (\sum_i m_i^2 \alpha_i)\mathcal{U})^b} \\ &= \int \frac{d^E\alpha}{GL(1)} \frac{1}{\mathcal{U}^a} \frac{1}{[(-s)(\mathcal{F}_s^0 - \mathcal{F}_u^0) + (-t)(\mathcal{F}_t^0 - \mathcal{F}_u^0) + (\sum_i m_i^2 \alpha_i)\mathcal{U})^b} \\ &= \int dM^2 \frac{\sum_q (-t)^q \tilde{\rho}^{(q)}(M^2)}{(M^2 - s)^b}, \end{aligned} \tag{B.45}$$

where

$$\tilde{\rho}^{(q)}(M^2) = \int \frac{d^E\alpha}{GL(1)} \frac{1}{\mathcal{U}^a} \frac{(\mathcal{F}_t^0 - \mathcal{F}_u^0)^q}{(\mathcal{F}_s^0 - \mathcal{F}_u^0)^{b+q}} \delta\left(M^2 - \frac{(\sum_i m_i^2 \alpha_i)\mathcal{U}}{(\mathcal{F}_s^0 - \mathcal{F}_u^0)}\right). \tag{B.46}$$

Now, the point is again that the δ function constraint forces either that for $M^2 > 0$, $M^2 > M_s^2$, or for $M^2 < 0$, that $M^2 < -M_u^2 - t$, so that we can write

$$I(s,t) \;=\; \int_{M_s^2}^{\infty} dM^2 \frac{\sum_q (-t)^q \tilde{\rho}_s^{(q)}(M^2)}{(M^2-s)^b} + \int_{M_u^2}^{\infty} dM^2 \frac{\sum_q (-t)^q \tilde{\rho}_u^{(q)}(M^2)}{(M^2-u)^b}\,. \tag{B.47}$$

By the same integration by parts idea, we arrive at our final form:

$$I(s,t) = \int_{M_s^2}^{\infty} dM^2 \frac{\rho_s(M^2,t)}{(M^2-s)} + \int_{M_u^2}^{\infty} dM^2 \frac{\rho_u(M^2,t)}{(M^2-u)}\,. \tag{B.48}$$

B.3 Partial wave expansion of unitarity cuts

As stressed in the main text, near the forward limit the singularities of the four-point amplitude are associated with threshold productions. We would like to demonstrate that contributions from these singularities, which are the imaginary part of the amplitude on the real s-axes, is given by a positive expansion on the Gegenbauer polynomial. We begin by considering scalar scattering in the C.O.M frame, with the spatial momenta of the incoming and out going particles given by $\hat{p}_{in} = p_1 - p_2$ and $\hat{p}_{out} = p_3 - p_4$ respectively, which span a $D-1$-dimensional space. As the singularites are associated with threshold production, in the C.O.M frame these are all single or multi-particle states forming irreducible representations under SO($D-1$). To this end, let us first build up general irreps of SO($n+1$), latter identifying $n = D-2$.

For a system with rotational SO($n+1$) symmetry, it is useful to consider operators as matrix elements on the Hilbert space of states that form irreducible representations of SO($n+1$). To this end, we introduce $n+1$-dimensional unit vectors x, i.e. points on an n-sphere. The states in the Hilbert space will be functions of these vectors, in particular we have states $|x\rangle$ equipped with the inner product $\langle x|y\rangle = \delta(x,y)$. To integrate these functions, we introduce the SO($n+1$) invariant measure $\langle x d^n x\rangle \equiv \frac{1}{\Omega_n}\varepsilon(x dx\cdots dx)$, where it is normalized with the solid angle Ω_n.

Now we will like to construct states that transforms as irreps under SO($n+1$), i.e. they transform linearly. To draw an analogy, consider the state labeled by coordinate X, $|X\rangle$. Under translations T_a, it transforms non-linearly, $T_a|X\rangle = |X+a\rangle$. For *linear* representations, we know we can define the Fourier transformed state $|k\rangle$ which transforms under translation as:

$$|k\rangle = \int dX e^{ikX}|X\rangle \to T_a|k\rangle = e^{-ika}|k\rangle\,. \tag{B.49}$$

We would like a similar representation for SO(n+1). Now clearly the state

$$| \rangle = \int \langle x d^n x\rangle \, |x\rangle \tag{B.50}$$

is invariant as $|x\rangle \to |Rx\rangle$, where R is a SO(n+1) rotation, while

$$|i\rangle = \int \langle x d^n x\rangle \; x^i |x\rangle \tag{B.51}$$

transforms as a vector. For $|ij\rangle$ we cannot simply use $\int \langle x d^n x\rangle \;\; x^i x^j |x\rangle$ since it is not reducible and contains a trace piece. This tells us that we should use $|ij\rangle = \int \langle x d^n x\rangle \;\left(x^i x^j - \frac{\delta^{ij}}{n+1}\right) |x\rangle$. Going onward it is clear that this is the same task we've encountered previously in deriving the Gegenbauer polynomial from tree-exchanges. Borrowing from that experience, we see that the irreducible states can be simply generated by expanding:

$$\int \frac{\langle x d^n x\rangle}{|x-y|^{n-1}} |x\rangle = \sum_\ell y^{i_1}\cdots y^{i_\ell} |i_1 \cdots i_\ell\rangle \,. \tag{B.52}$$

The states $|i_1 \cdots i_\ell\rangle$ are now irreps: symmetric traceless tensors of SO(n+1). Note that the Gegenbauer polynomials in this language is simply

$$G_\ell^{\frac{n-1}{2}}(\cos\theta) = \mathcal{A}_{n,\ell} y_{i_1}\cdots y_{i_\ell} \langle x | i_1 \cdots i_\ell\rangle, \quad \cos\theta = y\cdot x \tag{B.53}$$

where $\mathcal{A}_{n,\ell} := 2^\ell \frac{\Gamma(\ell+\frac{n-1}{2})}{\Gamma(\frac{n-1}{2})\ell!}$. The orthogonality property of Gegenbauer polynomials is then simply:

$$\begin{aligned} \int \langle z d^n z\rangle \frac{G_\ell^{\frac{n-1}{2}}(y\cdot z)}{\mathcal{A}_{n,\ell}} \frac{G_{\ell'}^{\frac{n-1}{2}}(w\cdot z)}{\mathcal{A}_{n,\ell'}} &= \int \langle z d^n z\rangle \; y_{i_1}\cdots y_{i_\ell} \langle i_1 \cdots i_\ell | z\rangle \langle z | j_1 \cdots j_{\ell'}\rangle w^{j_1}\cdots w^{j_{\ell'}} \\ &= \mathcal{B}_{n,\ell}\delta_{\ell,\ell'} \frac{G_\ell^{\frac{n-1}{2}}(y\cdot w)}{\mathcal{A}_{n,\ell}} \,, \end{aligned} \tag{B.54}$$

where we've used that the states $|i_1 \cdots i_\ell\rangle$ and $|j_1 \cdots j_{\ell'}\rangle$ are orthogonal to each other if $\ell \neq \ell'$ since they have different quantum numbers, and here $\mathcal{B}_{n,\ell} = \frac{2^{-\ell}\Gamma(n+\ell-1)\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n-1)\Gamma\left(\ell+\frac{n+1}{2}\right)}$. If we let $y = w$ and replace $\langle z d^n z\rangle$ by

$$\frac{\Omega_{n-1}}{\Omega_n} \sin^{n-2}\theta d\cos\theta \,, \tag{B.55}$$

we get the usual normalization factor for Gegenbauer polynomials:

$$\int G_\ell(\cos\theta) G_{\ell'}(\cos\theta) \sin^{n-2}\theta d\cos\theta = \mathcal{N}_{n,\ell}\delta_{\ell,\ell'} \tag{B.56}$$

with

$$\mathcal{N}_{n,\ell} = \frac{\Omega_n}{\Omega_{n-1}} \mathcal{A}_{n,\ell}\mathcal{B}_{n,\ell} = \frac{\pi 2^{2-n}\Gamma[\ell+n-1]}{\ell!\left(\ell+\frac{n-1}{2}\right)\Gamma^2\left[\frac{n-1}{2}\right]} . \tag{B.57}$$

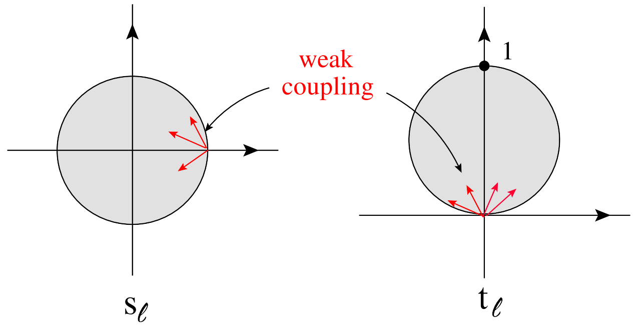


Figure B.1: The region allowed for $\mathbf{s}_\ell$ and $\mathbf{t}_\ell$ by unitarity. At weak coupling this constraint is only reflected in Re$[\mathbf{s}_\ell] \leq 1$ and Im$[\mathbf{t}_\ell] \geq 0$.

The orthogonality relation also implies that:

$$\begin{aligned}\langle x|i_1\cdots i_\ell\rangle\langle i_1\cdots i_\ell|y\rangle &= \mathcal{B}^{-1}_{n,\ell}\int\langle zd^nz\rangle\left(\langle x|i_1\cdots i_\ell\rangle z^{i_1}\cdots z^{i_\ell}\right)\left(z^{j_1}\cdots z^{j_\ell}\langle j_1\cdots j_\ell|y\rangle\right)\\ &= \mathcal{B}^{-1}_{n,\ell}\int\langle zd^nz\rangle x^{i_1}\cdots x^{i_\ell}\langle i_1\cdots i_\ell|z\rangle\langle z|j_1\cdots j_\ell\rangle y^{j_1}\cdots y^{j_\ell}\\ &= \mathcal{A}^{-1}_{n,\ell}G_\ell^{\frac{n-1}{2}}(x\cdot y)\,,\end{aligned} \tag{B.58}$$

where the first equality holds since the SO(n+1) invariant integration of $z^{i_1}\cdots z^{j_\ell}$ yields a polynomial of products of Kronecker deltas, and when acting on the irreps, only i, j contractions yield contributions as any trace pieces vanish.

Finally, these irreducible states also provide a basis for operators. A general operator can be expanded as:

$$\mathcal{O} = \sum \mathcal{O}^{i_1\cdots i_\ell;j_1\cdots j_{\ell'}}|i_1\cdots i_\ell\rangle\langle j_1\cdots j_{\ell'}|\,. \tag{B.59}$$

However, for SO(n+1) invariant ones, the operator $\mathcal{O}^{i_1\cdots i_\ell;j_1\cdots j_{\ell'}}$ can only be comprised of Kronecker deltas and since $\delta_{i_a i_b}$ contracted with the states $|i_1\cdots i_\ell\rangle$ vanishes, it can only be polynomials of $\delta_{i_a j_b}$. This tells us that $\ell = \ell'$, i.e. it is diagonal in spin space. In the last equality we've used eq.(B.54). Thus we conclude that SO(n+1) invariant operators can be written as

$$\langle y|\mathcal{O}^{Inv}|x\rangle = \sum_\ell \mathcal{N}_{n,\ell}\mathsf{p}_\ell G_\ell^{\frac{n-1}{2}}(x\cdot y)\,, \tag{B.60}$$

i.e. it is expandable on the Gegenbauer polynomials.

Now let's consider $\mathbb{S}$, the s-matrix of the full theory. Restricting ourselves to the $2\to 2$ elastic scattering, we can define the "little" matrix $\mathbf{s}$

$$\langle\hat{p}_{out}|\mathbf{s}|\hat{p}_{in}\rangle =_{out}\langle p_3, p_4|\mathbb{S}|p_1, p_2\rangle_{in}\,. \tag{B.61}$$

In other words $\mathbf{s}$ is only defined only on the $2 \to 2$ states. The full s-matrix satisfy $\mathbb{S}^\dagger\mathbb{S} = \mathbb{I}$, while the small s-matrix satisfy

$$\mathbf{s}^\dagger\mathbf{s} \leq \mathbb{I}\,, \tag{B.62}$$

as an operator statement, i.e. for any state $|\psi\rangle$, we have $\langle\psi|\mathbf{s}^\dagger\mathbf{s}|\psi\rangle \leq \langle\psi|\psi\rangle$. Now since $\mathbf{s}$ is rotationally invariant, we can write

$$\mathbf{s} = \sum_\ell \; \mathbf{s}_\ell \, |i_1 i_2 \cdots i_\ell\rangle\langle i_1 i_2 \cdots i_\ell|\,, \tag{B.63}$$

then $\mathbf{s}^\dagger\mathbf{s} \leq 1$ implies $|\mathbf{s}_\ell| \leq 1$. If we write $\mathbf{s} = 1 + i\mathbf{t}$, then this implies $|1 + i\mathbf{t}_\ell| \leq 1$. Note that

$$\mathbf{t} = \sum_\ell \; \mathbf{t}_\ell \, |i_1 i_2 \cdots i_\ell\rangle\langle i_1 i_2 \cdots i_\ell| \; \to \langle\hat{p}_{out}|\mathbf{t}|\hat{p}_{in}\rangle = \mathcal{N}_{n,\ell} \sum_\ell \; \mathbf{t}_\ell G_\ell^{\frac{n-1}{2}}(\hat{p}_{out} \cdot \hat{p}_{in}) \tag{B.64}$$

where $\langle\hat{p}_{out}|\mathbf{t}|\hat{p}_{in}\rangle$ is the four-point amplitude of interest. Since $|1 + i\mathbf{t}_\ell| \leq 1$,

$$1 + i(\mathbf{t}_\ell - \mathbf{t}_\ell^*) + |\mathbf{t}_\ell|^2 \leq 1 \; \to \; i(\mathbf{t}_\ell^* - \mathbf{t}_\ell) \geq |\mathbf{t}_\ell|^2\,. \tag{B.65}$$

More explicitly we have $1 + i\mathbf{t}_\ell = \eta_\ell e^{i\delta_\ell}$ with $\eta_\ell \leq 1$. Note that in a weakly coupled theory, eq.(B.65) just tells us that $i(\mathbf{t}_\ell - \mathbf{t}_\ell^*) \geq 0$, i.e. the imaginary part is positive. The full non-linear constraint is only present at strong coupling see fig.B.1. Since the imaginary part is positive, we have

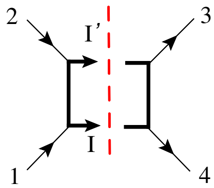
$$\mathrm{Im}[\langle\hat{p}_{out}|\mathbf{t}|\hat{p}_{out}\rangle] = \mathcal{N}_{n,\ell} \sum_\ell \; \mathrm{Im}[\mathbf{t}_\ell] G_\ell^{\frac{n-1}{2}}(\hat{p}_{out} \cdot \hat{p}_{in}) \tag{B.66}$$

i.e. the imaginary part of the amplitude is positively expandable on the Gegenbauer polynomials.

B.4 The spinning-spectral function for massive box

From appendix B.2 we've seen that near the forward limit, the four-point amplitude admits a Källén-Lehman representation representation, where the "spectral function" depends on t, i.e. $\rho(M^2, t)$. Since the spectral function is a polynomial in t near the forward limit, it has a partial wave expansion. Now from appendix B.3, we've seen that the discontinuity for $A, B \to A, B$ type scattering should be positively expandable on the Gegenbauer polynomials. Since the discontinuity in the dispersive representation is the spectral functions, we conclude that the "spinning spectral function" should be a positive function. Here we will use the massive box to demonstrate this fact.

Let us consider an explicit example, the discontinuity for the box-integral with massive internal propagators in four-dimensions. The integrand in the phase space integral is simply given by the product of two tree-propagators:



$$= \frac{1}{2(p_1 \cdot p_I)} \frac{1}{2(p_4 \cdot p_I)} = \frac{4}{s^2} \frac{1}{1 - \sqrt{1 - \frac{4m^2}{s}} \hat{p}_1 \cdot \hat{p}_I} \frac{1}{1 - \sqrt{1 - \frac{4m^2}{s}} \hat{p}_4 \cdot \hat{p}_I}, \tag{B.67}$$

where we are again considering the kinematics in center of mass frame. The discontinuity is now given as:

$$\langle \hat{p}_{in} | T^\dagger T | \hat{p}_{out} \rangle = \int_{4m^2}^{\infty} ds \frac{4J_s}{s^2} \int \langle \hat{p}_I d^2 \hat{p}_I \rangle F^*(\hat{p}_1 \cdot \hat{p}_I) F(\hat{p}_4 \cdot \hat{p}_I) \tag{B.68}$$

where $F(x) = \left(1 - \sqrt{1 - \frac{4m^2}{s}} x\right)^{-1}$, and J_s is the dimensionless Jacobian factor stemming from the phase space integral:

$$\int d^D \ell \delta\left(\ell^2 - m^2\right) \delta\left((\ell - p_{12})^2 - m^2\right) = \frac{(s - 4m^2)^{\frac{D-3}{2}}}{\sqrt{s}} \int d\Omega_{D-2}\,, \tag{B.69}$$

which for $D = 4$ is simply $J_s = \sqrt{1 - \frac{4m^2}{s}}$.

Let us write $F(x)$ as an expansion on the Gegenbauer polynomial with coefficient f_ℓ, $F(x) = \sum_\ell f_\ell(s) G_\ell^{\frac{1}{2}}(x)$. Then the two-dimensional angular integral simply reduces the corresponding product of $G_\ell^{\frac{1}{2}}(x)$s in eq.(B.68) to $\sum_\ell |f_\ell(s)|^2 \frac{2}{2\ell+1} G_\ell^{\frac{1}{2}}(\hat{p}_1 \cdot \hat{p}_4)$, where θ is precisely the scattering angle. Thus we conclude that the discontinuity is simply

$$\langle \hat{p}_{in} | T^\dagger T | \hat{p}_{out} \rangle = \int_{4m^2}^{\infty} ds \frac{4J_s}{s^2} \sum_\ell \mathsf{p}_\ell(s) \frac{2}{2\ell + 1} G_\ell^{\frac{1}{2}}(\cos\theta), \tag{B.70}$$

where $\mathsf{p}_\ell(s) \equiv |f_\ell(s)|^2$ is the positive definite "spinning" spectral function. Let us compute the $f_\ell(s)$s explicitly.

Using the generating function and the orthogonality of the Gegenbauer polynomials, we can write down the following generating function for $f_\ell(s)$,

$$\int_{-1}^{1} dx \frac{1}{(1 - ax)} \frac{1}{(1 - 2rx + r^2)^{\frac{1}{2}}} = \sum_\ell r^\ell \frac{2}{2\ell + 1} f_\ell(s)\,, \tag{B.71}$$

where $a = \sqrt{1 - \frac{4m^2}{s}}$. A straight forward integration yields for the LHS:

$$\frac{1}{ab} \log\left[\frac{(1 - r + b)}{(1 - r - b)} \frac{(1 + r - b)}{(1 + r + b)}\right], \qquad b = \sqrt{1 + r^2 - 2\frac{r}{a}} \tag{B.72}$$

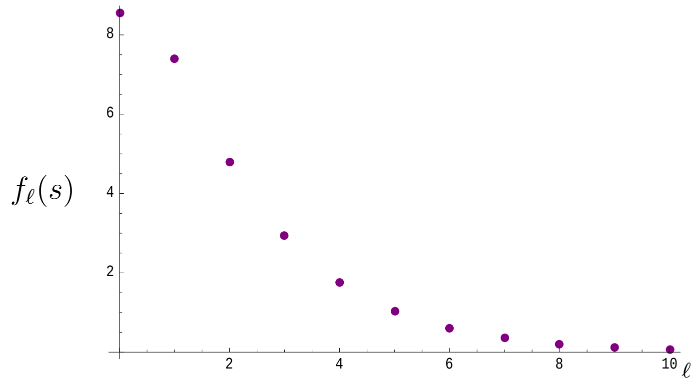


Figure B.2: We plot the coefficients $f_\ell(s)$ for $s = 14$. We see that the coefficients are suppressed for higher spins

As the generating function is non-polynomial in r, we have an infinite tower of spin in the expansion. The coefficient for the first few spins are:

$$f_0 = \frac{1}{2}\frac{\log\delta}{a}, \quad f_1 = \frac{3}{2}\frac{(-2a + \log\delta)}{a^2}, \quad f_2 = \frac{5}{2}\frac{(-6a + 3\log\delta - a^2\log\delta)}{2a^3}, \tag{B.73}$$

where $\delta = \frac{1+a}{1-a}$. Since a takes value between 0 and 1, one can straightforwardly see that the coefficient decreases for increasing spin.

Let us verify that eq.(B.70), combined with (B.71) and (B.72), indeed reproduces the correct discontinuity of eq.(3.20)

$$I_4[s,t] - I_4[s,t]|_{\beta_u \to -\beta_u} . \tag{B.74}$$

To compare, we first note that the coefficients $f_\ell(s)$ is suppressed for higher spin, see. fig B.2. Thus we should find a good approximation by truncating at $\ell = 10$. Indeed summing eq.(B.70) up to spin-10 the result matches with that of eq.(B.74) as shown in fig.(B.3), thus confirming eq.(B.70).

B.5 Positivities of the Gegenbauer matrix

The results on the total positivity of Gegenbauer polynomials follow from general theorems connecting total positivity to orthogonal polynomials with positive measure discovered in the 1960s [111]. Here, we will give elementary and explicit computations that show the positivity properties explicitly for the Gegnebauer polynomial case of immediate interest to us. For the simplest case of $d = 2$, where we just have Fourier expansion in $\cos(\theta)$, we will give an especially simple argument for positivity going back essentially to Chebyshev. We will then give a simple explicit computation of the determinants associated with the Taylor expansion of Gegenbauer polynomials, where they can explicitly be seen to be positive

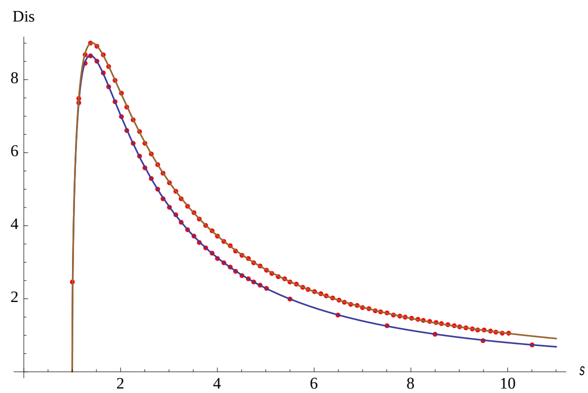


Figure B.3: We compare our Gegenbauer sum expression in eq.(B.70), truncating at $\ell = 10$, with the explicit discontinuity in eq.(B.3). We've normalized $\frac{s}{4m^2} \to s$, so that the discontinuity begins at $s = 1$ to ∞. We've compared the result of eq.(B.70), in red dots, to eq.(B.74) which is the colored curve. The brown curve is for $\cos\theta = \frac{1}{2}$, and the blue curve for $\cos\theta = 1/6$. Both exhibit perfect matching.

Total positivity of Chebyshev matrix

Let us consider a general strategy in proving the positivity of the determinant of matrices constructed from specific functions $V_\ell(y)$. In particular, the columns of the matrix are given by evaluating the function at n distinct ordered points $y_1 < y_2 < \cdots < y_n$, i.e. $\mathbf{V}_\ell = (V_\ell(y_1), V_\ell(y_2), \cdots, V_\ell(y_n))$. Our task is to prove that for a collection of n such vectors,

$$\mathrm{D}et(\mathbf{V}_{\ell_1}, \mathbf{V}_{\ell_2}, \cdots, \mathbf{V}_{\ell_n}) = \mathrm{D}et \begin{pmatrix} V_{\ell_1}(y_1) & V_{\ell_2}(y_1) & \cdots & V_{\ell_n}(y_1) \\ V_{\ell_1}(y_2) & V_{\ell_2}(y_2) & \cdots & V_{\ell_n}(y_2) \\ \vdots & \vdots & \cdots & \vdots \\ V_{\ell_1}(y_n) & V_{\ell_2}(y_n) & \cdots & V_{\ell_n}(y_n) \end{pmatrix} > 0\,. \tag{B.75}$$

The general strategy, as also discussed in [8], is to show that the above can never be zero for any choice of distinct y_is. In other words, the sign of the determinant is fixed. Then the vanishing of the determinant implies that the column vectors are now linearly dependent, or

$$\sum_{i=1}^{n} c_i V_{\ell_i}(y_j) = 0 \ . \tag{B.76}$$

for $j = 1, 2\cdots, n$. Said in another way, the function $\sum_{i=1}^{n} c_i V_{\ell_i}(y)$ have n roots on the real axes. Thus proving the definite sign of eq.(B.75) amounts to proving that eq.(B.76) cannot have n real solutions.

Before considering Chebyshev polynomials, let's first begin with $V_\ell(y) = e^{\ell y}$. Choose a sets of n ℓ_is conveniently labelled with $\ell_1 < \ell_2 < \cdots < \ell_n$, where the goal is to show that

$$f_n(y) = \sum_{i=1}^{n} c_i e^{\ell_i y} \tag{B.77}$$

cannot have n real roots for any c_i. We will prove this by induction. First for n=1, indeed $f_1(y) = e^{\ell_1 y}$ does not have a root. Next, lets assume that there are at most $n-2$ roots for $f_{n-1}(y)$, but $f_n(y)$ has n roots. We will show that this leads to a contradiction. If $f_n(y)$ has n roots, then multiplied by $e^{-\ell_1 y}$ will not change that. That is,

$$e^{-\ell_1 y} f_n(y) = c_1 + c_2 e^{(\ell_2-\ell_1)y} + \cdots, + c_n e^{(\ell_n-\ell_1)y} \tag{B.78}$$

will also have n roots. Now the derivative of a function with n roots on the real axes must have at least $n-1$ real roots. Taking the derivative we find,

$$\left(e^{-\ell_1 y} f_n(y)\right)' = c_2(\ell_2-\ell_1) e^{(\ell_2-\ell_1)y} + \cdots, + c_n(\ell_n-\ell_1) e^{(\ell_n-\ell_1)y}\,. \tag{B.79}$$

But this is nothing but f_{n-1} with another set of ordered ℓ_i, which now has $n-1$ real roots, a contradiction to our initial assumption! Thus we conclude that $f_n(y)$ cannot have n-roots and the determinant in eq.(B.75) can never be zero. Note that if one replaces $y = \log x$, then the functions we are considering are simply moments x^ℓ. As we assume that y is real, we have $x > 0$ and thus the positivity of eq.(B.75) also leads to the total positivity of the Vandermonde matrix for half moment curves.

We are interested in the Chebyshev polynomials $\cos \ell y$. Since we will be interested in cases where $\cos y > 1$, y is purely imaginary and the Chebyshev polynomial becomes $\cosh \ell y$ with y being real. Now we want to show that

$$\sum_{i=1}^{n} c_i \cosh \ell_i y = 0 \tag{B.80}$$

cannot have $2n$ real roots (or n positive roots since its a even function). But we've already shown that any linear combination of $2n$ distinct $e^{\ell y}$ cannot have $2n$ roots—thus a contradiction! This therefore proves that

$$\mathrm{D}et \begin{pmatrix} \cosh \ell_1 y_1 & \cosh \ell_2 y_1 & \cdots & \cosh \ell_n y_1 \\ \cosh \ell_1 y_2 & \cosh \ell_2 y_2 & \cdots & \cosh \ell_n y_2 \\ \vdots & \vdots & \cdots & \vdots \\ \cosh \ell_1 y_n & \cosh \ell_2 y_n & \cdots & \cosh \ell_n y_n \end{pmatrix} \neq 0. \tag{B.81}$$

i.e. it has a definite sign. Finally since all that we assumed for our Chebyshev matrix is that the spin is ordered, the minors of a given matrix obviously satisfies the same criteria, and hence we conclude that the Chebyshev matrix is a totally positive matrix.

Positivity of the Taylor scheme Gegenbauer matrix

Here we analytically prove that the determinant of the Gegenbauer matrix in the derivative scheme. Starting with the Taylor coefficients defined in eq.(3.127), first we reorganize the analytic expression as:

$$v^{\mathrm{D}}_{\ell,q} = \frac{1}{q!(\ell-q)!}\frac{(\Delta)_{\ell+q}}{\prod_{a=1}^{q}(\Delta+2a-1)} = \frac{(\Delta)_\ell}{(q!)(\ell!)}\frac{1}{\prod_{a=1}^{q}(\Delta+2a-1)}\left[(\ell)_{-q}(\ell+\Delta)_q\right], \tag{B.82}$$

where $\Delta = D{-}3$, $(a)_{-q} = a(a-1)\cdots(a-q+1)$ and $(a)_0 = 1$. Now consider the determinant of $n{+}1$ Taylor vectors. Due to our rearrangement, the determinant can be written in a factorized form:

$$\begin{aligned} \mathrm{D}et\begin{bmatrix} v^{\mathrm{D}}_{\ell_1,0} & v^{\mathrm{D}}_{\ell_2,0} & \cdots \\ v^{\mathrm{D}}_{\ell_1,1} & v^{\mathrm{D}}_{\ell_2,1} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix} &= \left(\prod_{i=1}^{n+1}\frac{(\Delta)_{\ell_i}}{\ell_i!}\frac{1}{\prod_{a=1}^{i-1}(\Delta+2a-1)a!}\right) \\ &\times \mathrm{D}et\begin{pmatrix} (\ell_1)_0(\ell_1+\Delta)_0 & (\ell_1)_{-1}(\ell_1+\Delta)_1 & ... \\ (\ell_2)_0(\ell_2+\Delta)_0 & (\ell_2)_{-1}(\ell_2+\Delta)_1 & ... \\ ... & ... & ... \end{pmatrix}. \end{aligned} \tag{B.83}$$

Now we know that the remaining determinant must have the factor $\prod_{i<j}(\ell_j-\ell_i)$ since the result vanishes if $\ell_i = \ell_j$. Furthermore, using

$$(-a)_b = (-a)(-a+1)...(-a+b-1) = (-1)^b(a)_{-b}\,, \tag{B.84}$$

we can see that the remaining determinant is invariant under $\ell \to -\ell-\Delta$. This together with power counting leads to

$$\mathrm{D}et\begin{pmatrix} (\ell_1)_0(\ell_1+\Delta)_0 & (\ell_1)_{-1}(\ell_1+\Delta)_1 & ... \\ (\ell_2)_0(\ell_2+\Delta)_0 & (\ell_2)_{-1}(\ell_2+\Delta)_1 & ... \\ ... & ... & ... \end{pmatrix} = \prod_{i<j}(\ell_j-\ell_i)(\Delta+\ell_j+\ell_i). \tag{B.85}$$

Thus we find that

$$(\prod_i v^{\mathrm{D}}_{\ell_i,\sigma_i})\epsilon^{\sigma_1\sigma_2\cdots} = \left(\prod_{i=1}^{n+1}\frac{(\Delta)_{\ell_i}}{\ell_i!}\frac{1}{\prod_{a=1}^{i-1}(\Delta+2a-1)a!}\right)\prod_{i<j}(\ell_j-\ell_i)(\Delta+\ell_j+\ell_i). \tag{B.86}$$

As one can see, the result is positive so long as $\ell_1 < \ell_2 < \cdots < \ell_{n+1}$!

B.6 The true boundary of the $\mathbb{P}^1$ EFT-hedron

The EFT-hedron constraint relies on two aspects, the wall $\vec{\mathcal{W}}_I$ and the resulting deformation parameters $\{\alpha_i\}$. Let us consider dotting $\vec{a}$ in to some wall $\mathcal{W} = (-w, 1)$, then the RHS of eq.(3.228) then tells us that:

$$\begin{pmatrix} \vec{a}_2 \cdot \mathcal{W} \\ \vec{a}_4 \cdot \mathcal{W} \\ \vec{a}_6 \cdot \mathcal{W} \end{pmatrix} = \begin{pmatrix} a_2(\beta_2 - w) \\ a_4(\beta_4 - w) \\ a_6(\beta_6 - w) \end{pmatrix} = \sum_a \mathsf{p}_a \begin{pmatrix} (u^{(2)}_{\ell_a} - w) \\ (u^{(4)}_{\ell_a} - w)y_a \\ (u^{(6)}_{\ell_a} - w)y_a^2 \end{pmatrix}, \tag{B.87}$$

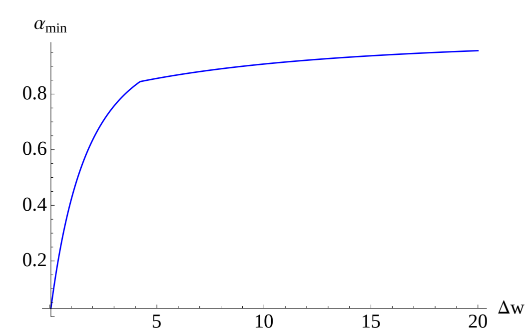
where we absorbed factors of x_a into p_a, and $y_a = x_a^2$. We see that the inner product lives in the hull of multiple deformed curves. To ensure that the hull is non-trivial, we would like to ensure all entires of the deformed moment curves to be non-negative. In other words we want $\mathcal{W}$ to satisfy

$$(u^{(k)}_\ell - w) > 0, \quad \forall \ell. \tag{B.88}$$

As the minimum of $u^{(k)}_\ell$ listed in eq.(3.229) is $-\frac{21}{4}$, we write $w = -\frac{21}{4} - \Delta w$ with $\Delta w \geq 0$. Since we have a collection of deformed curves, the constraint for $\vec{a}_k \cdot \mathcal{W}$ should be derived from a curve that encapsulate all the other curves. i.e. the master moment curve. In other words, we want to find $(1, x, \alpha x^2)$ such that its convex hull contains all the individual moment curves, or,

$$\frac{(u^{(2)}_\ell - w)(u^{(6)}_\ell - w)}{\alpha} - (u^{(4)}_\ell - w)^2 \geq 0, \quad \forall \ell\,. \tag{B.89}$$

This tells us that there is an upper bound for α, corresponding to the the minimum of $\frac{(u^{(6)}_\ell - w)(u^{(2)}_\ell - w)}{(u^{(4)}_\ell - w)^2}$, which we denote as $\alpha_{min}[\Delta w]$, reflecting the fact that it is a function of Δw. Explicitly plotting $\alpha_{min}[\Delta w]$ we find:



We see that α_{min} rises approximately linear with Δw up to around $\Delta w \sim 5$, after which $\alpha_{min} \sim 1$ for all ℓ.

Equipped with $\alpha_{min}[\Delta w]$ we can now write down the non-linear constraint for $\vec{a}_k \cdot \mathcal{W}$:

$$(\vec{a}_2 \cdot \mathcal{W})(\vec{a}_6 \cdot \mathcal{W}) - \alpha_{min}[\Delta w](\vec{a}_4 \cdot \mathcal{W})^2 > 0 \tag{B.90}$$

It is important to see if above gives constraints that go beyond those in eq.(3.230). To this end we write $\beta_2 = -\frac{3}{4} + \hat{\beta}_2$, $\beta_4 = -\frac{3}{2} + \hat{\beta}_4$ and $\beta_6 = -\frac{21}{4} + \hat{\beta}_6$, so that the original polytope bound is simply that $\hat{\beta}_i \geq 0$. In terms of these new parameters, eq.(B.90) becomes,

$$\left(\hat{\beta}_6 + \Delta w\right)\left(\hat{\beta}_2 + \frac{9}{2} + \Delta w\right) - \alpha_{min}[\Delta w]\frac{a_4^2}{a_2 a_6}\left(\hat{\beta}_4 + \frac{15}{4} + \Delta w\right)^2 \geq 0\,. \tag{B.91}$$

If the above leads to any constraint for $\hat{\beta}_i$ beyond that it is non-negative, or $\epsilon \equiv \frac{a_4^2}{a_2 a_6} < 1$ then we have found new constraints beyond eq.(3.230). For example, for $\Delta w = 0$, $\alpha_{min}[0] = 0$ and eq.(B.91) does not implement anything new.

However, for non-zero $\alpha_{min}[\Delta w]$ we will always obtain new constraints! For example, since $\hat{\beta}_4 \geq 0$, eq.(B.91) implies

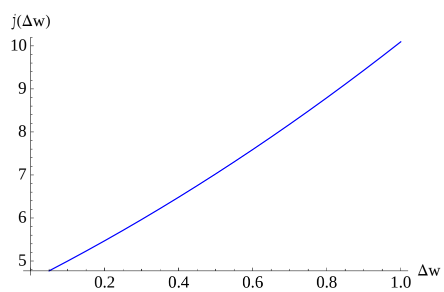
$$\boxed{\frac{\left(\hat{\beta}_6 + \Delta w\right)\left(\hat{\beta}_2 + \frac{9}{2} + \Delta w\right)}{\alpha_{min}[\Delta w]\epsilon} \geq (\frac{15}{4} + \Delta w)^2}\,, \tag{B.92}$$

and we see that $(\hat{\beta}_2, \hat{\beta}_6)$ is bounded from below. Let's set $\hat{\beta}_2, \hat{\beta}_6 = 0$, and consider

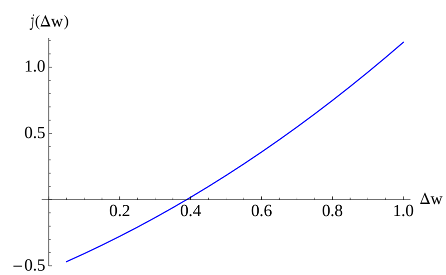
$$j(\Delta w) = \frac{\Delta w\left(\frac{9}{2} + \Delta w\right)}{\alpha_{min}[\Delta w]\epsilon} - \left(\frac{15}{4} + \Delta w\right)^2\,. \tag{B.93}$$

We have non-trivial lower bounds for $(\hat{\beta}_2, \hat{\beta}_6)$ if $j(\Delta w) < 0$. Plotting $j(\Delta w)$ for fixed ϵ with respect to Δw we find

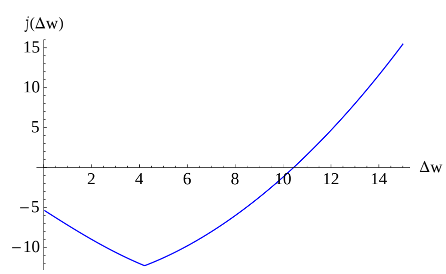
$\epsilon = 0.4$:



$\epsilon = 0.55$:



$\epsilon = 0.85$:



We see that if ϵ is above a critical value $\epsilon_c = 0.54$, there are ranges of Δw where the constraint is non-trivial. Thus we either have a non-trivial lower bound for $(\hat{\beta}_2, \hat{\beta}_6)$, or that we have an upper bound for $\epsilon < \epsilon_c$. Note that these non-trivial bounds are derived from walls that are not the walls of the original polytopes.

From eq.(B.91) one can also derive an upper bound for $\hat{\beta}_4$:

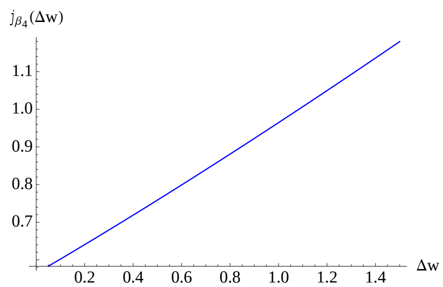
$$\boxed{\sqrt{\frac{\left(\hat{\beta}_6+\Delta w\right)\left(\hat{\beta}_2+\frac{9}{2}+\Delta w\right)}{\alpha_{min}[\Delta w]\epsilon}}-\frac{15}{4}-\Delta w \geq \hat{\beta}_4} \,. \tag{B.94}$$

Obviously, the bound is most stringent when $\hat{\beta}_6=\hat{\beta}_2=0$. Thus we consider

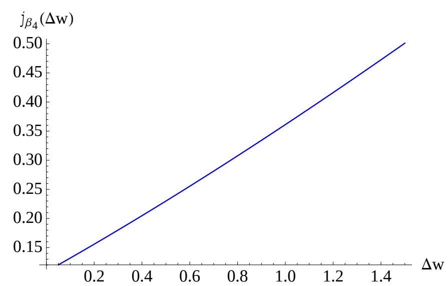
$$j_{\beta_4}(\Delta w)=\sqrt{\frac{\Delta w\left(\frac{9}{2}+\Delta w\right)}{\alpha_{min}[\Delta w]\epsilon}}-\frac{15}{4}-\Delta w \,. \tag{B.95}$$

We plot the above function with respect to Δw and look for the upper bound for $\hat{\beta}_4$ as the minimum of $j_{\beta_4}(\Delta w)$. The result depends on ϵ:

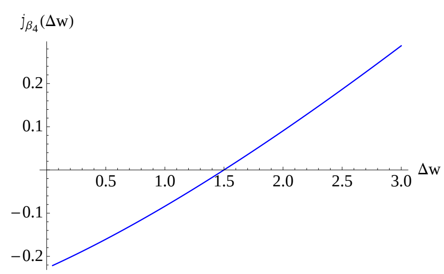
$\epsilon = 0.4$:



, $\epsilon = 0.5$:



$\epsilon = 0.6$:



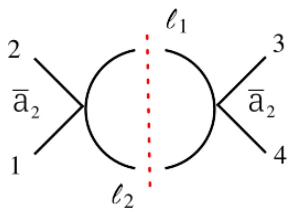
For the first two graphs we consider $\epsilon < \epsilon_c$, where no lower bounds on $(\hat{\beta}_6, \hat{\beta}_2)$ were imposed from eq.(B.92), we see that there is always an upper bound for $\hat{\beta}_4$. For $\epsilon > \epsilon_c$, we have a region of walls, $\Delta w < 15$, where there's no new bounds on $\hat{\beta}_4$, however for these cases, there are lower bounds on $\hat{\beta}_6, \hat{\beta}_2$.

In summary, we find that using walls that are "outside" the walls of $Conv[\vec{u}_{\ell,k}]$, imposes further constraint through eq.(B.91) either as a upper bound on $\hat{\beta}_4$, or lower bound on $(\hat{\beta}_2, \hat{\beta}_6)$, depending on whether ϵ is above or below ϵ_c. Thus eq.(B.92) and eq.(B.94) characterizes the $\mathbb{P}^1$ EFT-hedron.

B.7 Beta function for eq.(3.289)

Here we present the details for the computation of the beta functions from two-particle cuts in eq.(3.289).

We compute the two-particle cut by taking the product of the two tree-amplitudes parameterized in the center of mass frame as illustrated in fig.(B.4) : (θ', ϕ') is the angular dependence of the phase space for the cut propagators, and θ is the scattering angle for the external momenta. For example, for the $\bar{a}_2^2$ coupling the bubble coefficient is given by:



$$= -\frac{\bar{a}_2^2}{\Lambda^8(4\pi)} \int d\phi' d\cos\theta' \ \ (s^2+u^2+t^2)_L(s^2+u^2+t^2)_R$$

$$= -\frac{\bar{a}_2^2}{\Lambda^8} s^4 \int d\phi' d\cos\theta' \ \ F_{2,L}F_{2,R} = \frac{\bar{a}_2^2}{\Lambda^8}\frac{s^4}{60}(167+\cos 2\theta) \quad \text{(B.96)}$$

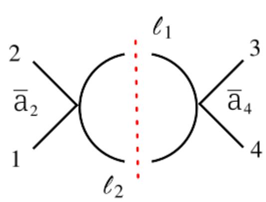
where we've defined the short hand notation:

$$\begin{aligned} F_{n,L} &= \left(1+\left(\frac{1+\cos\theta'}{2}\right)^n + \left(\frac{1-\cos\theta'}{2}\right)^n\right) \\ F_{n,R} &= \left(1+\left(\frac{1+\cos\theta'\cos\theta+\sin\theta'\cos\phi'\sin\theta}{2}\right)^n + \left(\frac{1-\cos\theta'\cos\theta-\sin\theta'\cos\phi'\sin\theta}{2}\right)^n\right). \end{aligned} \quad \text{(B.97)}$$

Changing back to Mandelstam variables we find the coefficient for the $\frac{\bar{a}_2^2}{15}(41s^2+t^2+u^2)s^2$ for the s-channel coefficient. Summing over the three channels we obtain

$$-\frac{14\bar{a}_2^2}{5\Lambda^8(4\pi)^2}(s^4+t^4+u^4)\log\frac{p^2}{\mu^2},$$

and hence $\beta_1 = \frac{14}{5(4\pi)^2}$. Similarly for s^6 we have:



$$= -\frac{\bar{a}_2\bar{a}_4 s^6}{\Lambda^{12}(4\pi)} \int d\phi' d\cos\theta' \ \ F_{2,L}F_{4,R}+F_{4,L}F_{2,R}$$

$$= \frac{2\bar{a}_2\bar{a}_4}{\Lambda^{12}35}s^6(82+\cos 2\theta) = \frac{2\bar{a}_2\bar{a}_4}{35\Lambda^{12}}s^4(83s^2+8st+8t^2)\,. \quad \text{(B.98)}$$

Again summing over all three channels we obtain

$$-\frac{2\bar{a}_2\bar{a}_4}{35\Lambda^{12}(4\pi)^2}\left(83(s^6+t^6+u^6)-24(stu)^2\right)\log\frac{p^2}{\mu^2},$$

and thus$\beta_2 = \frac{166}{35(4\pi)^2}$.

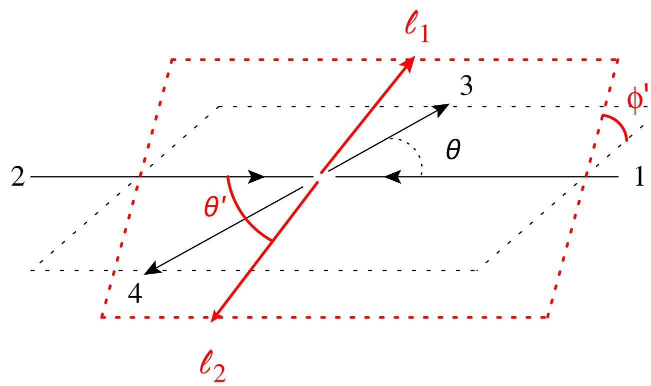


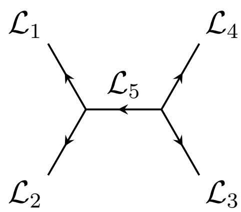
Figure B.4: We represent the internal loop momentum in the center of mass frame. The angle between the loop momentum and $\vec{p}_1 = -\vec{p}_2$ is θ', while the angle between the plane spanned by $(\vec{p}_1, \vec{\ell}_1)$ and the plane $(\vec{p}_1, \vec{p}_2)$ is ϕ'. θ is then the usual scattering angle.

Appendix C

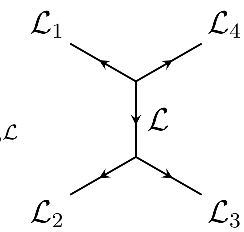
APPENDICE TO CHAPTER 4

C.1 F-symbols and tetrahedra

This appendix contains a derivation of the relation (4.3) between F-symbols and tetrahedra. On both sides of the F-move equation (5.6)

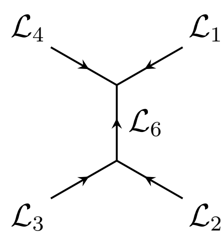


$$= \sum_{\mathcal{L}} (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\overline{\mathcal{L}}_4})_{\mathcal{L}_5,\mathcal{L}}$$



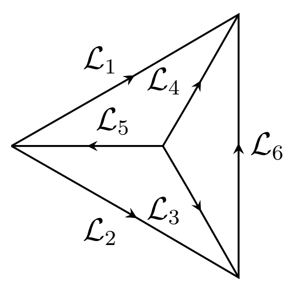
,

join



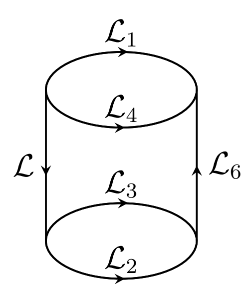
(C.1)

from the right. The resulting graph on the left side of the F-move equation can be adjusted into a tetrahedron

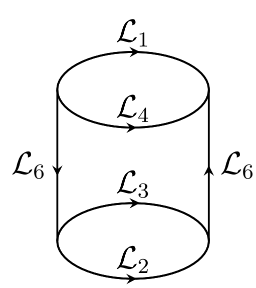


,

whereas the graph on the right side of the F-move equation can be adjusted into

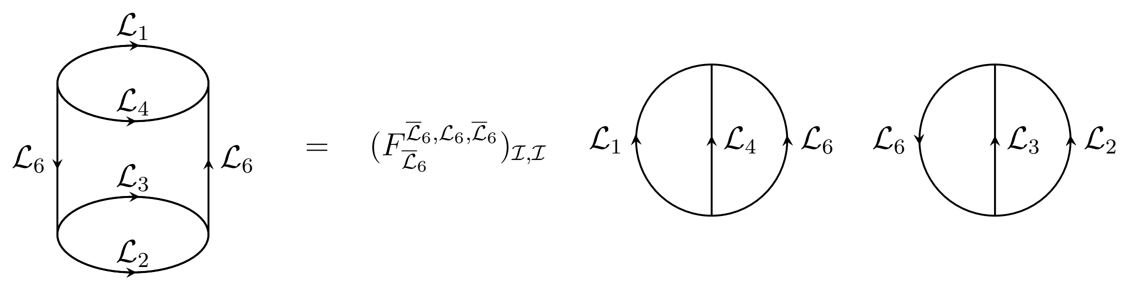


$$= \delta_{\mathcal{L},\mathcal{L}_6} \times$$



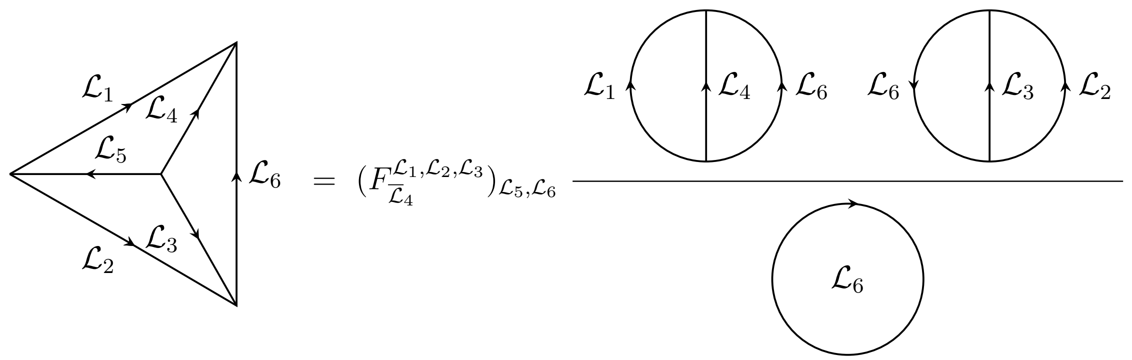
,

which vanishes if $\mathcal{L} \neq \mathcal{L}_6$ because the top and bottom loops can be shrunk but the vector space $V_{\mathcal{L},\mathcal{L}_6}$ is empty. Applying the F-move to a unit object connecting the two $\mathcal{L}_6$ edges gives

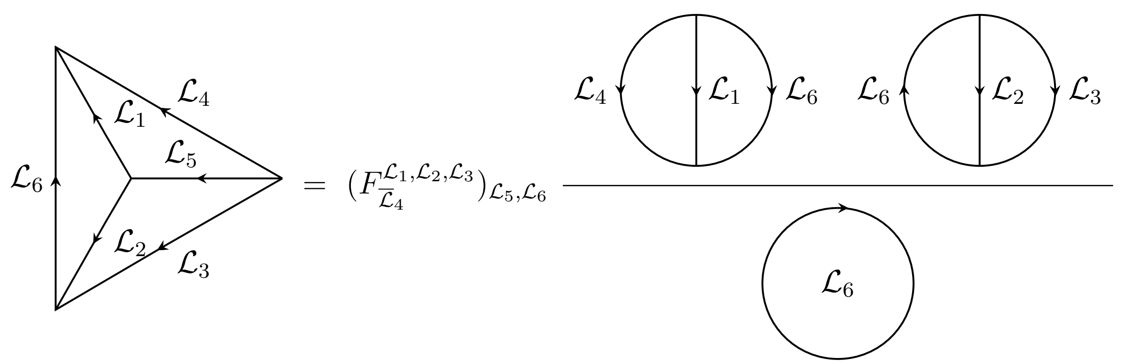


Again, no non-unit object $\mathcal{L}$ can bridge the two Θ graphs on the right because the Θ graphs can be shrunk, but the vector space $V_{\mathcal{L},\mathcal{I}}$ is empty if $\mathcal{L} \neq \mathcal{I}$.

Putting things together,



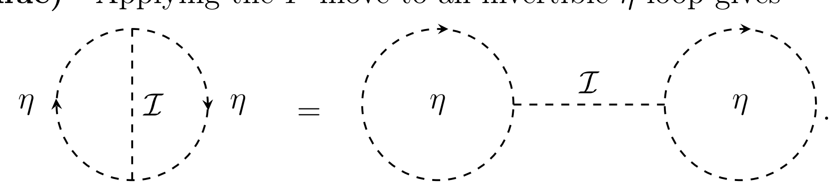
A similar derivation by joining (C.1) from the left with the F-move equation shows that



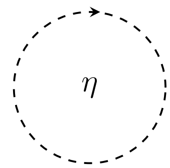
C.2 Transparent graph equivalences

Let $\mathcal{C}$ be a transparent fusion category, and η an invertible object. There are the following graph equivalences.

1. **(Loop Value)** Applying the F-move to an invertible η loop gives



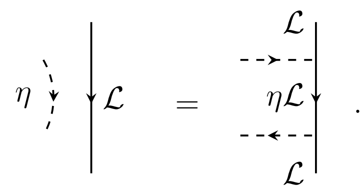
Thus



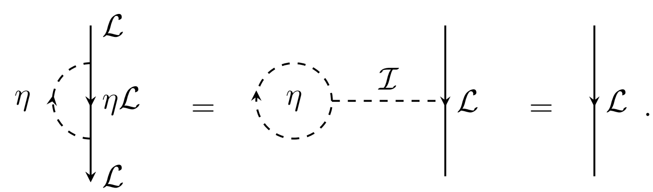
$$= 1\,, \tag{C.2}$$

i.e. invertible loops have value 1.

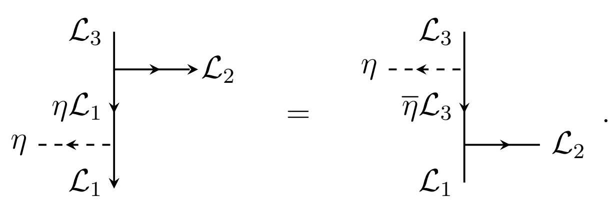
2. **(Attachment)** An invertible object can be attached to a simple object $\mathcal{L}$



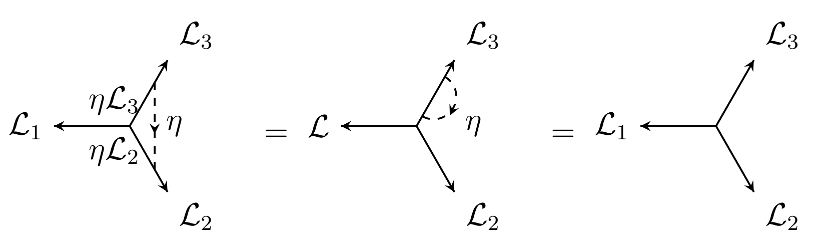
3. **(Detachment)** An invertible object with two ends attached to a non-invertible simple object $\mathcal{L}$ can be detached



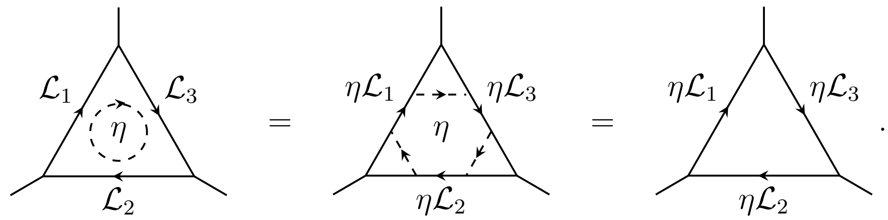
4. **(Swap)** An invertible object attached to an edge can be swapped across a trivalent vertex



5. **(Contraction)** An invertible object bridged across a trivalent vertex can be contracted. It can be regarded as a swap followed by a detachment



6. **(Symmetry nucleation)** Given a graph, an invertible loop can be nucleated on any face and merged with the bordering edges, where the merging can be regarded as attachments followed by contractions. For example, on a triangular face,



C.3 Polynomials with F-symbols as roots

$G = \mathbb{Z}_7$

$$
\begin{aligned}
P_y^{\mathbb{Z}_7}(y) =& \, 117649y^{12} - 453789y^{11} + 1145277y^{10} - 1070503y^9 + 882588y^8 - 284732y^7 \\
& - 89977y^6 + 31488y^5 - 1828y^4 - 849y^3 + 381y^2 + 45y - 1 \,, \\
P_z^{\mathbb{Z}_7}(z) =& \, 343z^6 + 196z^5 - 371z^4 + 27z^3 + 56z^2 - 9z - 1 \,, \\
P_w^{\mathbb{Z}_7}(w) =& \, 49w^4 - 63w^3 + 15w^2 + 10w - 4 \,.
\end{aligned}
$$

$G = \mathbb{Z}_9$

$$
\begin{aligned}
P_y^{\mathbb{Z}_9}(y) =& \, 282429536481y^{24} - 2541865828329y^{23} + 13891349053584y^{22} - 42375665666331y^{21} \\
& + 93048845085738y^{20} - 163017616751046y^{19} + 191382870385035y^{18} \\
& - 91749046865085y^{17} - 71565147070767y^{16} + 121393466114850y^{15} \\
& - 42556511453652y^{14} - 23330326470255y^{13} + 20787803433577y^{12} \\
& - 1805958554210y^{11} - 2533403044422y^{10} + 632950992624y^9 \\
& + 91558817982y^8 - 30315392921y^7 - 4655443748y^6 + 986603649y^5 \\
& + 182920180y^4 - 28268573y^3 - 1118977y^2 - 127236y - 1801 \,, \\
P_z^{\mathbb{Z}_9}(z) =& \, 531441z^{12} + 885735z^{11} - 1535274z^{10} - 121014z^9 + 647352z^8 - 79407z^7 \\
& - 92863z^6 + 18139z^5 + 4928z^4 - 1208z^3 - 64z + 25z - 1 \,, \\
P_w^{\mathbb{Z}_9}(w) =& \, 282429536481w^{24} - 1129718145924w^{23} + 1997927461773w^{22} - 1984755165147w^{21} \\
& + 1330918519878w^{20} - 791614850283w^{19} + 459695402118w^{18} - 222483700269w^{17} \\
& + 99182263023w^{16} - 47943836820w^{15} + 17026501158w^{14} - 3348784053w^{13} \\
& + 1374949378w^{12} - 621445880w^{11} - 329500476w^{10} + 412571852w^9 - 148134014w^8 \\
& + 18260969w^7 + 2110023w^6 - 806198w^5 + 47683w^4 + 6215w^3 - 711w^2 + 4w + 1 \,. \\
P_r^{\mathbb{Z}_9}(r) =& \, 6561r^8 - 8019r^7 + 1377r^6 - 792r^5 + 3349r^4 + 4r^3 - 662r^2 + 52r + 19 \,, \\
P_s^{\mathbb{Z}_9}(s) =& \, 81s^4 + 99s^3 + 17s^2 - 14s - 4 \,.
\end{aligned}
$$

$G = \mathbb{Z}_{11}$

For the unitary orbit with two solutions

$$
\begin{aligned}
P^{\mathbb{Z}_{11}}_{2|y}(y) &= 121y^4 - 209y^3 + 82y^2 + 24y - 9\,,\\
P^{\mathbb{Z}_{11}}_{2|z}(z) &= 11z^2 + 7z + 1\,,\\
P^{\mathbb{Z}_{11}}_{2|w}(w) &= 121w^4 - 88w^3 + 38w^2 - 13w + 1\,.
\end{aligned}
$$

For the unitary orbit with ten solutions,

$$
\begin{aligned}
P^{\mathbb{Z}_{11}}_{10|y}(y) = {} & 25937424601y^{20} - 47158953820y^{19} + 1064291844165y^{18} + 4808654315960y^{17}\\
& + 35564388240370y^{16} + 114903432126461y^{15} + 194232171940290y^{14}\\
& + 126582540515475y^{13} - 21851286302395y^{12} - 65093840585730y^{11}\\
& - 20230205549333y^{10} + 6813959963720y^{9} + 4785911566905y^{8} + 360322446200y^{7}\\
& - 303249779065y^{6} - 76228721396y^{5} - 379548930y^{4} + 2142467760y^{3}\\
& + 324308000y^{2} + 19299130y + 40207\,,\\
P^{\mathbb{Z}_{11}}_{10|z}(z) = {} & 161051z^{10} + 658845z^{9} - 971630z^{8} - 542080z^{7} + 322135z^{6}\\
& + 105612z^{5} - 39815z^{4} - 6570z^{3} + 1960z^{2} + 70z - 19\,,\\
P^{\mathbb{Z}_{11}}_{10|w}(w) = {} & 25937424601w^{20} - 176846076825w^{19} + 592702305965w^{18} - 1134445659765w^{17}\\
& + 1534818445765w^{16} - 1765089648718w^{15} + 1769544129045w^{14}\\
& - 1394768735745w^{13} + 776013578560w^{12} - 263088585485w^{11} + 20179458718w^{10}\\
& + 32370728245w^{9} - 20820136235w^{8} + 6982550700w^{7} - 1450721110w^{6}\\
& + 175316847w^{5} - 7539540w^{4} - 877925w^{3} + 133550w^{2} - 5960w + 71\,.
\end{aligned}
$$

$G = \mathbb{Z}_{13}$

$$\begin{aligned}
P_y^{\mathbb{Z}_{13}}(y) = {}& 23298085122481y^{24} + 80647217731665y^{23} + 3069557179509834y^{22} \\
&+ 41919543603471508y^{21} + 536909384312855190y^{20} + 4259352400707950897y^{19} \\
&+ 19179161744641728596y^{18} + 47561155144008593243y^{17} + 63626358551986353149y^{16} \\
&+ 40207662041712799114y^{15} + 12576352168592287 66y^{14} - 13522223195096193305y^{13} \\
&- 6598116247933199625y^{12} + 128413711306511340y^{11} + 938990747292838888y^{10} \\
&+ 202797783582401196y^{9} - 32756778784407789y^{8} - 16526752437401584y^{7} \\
&- 933201395423678y^{6} + 349378912529867y^{5} + 53761577382743y^{4} + 1555890743172y^{3} \\
&- 87453542726y^{2} - 2773486466y + 28678361\,, \\
P_z^{\mathbb{Z}_{13}}(z) = {}& 4826809z^{12} + 34901542z^{11} - 124183228z^{10} - 57416398z^{9} + 51122838z^{8} + 3476850z^{7} \\
&- 4988283z^{6} + 418090z^{5} + 93250z^{4} - 14139z^{3} + 205z^{2} + 38z - 1\,, \\
P_w^{\mathbb{Z}_{13}}(w) = {}& 23298085122481w^{24} - 268824059105550w^{23} + 1610738618763716w^{22} \\
&- 4730805028787149w^{21} + 8265875258850053w^{20} - 9798763675027379w^{19} \\
&+ 8948312751528579w^{18} - 6464842564613641w^{17} + 3087209293878385w^{16} \\
&- 284952516401007w^{15} - 771813881083466w^{14} + 531872957583864w^{13} \\
&- 107361616574952w^{12} - 39739582655570w^{11} + 27485052167132w^{10} \\
&- 4323332693485w^{9} - 1159653323459w^{8} + 583780092624w^{7} - 51758752951w^{6} \\
&- 19939454943w^{5} + 4746063302w^{4} + 131285807w^{3} - 111025779w^{2} + 2170222w \\
&+ 898159\,, \\
P_s^{\mathbb{Z}_{13}}(s) = {}& 28561s^{8} - 24167s^{7} + 163930s^{6} - 225693s^{5} + 119817s^{4} - 26999s^{3} + 1045s^{2} + 546s \\
&- 67\,.
\end{aligned}$$

Appendix D

APPENDICE TO CHAPTER 5

D.1 The F-symbols for the Haagerup $\mathcal{H}_3$ fusion category

This appendix presents the F-symbols for the transparent Haagerup $\mathcal{H}_3$ fusion category found in [102]. We first present the unitary gauge, and then transit to a slightly more convenient gauge for this note.

Let $I = \{\mathcal{I}, \alpha, \alpha^2\}$ be the set of invertible objects, $N = \{\rho, \alpha\rho, \alpha^2\rho\}$ be the set of non-invertible simple objects of the Haagerup fusion ring, and define $\zeta \equiv \frac{3+\sqrt{13}}{2}$. For a unitary fusion category, the F-symbols involving at least one invertible object can be set to

$$(F^{\eta\mathcal{L}\theta,\overline{\theta}\mathcal{L},\mathcal{L}}_{\eta\mathcal{L}})_{\eta,\overline{\theta}} \;=\; \zeta^{-1}\,, \qquad (F^{\mathcal{L}_1,\mathcal{L}_3,\eta\mathcal{L}_3}_{\mathcal{L}_1\overline{\eta}})_{\mathcal{L}_2,\overline{\eta}} \;=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_1,\eta\mathcal{L}_3}_{\mathcal{L}_3})_{\overline{\eta},\mathcal{L}_2} \;=\; \zeta^{-\frac{1}{2}}\,, \tag{D.1}$$

where $\eta, \theta \in I$ and $\mathcal{L}_i \in N$. The remaining nontrivial F-symbols are the ones where all six simple objects are non-invertible. It suffices to specify the nine components $(F^{\rho,\rho,\rho}_{*})_{\rho,*}$ with $*$ running over the non-invertible simple objects, since via the transparency relations

$$\begin{aligned}(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;&=\; (F^{\eta\mathcal{L}_1,\eta\mathcal{L}_2,\eta\mathcal{L}_3}_{\eta\mathcal{L}_4})_{\eta\mathcal{L}_5,\eta\mathcal{L}_6} \;=\; (F^{\eta\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\eta}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \\ &=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4\eta})_{\mathcal{L}_5,\mathcal{L}_6} \;=\; (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\eta\mathcal{L}_5,\mathcal{L}_6\eta}\end{aligned} \tag{D.2}$$

the values of all other F-symbols are determined. An equivalent, and sometimes more convenient, set of relations that also allows the generation of all F-symbols is

$$\begin{aligned}(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} \;&=\; (F^{\mathcal{L}_1\overline{\eta},\eta\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\eta\mathcal{L}_6} \;=\; (F^{\mathcal{L}_1,\mathcal{L}_2\overline{\eta},\eta\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5\overline{\eta},\mathcal{L}_6} \\ &=\; (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\overline{\eta}}_{\mathcal{L}_4\overline{\eta}})_{\mathcal{L}_5,\mathcal{L}_6\overline{\eta}} \;=\; (F^{\eta\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\eta\mathcal{L}_4})_{\eta\mathcal{L}_5,\mathcal{L}_6}\,.\end{aligned} \tag{D.3}$$

We provide an explicit algorithm to turn any nontrivial F-symbol into $(F^{\rho,\rho,\rho}_{*})_{\rho,*}$ form.

1. Use $(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} = (F^{\eta\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\eta}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6}$ to turn $\mathcal{L}_1$ into ρ.
2. Use $(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} = (F^{\mathcal{L}_1,\eta\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4\eta})_{\mathcal{L}_5,\mathcal{L}_6}$ to turn $\mathcal{L}_2$ into ρ.
3. Use $(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} = (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3\overline{\eta}}_{\mathcal{L}_4\overline{\eta}})_{\mathcal{L}_5,\mathcal{L}_6\overline{\eta}}$ to turn $\mathcal{L}_3$ into ρ.
4. Use $(F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\mathcal{L}_5,\mathcal{L}_6} = (F^{\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3}_{\mathcal{L}_4})_{\eta\mathcal{L}_5,\mathcal{L}_6\eta}$ to turn $\mathcal{L}_5$ into ρ.

Solving the pentagon identity in the gauge (D.1) and under transparency and S_4 full tetrahedral symmetry, there are exactly two solutions, both of which are unitary. One of them is given by

$$\begin{array}{c|ccc} (F^{\rho,\rho,\rho}_{*})_{\rho,*} & \rho & \alpha\rho & \alpha^2\rho \\ \hline \rho & x & y_1 & y_2 \\ \alpha\rho & y_1 & y_2 & z \\ \alpha^2\rho & y_2 & z & y_1 \end{array} \tag{D.4}$$

where

$$x = \frac{2-\sqrt{13}}{3}\,, \quad y_{1,2} = \frac{1}{12}\left(5-\sqrt{13}\mp\sqrt{6\left(1+\sqrt{13}\right)}\right)\,, \quad z = \frac{1+\sqrt{13}}{6}\,. \tag{D.5}$$

The other solution is related by the $\mathrm{Aut}(\mathbb{Z}_3) \cong \mathbb{Z}_2$ action that exchanges y_1 and y_2. For the first solution, some of the F-symbols can be presented as

$$\begin{aligned}
F^{\rho_i,\rho_j,\rho_j}_{\rho_i} &= \begin{pmatrix} \zeta^{-1} & \zeta^{-\frac{1}{2}} & \zeta^{-\frac{1}{2}} & \zeta^{-\frac{1}{2}} \\ \zeta^{-\frac{1}{2}} & \mathfrak{f}_{i+j} & \mathfrak{f}_{i+j-1} & \mathfrak{f}_{i+j-2} \\ \zeta^{-\frac{1}{2}} & \mathfrak{f}_{i+j-1} & \mathfrak{f}_{i+j-2} & \mathfrak{f}_{i+j} \\ \zeta^{-\frac{1}{2}} & \mathfrak{f}_{i+j-2} & \mathfrak{f}_{i+j} & \mathfrak{f}_{i+j-1} \end{pmatrix}\,, \qquad \mathfrak{f}_0 = x\,, \quad \mathfrak{f}_1 = y_2\,, \quad \mathfrak{f}_2 = y_1\,, \\
F^{\rho_i,\rho_j,\rho_j}_{\rho_{k\neq i}} &= \begin{pmatrix} \mathfrak{f}'_{j-i-k} & \mathfrak{f}'_{j-i-k-1} & \mathfrak{f}'_{j-i-k-2} \\ \mathfrak{f}'_{j-i-k-1} & \mathfrak{f}'_{j-i-k-2} & \mathfrak{f}'_{j-i-k} \\ \mathfrak{f}'_{j-i-k-2} & \mathfrak{f}'_{j-i-k} & \mathfrak{f}'_{j-i-k-1} \end{pmatrix}\,, \qquad \mathfrak{f}'_0 = z\,, \quad \mathfrak{f}'_1 = y_2\,, \quad \mathfrak{f}'_2 = y_1\,, \\
F^{\rho_i,\rho_j,\rho_i}_{\rho_{j\neq i}} &= \begin{pmatrix} \mathfrak{f}'_{i+j} & \mathfrak{f}'_{i+j-1} & \mathfrak{f}'_{i+j-2} \\ \mathfrak{f}'_{i+j-1} & \mathfrak{f}'_{i+j-2} & \mathfrak{f}'_{i+j} \\ \mathfrak{f}'_{i+j-2} & \mathfrak{f}'_{i+j} & \mathfrak{f}'_{i+j-1} \end{pmatrix}\,,
\end{aligned} \tag{D.6}$$

where the subscripts of $\mathfrak{f}$ and $\mathfrak{f}'$ are defined modulo 3.

In this note we adopt a different, non-unitary gauge to eliminate the appearance of some factors of $\zeta^{-\frac{1}{2}}$ in the bootstrap equations. The only difference from the unitary gauge is that (D.1) is replaced by

$$(F^{\eta\mathcal{L}\theta,\overline{\mathcal{L}\theta},\mathcal{L}}_{\eta\mathcal{L}})_{\eta,\overline{\theta}} \;=\; (F^{\mathcal{L}_1,\eta\mathcal{L}_1,\eta\mathcal{L}_3}_{\overline{\mathcal{L}_3}})_{\overline{\eta},\mathcal{L}_2} \;=\; \zeta^{-1}\,, \qquad (F^{\mathcal{L}_1,\mathcal{L}_3,\eta\mathcal{L}_3}_{\overline{\eta\mathcal{L}_1}})_{\mathcal{L}_2,\overline{\eta}} \;=\; 1\,. \tag{D.7}$$

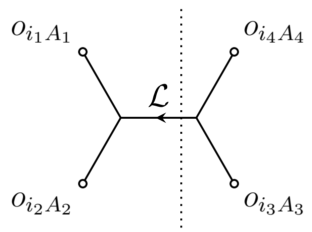
Consequently, the 4×4 F-symbols become

$$F^{\rho_i,\rho_j,\rho_j}_{\rho_i} = \begin{pmatrix} \zeta^{-1} & \zeta^{-1} & \zeta^{-1} & \zeta^{-1} \\ 1 & \mathfrak{f}_{i+j} & \mathfrak{f}_{i+j-1} & \mathfrak{f}_{i+j-2} \\ 1 & \mathfrak{f}_{i+j-1} & \mathfrak{f}_{i+j-2} & \mathfrak{f}_{i+j} \\ 1 & \mathfrak{f}_{i+j-2} & \mathfrak{f}_{i+j} & \mathfrak{f}_{i+j-1} \end{pmatrix}\,, \tag{D.8}$$

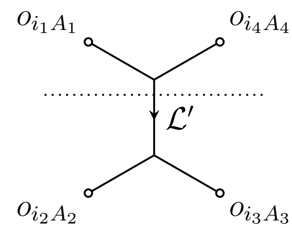
while the 3×3 F-symbols remain the same.

D.2 Crossing symmetry of ρ defect operators

General crossing symmetry involving topological defect lines (TDLs) was discussed in Section 5.2. In search for a defect topological field theory (TFT) whose TDLs realize the Haagerup $\mathcal{H}_3$ fusion category, the subset of crossing symmetry constraints that are equivalent to the associativity with at least one local operator was delineated in Section 5.5, and solved in Section 5.6 to obtain part of the defining data of the TFT. In this appendix, we study other crossing symmetry constraints that encode more data of the TFT. These constraints can be depicted graphically as



$$= \sum_{\mathcal{L}'}$$



$$(F^{\rho_{i_1},\rho_{i_2},\rho_{i_3}}_{\rho_{i_4}})_{\mathcal{L},\mathcal{L}'}\,, \tag{D.9}$$

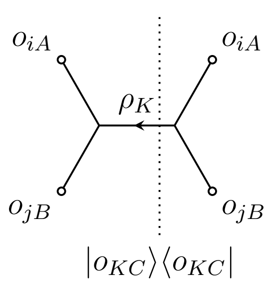
and cutting along the dotted line gives

$$\begin{aligned} &\sum_{\mathcal{O}\in\mathcal{H}_{\mathcal{L}}} c(o_{i_1A_1}, o_{i_2A_2}, \mathcal{O})\, c(o_{i_3A_3}, o_{i_4A_4}, \overline{\mathcal{O}}) \\ &= \sum_{\mathcal{L}'} (F^{\rho_{i_1},\rho_{i_2},\rho_{i_3}}_{\rho_{i_4}})_{\mathcal{L},\mathcal{L}'} \sum_{\mathcal{O}'\in\mathcal{H}_{\mathcal{L}'}} c(o_{i_2A_2}, o_{i_3A_3}, \mathcal{O}')\, c(o_{i_4A_4}, o_{i_1A_1}, \overline{\mathcal{O}}')\,. \end{aligned} \tag{D.10}$$

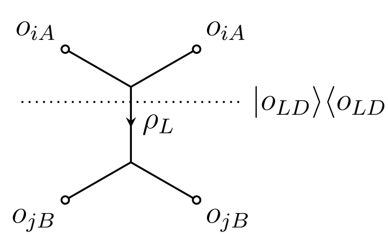
Depending on the quadruple (i_1, i_2, i_3, i_4), the internal TDLs $\mathcal{L}$, $\mathcal{L}'$ run over either the three non-invertible TDLs $\rho_0 \equiv \rho,\ \rho_1 \equiv \alpha\rho,\ \rho_2 \equiv \alpha^2\rho$, or an additional invertible TDL. It is convenient to introduce a capital I index such that $\rho_{I=-1}$ denotes the invertible TDL whenever applicable, and $\rho_{I=i} = \rho_i$ for $i = 0, 1, 2$. In particular, if $\rho_{I=-1}$ is the trivial TDL $\mathcal{I}$, then $o_{I=-1,A}$ with $A = 1, \ldots, n_{\rm V}$ represent the local operators.

Two pairs of identical external operators o_{iA} and o_{jB} in the 1221 configuration

In this case, the defect crossing equation (D.10) in the newly introduced notation reads



$$= \sum_{L=-1}^{2}$$



$$(F^{\rho_{i_1},\rho_{i_2},\rho_{i_3}}_{\rho_{i_4}})_{\rho_K,\rho_L}\,,$$

$$\sum_{C=-1}^{2} |c(o_{iA}, o_{jB}, o_{KC})|^2 = \sum_{L=-1}^{2} (F^{\rho_i,\rho_j,\rho_j}_{\rho_i})_{\rho_K,\rho_L} \sum_{D} c(o_{iA}, o_{iA}, o_{LD})\, c(o_{jB}, o_{jB}, o_{LD})\,. \tag{D.11}$$

- Setting $K=-1$ gives

$$\sum_C c(o_{iA},o_{jB},o_{-1,C})^2 = \zeta^{-1}\sum_D c(o_{iA},o_{iA},o_{-1,D})\,c(o_{jB},o_{jB},o_{-1,D}) + \zeta^{-1}\sum_{\ell=0}^{2}\sum_D c(o_{iA},o_{iA},o_{\ell D})\,c(o_{jB},o_{jB},o_{\ell D})\,, \tag{D.12}$$

where we have used the explicit values of F-symbols given in (D.8).

- If we sum (D.11) over $K=k=0,1,2$ (but not $K=-1$), and use the explicit values of F-symbols given in (D.8), then we obtain[1]

$$\sum_{k=0}^{2}\sum_C |c(o_{iA},o_{jB},o_{kC})|^2 = 3\sum_D c(o_{iA},o_{iA},o_{-1,D})\,c(o_{jB},o_{jB},o_{-1,D}) - \zeta^{-1}\sum_{\ell=0}^{2}\sum_D c(o_{iA},o_{iA},o_{\ell D})\,c(o_{jB},o_{jB},o_{\ell D})\,. \tag{D.14}$$

- Let us set $i=j$ and $A\neq B$. Using the explicit values of κ^i_{AB} and $\lambda^i_{AB;a}$ in (5.112) and (5.139) to evaluate the contributions from local operators,

$$\sum_C c(o_{iA},o_{iB},o_{-1,C})^2 = \delta_{AB} + (\kappa^i_{AB})^2 + 2\sum_a \lambda^i_{AB;a}\bar{\lambda}^i_{AB;a} = \begin{cases} 3\sigma_{111}^2 & A=B\,,\\ 0 & A\neq B\,,\end{cases} \tag{D.15}$$

$$\begin{aligned}&\sum_D c(o_{iA},o_{iA},o_{-1,D})\,c(o_{jB},o_{jB},o_{-1,D})\\ &= 1+\kappa^i_{AA}\kappa^j_{BB} + \sum_a\left(\lambda^i_{AA;a}\bar{\lambda}^j_{BB;a} + \bar{\lambda}^i_{AA;a}\lambda^j_{BB;a}\right) = \begin{cases} 3\sigma_{111}^2 & A-i=B-j\,,\\ 0 & A-i\neq B-j\,,\end{cases}\end{aligned} \tag{D.16}$$

where

$$\sigma_{111} = \sqrt{1+\zeta^{-2}}\,, \tag{D.17}$$

the preceding two equations (D.12) and (D.14) become

$$\begin{aligned} 0 &= \sum_{\ell=0}^{2}\sum_D c(o_{iA},o_{iA},o_{\ell D})\,c(o_{iB},o_{iB},o_{\ell D})\,,\\ \sum_{k=0}^{2}\sum_C c(o_{iA},o_{iB},o_{kC})^2 &= -\zeta^{-1}\sum_{\ell=0}^{2}\sum_D c(o_{iA},o_{iA},o_{\ell D})\,c(o_{iB},o_{iB},o_{\ell D})\,.\end{aligned} \tag{D.18}$$

[1] We used

$$\sum_{k=0}^{2}(F^{\rho_i,\rho_j,\rho_j}_{\rho_i})_{\rho_k,\mathcal{I}} = 3\,, \qquad \sum_{k=0}^{2}(F^{\rho_i,\rho_j,\rho_j}_{\rho_i})_{\rho_k,\rho_\ell} = x+y_+ + y_- = -\zeta^{-1}\quad \forall\,\ell\,. \tag{D.13}$$

It follows that

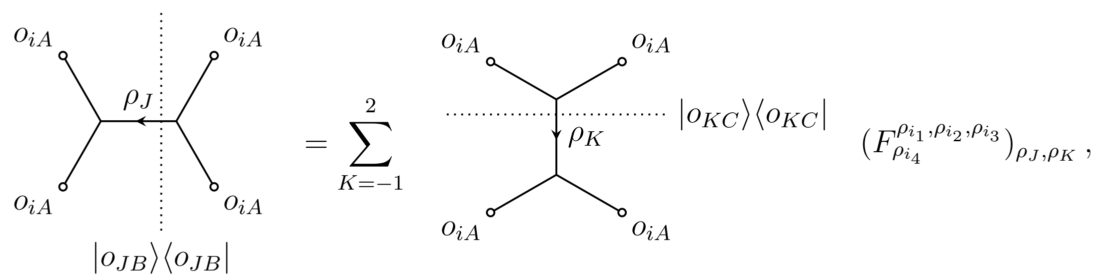
$$\sum_{k=0}^{2} \sum_{C} c(o_{iA}, o_{iB}, o_{kC})^2 = 0\,, \tag{D.19}$$

and we arrive at a selection rule:

$$c(o_{iA}, o_{iB}, o_{kC}) = 0 \quad \forall\, i,\, k,\, C \quad \text{if } A \neq B. \tag{D.20}$$

Four identical external defect operators o_{iA}

In this case, the defect crossing equation (D.10) becomes



$$\sum_{K=-1}^{2} \left[\delta_{JK}\delta_{BC} - (F^{\rho_i,\rho_i,\rho_i}_{\rho_i})_{\rho_J,\rho_K}\right] \sum_{C} c(o_{iA}, o_{iA}, o_{KC})^2 = 0\,, \tag{D.21}$$

which says that the four-dimensional vector $\sum_C c(o_{iA}, o_{iA}, o_{KC})^2$ is a non-negative four-dimensional eigenvector of the matrix $F^{\rho_i,\rho_i,\rho_i}_{\rho_i}$ with eigenvalue one. Using the explicit values of the F-symbols given in (D.8), we determine that such an eigenvector is in the two-dimensional subspace spanned by

$$(1+\zeta,\, 1,\, 1,\, 1)\;, \qquad \begin{cases} (0,\, 1,\, \psi_+,\, \psi_-) & i = 0\,, \\ (0,\, \psi_-,\, 1,\, \psi_+) & i = 1\,, \\ (0,\, \psi_+,\, \psi_-,\, 1) & i = 2\,, \end{cases} \tag{D.22}$$

where

$$\psi_\pm = \frac{-1 \pm \sqrt{7 + 2\sqrt{13}}}{2}\,. \tag{D.23}$$

Equivalently, $\sum_C c(o_{iA}, o_{iA}, o_{KC})^2$ is orthogonal to

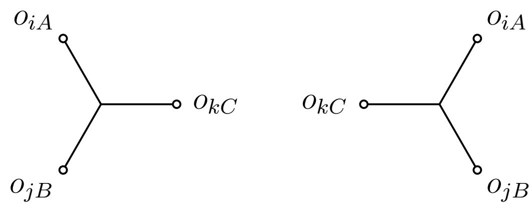
$$(-1+\zeta^{-1},\, 1,\, 1,\, 1)\,, \qquad \begin{cases} (0,\, 1,\, \eta_-,\, \eta_+) & i = 0\,, \\ (0,\, \eta_+,\, 1,\, \eta_-) & i = 1\,, \\ (0,\, \eta_-,\, \eta_+,\, 1) & i = 2\,, \end{cases} \tag{D.24}$$

where

$$\eta_\pm = \frac{-1 \pm \sqrt{3\left(2\sqrt{13} - 7\right)}}{2}\,. \tag{D.25}$$

Two pairs of identical external operators o_{iA} and o_{jB} in the 1212 configuration

We first recall from (5.17) and (5.18) that the defect three-point coefficients are invariant under cyclic permutations and complex conjugate under reflections. Thus the trivalent vertices



are complex conjugates of each other, and can differ by a phase $2\phi_{ijk}$. The corresponding three-point coefficients of defect operators can be parameterized as

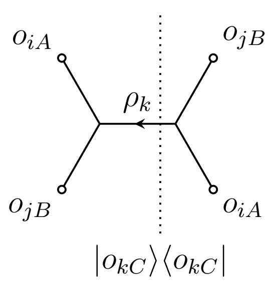
$$\begin{aligned} c(o_{iA}, o_{jB}, o_{kC}) &= |c(o_{iA}, o_{jB}, o_{kC})| e^{i\phi_{ijk}} \,, \\ c(o_{jB}, o_{iA}, o_{kC}) &= |c(o_{iA}, o_{jB}, o_{kC})| e^{-i\phi_{ijk}} \,. \end{aligned} \tag{D.26}$$

Since the phase is trivial when two indices coincide, $\phi_{iij} = 0$, the only nontrivial phase is $\phi \equiv \phi_{012}$. Let us define

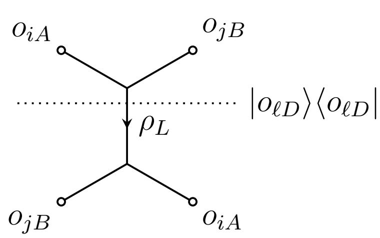
$$\Phi_{ij} = \begin{pmatrix} e^{i\phi_{ij0}} & & \\ & e^{i\phi_{ij1}} & \\ & & e^{i\phi_{ij2}} \end{pmatrix} , \tag{D.27}$$

which is the identity matrix if $i = j$, and has a single possibly nontrivial entry if $i \neq j$.

In the current case, the crossing equation (D.10) becomes



$$= \sum_{\ell=0}^{2}$$



$$(F^{\rho_{i_1},\rho_{i_2},\rho_{i_3}}_{\rho_{i_4}})_{\rho_k,\rho_\ell} \,, \tag{D.28}$$

$$\sum_C c(o_{iA}, o_{jB}, o_{kC})^2 = \sum_\ell (F^{\rho_i,\rho_j,\rho_i}_{\rho_j})_{\rho_k,\rho_\ell} \sum_D c(o_{jB}, o_{iA}, o_{\ell D})^2 \,.$$

By factoring out the phase using (D.26), the above crossing equation can be reexpressed as

$$\sum_C |c(o_{iA}, o_{jB}, o_{kC})|^2 = \sum_{\ell=0}^{2} (\bar{\Phi}^2_{ij} \, F^{\rho_i,\rho_j,\rho_i}_{\rho_j} \, \bar{\Phi}^2_{ij})_{k\ell} \sum_D |c(o_{iA}, o_{jB}, o_{\ell D})|^2 \,. \tag{D.29}$$

The three-dimensional vector $\sum_C |c(o_{iA}, o_{jB}, o_{kC})|^2$, if nonzero, is a non-negative eigenvector of $\bar{\Phi}^2_{ij}\, F^{\rho_i,\rho_j,\rho_i}_{\rho_j}\, \bar{\Phi}^2_{ij}$ with eigenvalue 1.

- If ϕ is a generic phase, by which we mean $\phi \neq 0,\, \pi$, then such an eigenvector is unique up to overall normalization, given by

$$\begin{cases} (\psi_+,\, -\psi_-,\, 0) & \{i,j\} = \{0,1\}\,, \\ (0,\, \psi_+,\, -\psi_-) & \{i,j\} = \{1,2\}\,, \\ (-\psi_-,\, 0,\, \psi_+) & \{i,j\} = \{0,2\}\,, \end{cases} \tag{D.30}$$

  which in particular implies that

$$c(o_{0A}, o_{1B}, o_{2C}) = 0\,. \tag{D.31}$$

  In other words, the only three-point coefficient that is allowed to have a nontrivial phase vanishes. Then without loss of generality, we can assume $\phi = 0,\, \pi$.

- If $\phi = 0,\, \pi$, then the eigenvector is in the two-dimensional subspace that is orthogonal to

$$\begin{cases} \left(\psi_-,\, \psi_+,\, e^{i\phi}\right) & \{i,j\} = \{0,1\}\,, \\ \left(e^{i\phi},\, \psi_-,\, \psi_+\right) & \{i,j\} = \{1,2\}\,, \\ \left(\psi_+,\, e^{i\phi},\, \psi_-\right) & \{i,j\} = \{0,2\}\,. \end{cases} \tag{D.32}$$